# Subjunctive Reasoning

*John L. Pollock*

SUBJUNCTIVE REASONING

PHILOSOPHICAL STUDIES SERIES
IN PHILOSOPHY



VOLUME 8

JOHN L. POLLOCK
*University of Rochester*

# SUBJUNCTIVE REASONING

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND/BOSTON-U.S.A.

Library of Congress Cataloging in Publication Data

Pollock, John L
Subjunctive reasoning.

(Philosophical studies series in philosophy ; v. 8)
Bibliography: p.
Includes index.
1. Conditionals (Logic) 2. Reasoning. 3. Counterfactuals (Logic) 4. Probabilities. I. Title.
BC199.C56P64 160 76-19095
ISBN 90-277-0701-4

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
Lincoln Building, 160 Old Derby Street, Hingham,
Mass. 02043, U.S.A.



Printed in The Netherlands

*TO CAROL*
*who puts up*
*with me*

# TABLE OF CONTENTS

# PREFACE

I am indebted to many people for the help they gave me in the writing of this book. I owe a large debt to David Lewis and Robert Stalnaker, on both general and specific grounds. As becomes apparent from reading the notes, the book would not have been possible without their pioneering work on subjunctive conditionals. In addition, both were kind enough to provide specific comments on earlier versions of different parts of the book, and Stalnaker read and commented on the entire manuscript. Closer to home, I am indebted to my colleagues Rolf Eberle and Henry Kyburg, Jr., my erstwhile colleague Keith Lehrer, and numerous graduate students for their helpful comments on various parts of the manuscript. Some of the material contained herein appeared first in the form of journal articles, and I wish to thank the journals in question for allowing the material to be reprinted here. Chapter One contains material taken from 'The "Possible Worlds" Analysis of Counter-factuals', published in *Phil. Studies* **29** (1976), 469 (Reidel); Chapter Two contains material much revised from 'Four Kinds of Conditionals', *Am. Phil. Quarterly* **12** (1975), and Chapter Three contains much revised material from 'Subjunctive Generalizations', *Synthese* **28** (1974), 199 (Reidel).

CHAPTER I

# INTRODUCTION

## 1. Subjunctive reasoning

There exists quite a variety of statements which are in some sense 'subjunctive'. The best known of these are the so-called 'counterfactual conditionals' which state that if something which is not the case had been the case, then something else would have been true. An example is 'If Kennedy had been president in 1972, the Watergate scandal would not have occurred'. Ordinary people use counterfactuals all the time, and philosophers use them freely in ordinary situations. However, when they are being careful, philosophers have traditionally felt uncomfortable about counterfactuals and eschewed their employment in philosophical analysis. Such philosophical squeamishness is on the whole meritorious and results from the recognition that counterfactual conditions have themselves stubbornly resisted philosophical analysis.

Although counterfactual conditionals constitute the kind of subjunctive statement which comes most readily to the mind of a philosopher, it is far from being the only philosophically important kind of subjunctive statement. Philosophers have long recognized that laws of nature cannot really be formulated using universally quantified material conditionals, but they have not usually been prepared to go the extra distance of admitting that statements expressing such laws are really subjunctive. It turns out that laws of nature must be formulated using a special kind of subjunctive statement, herein called a 'subjunctive generalization'. Subjunctive generalizations prove to be of pre-eminent importance in discussing inductive confirmation.

Causal statements constitute another category of statements which are in the requisite sense 'subjunctive'. The analysis of causal statements has always been recognized to be an extraordinarily difficult philosophical problem, but I think it has rarely been appreciated that the main source of this difficulty lies in the subjunctive nature of causal statements.

A traditional philosophical problem is the analysis of probability statements. There are in fact a number of different concepts equally deserving of being called 'probability'. There is not just one legitimate concept of probability. Philosophers have succeeded in sorting out and distinguishing between a number of different probability concepts, including 'degree of confirmation', 'degree of belief', 'degree of rational belief', and others. However, there are many probability statements which cannot be formulated using any of these recognized 'indicative' probability concepts. It will turn out that there are some extremely important probability concepts that are in essential ways 'subjunctive' and which have been almost entirely overlooked by philosophers bent upon the avoidance of suspicious (to them) subjunctive reasoning.

A problematic concept which has usually been recognized to have a subjunctive core is that of a disposition. Dispositions constitute an important tool of philosophical analysis. Particularly in the philosophy of mind, philosophers have felt that through the use of dispositions they could clarify the structure of interesting concepts. But, of course, the extent of such clarification has been strictly limited by the apparent need for clarifying dispositions themselves.

These various subjunctive concepts map out areas of what might be called 'subjunctive reasoning'. Subjunctive reasoning in general has been deemed philosophically problematic because it seems to presuppose a strange metaphysically suspicious sort of logically contingent necessity. To say that the Watergate scandal would not have occurred had Kennedy been president in 1972, seems to be to assert some kind of necessary connection between those two states of affairs. If there were no such connection, how could the occurrence of the one possibly effect the occurrence of the other? This same kind of necessity rears its ugly head repeatedly throughout subjunctive reasoning. The necessity in question is clearly not logical necessity, but what other kind is there? Surely this is a very suspicious notion, and philosophers would be well advised to avoid subjunctive concepts at all cost! But upon sober reflection, it must be admitted that this attitude is preposterous. Subjunctive concepts do make sense – we use them all the time. The problem cannot be whether they make sense, but what sense they

make. The real issue must be how they are to be analyzed, and not their legitimacy. We cannot in good conscience deny that there is this logically contingent kind of necessity which is involved in subjunctive reasoning. We cannot expunge it from our conceptual framework without leaving that framework seriously impoverished. What we must seek is an understanding of subjunctive reasoning, an analysis of subjunctive concepts in terms of other clearer concepts. Thus the task of this book will be to provide an analysis of the above subjunctive concepts in terms of relatively less problematic indicative concepts. If successful, such analyses will free philosophers to use subjunctive concepts with impunity in the course of philosophical analysis. There can be no doubt that subjunctive concepts, clearly understood, would constitute an extremely powerful logical tool.

It will turn out that, in important respects, the key to understanding subjunctive reasoning in general is to understand subjunctive conditionals. There are a number of reasons why subjunctive conditionals have so stubbornly resisted philosophical analysis, not the least of which is that our language systematically conflates two importantly different kinds of subjunctive conditionals. However, in recent years the first chinks have appeared in the armor protecting subjunctive conditionals from philosophical understanding. These chinks have resulted in large part from the philosophical onslaughts of Robert Stalnaker (1968, 1972) and David Lewis, (1972, 1973, 1973a). Building upon the insights which they have wrought, I believe that it may prove possible to achieve a complete understanding of subjunctive conditionals, and thereby of subjunctive reasoning in general.

Historically, there have been two popular approaches to the analysis of subjunctive conditionals. These are the older 'linguistic' approach, and the more recent 'possible worlds' approach initiated by Stalnaker and Lewis. In the rest of this introductory chapter I will examine briefly both of these approaches and use this examination to set the stage for my own analysis which builds upon them both. The plan of the book will then be to complete the analysis of subjunctive conditionals, in the course of which I will propose an analysis of subjunctive generalizations, and then to apply our results on subjunctive conditionals to the analysis of causes, probabilities, and dispositions.

## 2. The linguistic approach

This is a traditional approach to the analysis of subjunctive conditionals according to which they are to be explained ultimately in terms of physical laws. Let us symbolize the subjunctive conditional ⌜If it were true that $P$, then it would be true that $Q$⌝ as ⌜$P > Q$⌝. Then according to the linguistic theory, ⌜$P > Q$⌝ is true just in case there is a physical law, or conjunction of physical laws $L$, and there are true circumstances $C$, such that the conjunction ⌜$C$ & $L$ & $P$⌝ logically entails $Q$. For example, suppose we have a dry match under ordinary circumstances. If it were struck ($P$), it would light ($Q$). The reason ⌜$P > Q$⌝ is true, according to this theory, is that there are physical laws regarding the chemical structure of the match which, when taken together with the circumstances in which we find the match (e.g., the match is dry, there is plenty of oxygen, etc.), and taken together with the match's being struck, entail that the match will light.

As we have stated it, the linguistic theory would make too many subjunctive conditionals true. Without some restriction on what we can put into the circumstances $C$, it would turn out that whenever the material conditional ⌜$P \supset Q$⌝ is true, so is the subjunctive conditional ⌜$P > Q$⌝. This is because if ⌜$P \supset Q$⌝ is true, we could put it among the circumstances $C$, and then $C$ together with $P$ (without even making use of any laws) would automatically entail $Q$. Thus some restrictions must be put upon the contents of $C$. It is surprisingly difficult to see what restrictions are called for.

One popular view regarding what restrictions should be placed on $C$ is that it is entirely a matter of convention. This was advocated by Chisholm (1955). Chisholm suggested that we are free to select any true statements we want to put into $C$, calling the selected ones our 'presuppositions', and maintained that the subjunctive conditional is correspondingly ambiguous. He supported his contention as follows:

> Let us suppose a man accepts the following statements, taking the universal statements to be law statements: (1) All gold is malleable; (2) No cast-iron is malleable; (3) Nothing is both gold and cast-iron; (4) Nothing is both malleable and not malleable; (5) That is cast-iron; (6) That is not gold; and (7) That is not malleable. We may contrast three different situations in which he asserts three different counterfactuals having the same antecedents.
>
> First, he asserts, pointing to an object his hearers don't know to be gold and don't know not to be gold, 'If that *were* gold, it would be malleable'. In this case, he is

supposing the denial of (6); he is excluding from his presuppositions (5), (6), and (7); and he is concerned to emphasize (1).

Secondly, he asserts, pointing to an object he and his hearers agree to be cast-iron, 'If *that* were gold, then some gold things would not be malleable'. He is again supposing the denial of (6); he is excluding (1) and (6), but he is no longer excluding (5) or (7); and he is concerned to emphasize either (5) or (2).

Thirdly, he asserts, 'If that were gold, then some things would be both malleable and not malleable'. He is again supposing the denial of (6); he is now excluding (3) and no longer excluding (1), (5), (6), or (7); and he is now concerned to emphasize (1), (2), or (5).

Still other possibilities readily suggest themselves (Chisholm, 1955, p. 103).

Chisholm is certainly right that the first two conditionals he mentions are both true. But the third cannot be true. There are no circumstances under which one could reasonably assert, 'If that were gold, then some things would be both malleable and not malleable.' Similarly, one could not reasonably assert of the cast-iron. 'If that were gold, it would not be malleable' (emphasizing (7) and excluding (1) from our presuppositions). Apparently there have to be some constraints on what goes into *C*. It cannot be entirely a matter of convention.

But the view that the contents of *C* are at least partly conventional is one that has recurred many times in the discussion of counterfactuals. Stalnaker (1968) calls the ambiguity stemming from this supposedly conventional aspect of *C* 'pragmatic ambiguity'. The view that subjunctive conditionals are subject to pragmatic ambiguity seems to be supported by the fact that one could assert either of Chisholm's first two conditionals:

(A) If that *were* gold, it would be malleable.

(B) If *that* were gold, then some gold things would not be malleable.

It seems that the set *C* must differ from (A) to (B). But, popular though this view is, I think it is wrong. Although (A) and (B) *look*, at least at first, like they have the same antecedent, they do not. The antecedents consist of the same words in the same order, but there is a difference in what word is emphasized, and that difference in emphasis makes a difference in meaning. It seems eminently reasonable to paraphrase (B) as follows:

(B*) If some gold things were like that (i.e., had the properties that has), then some gold things would not be malleable.

The antecedent of (B*) is quite different from that of (A), and it is not unreasonable to maintain that (B*) is what we *mean* when we write (B).

There are many other pairs of conditionals which have traditionally been thought to illustrate the pragmatic ambiguity of subjunctive conditionals. One fruitful source consists of what have been called 'counter-identicals', which are conditionals whose antecedents are false identity statements. For example, the following two conditionals appear to have logically equivalent antecedents and incompatible consequents:

(C) If Richard Nixon were Golda Meir, he would be a woman.

(D) If Golda Meir were Richard Nixon, she would be a man.

But these conditionals do not really have equivalent antecedents. Contrary to initial appearance, they are not counter-identicals at all. This can be seen by contrasting them with the following conditional which really is a counter-identical:

(E) If Richard Nixon and Golda Meir were really one and the same person (who had been fooling us by flying back and forth between countries and wearing makeup, etc.), then the United States would be less friendly towards the Arab nations.

Unlike (E), it seems that (C) and (D) can be paraphrased roughly as follows:

(C*) If Richard Nixon had the non-relational properties presently possessed by Golda Meir, he would be a woman.

(D*) If Golda Meir had the non-relational properties presently possessed by Richard Nixon, she would be a man.

It is worth noting that not all putative counter-identicals can be paraphrased in the same way. For example,

(F) If I were Gerald Ford, I would be athletic.

can be paraphrased as:

(F*) If I had the non-relational properties presently possessed by Gerald Ford, I would be athletic.

But

(G) If I were Gerald Ford, I would sign the education bill.

is paraphrased differently as:

(G*) If I were in the role of Gerald Ford, I would sign the education bill.

That (F) and (G) really do have different antecedents is indicated by the fact that they do not jointly imply:

(H) If I were Gerald Ford, I would be athletic and sign the education bill.

which they would imply if their antecedents had the same meaning.

Another pair of examples is provided by Goodman (1955):

(I) If New York City were in Georgia, then New York City would be in the South.

(J) If Georgia included New York City, then Georgia would not be entirely in the South.

Initially, (I) and (J) appear to have equivalent antecedents, but they can be paraphrased in such a way that it is evident they do not:

(I*) If New York City were included within the present bounds of Georgia, then New York City would be in the South.

(J*) If Georgia included the area presently occupied by New York City, then Georgia would not be entirely in the South.

Many subjunctive conditionals are ambiguous. The above examples illustrate this. But what is important is whether subjunctive conditionals are subject to a special kind of 'pragmatic' ambiguity arising out of a conventional element in their meaning, or whether the ambiguity in subjunctive conditionals is just the normal kind of ambiguity present in all natural language which arises naturally from our tendency to say less than we mean. The availability of the above paraphrases suggests that the latter is the case. The ambiguity can be resolved simply by being more careful and saying precisely what we mean.

However, it must be admitted that some of the paraphrases employed above are not as clear as we might desire. This is primarily true of those which proceed in terms of the 'non-relational properties' of an object. Just what properties are these? I believe that the above paraphrases are essentially correct, but they do not tell the whole story. In particular, they suggest that the ambiguity in subjunctive conditionals is simply a matter of ambiguity in their antecedents, because in all of the above examples, the ambiguity is resolved in that way. If we allow ourselves to talk without further elucidation about the non-relational properties of an object, then the ambiguity in the above examples can be resolved in this way. But it will be found in Chapter IV that some of these examples are really special cases of a more general phenomenon which I will call 'subject preference'. Subject preference generates a kind of ambiguity in subjunctive conditionals which is often resolved in oral communication through the use of emphasis, e.g., by saying 'If *that* were gold . . .' rather than 'If that *were* gold . . .'. It will be found that not all examples of ambiguity arising from subject preference can be resolved so simply as that in the above examples. In some cases we must regard the conditional itself as genuinely ambiguous. However, this ambiguity is rule-governed and has nothing to do with any supposedly conventional element in the meaning of subjunctive conditionals. Thus it will turn out that there is still no reason to regard subjunctive conditionals as pragmatically ambiguous in the sense of Chisholm and Stalnaker.

Having allowed ourselves the freedom to paraphrase subjunctive conditionals as much as we have, it might be supposed that we can resolve the problem of counterfactuals altogether through the use of paraphrase. It might be supposed that counterfactuals are really just enthymematic conditionals. Or to be more precise, it might be supposed that when we assert a subjunctive conditional like 'If this match had been struck, it would have lit', we do not completely express our meaning – what we write is short for a longer conditional whose antecedent contains as explicit conjuncts those statements in the set $C$ which we use in getting from the supposition 'This match is struck' to the conclusion 'It lights'. This proposal is very much in the spirit of Chisholm's proposal. According to this proposal, the set $C$ consists

merely of statements that were implicitly part of our meaning in formulating the antecedent of the conditional but which we did not bother to state precisely because we assumed that our audience would understand what we had in mind. This implies that there is really no problem regarding what we can put into $C$ and what we cannot put into $C$. Membership in $C$ is simply a matter of what we mean when we assert the conditional.

This account would be very attractive if it worked, but it doesn't work. Although it seems indisputable that in many cases paraphrase is required to make explicit what a conditional means, such paraphrase cannot resolve the proplem of membership in $C$. This can be seen by coming down heavily on the requirement that the paraphrase must really *mean* what the original conditional meant. For example, consider the conditional, 'If this match had been struck, it would have lit'. The set $C$ must contain whatever conditions are necessary for a match to light when struck. Among these are its being dry, there being plenty of oxygen present, etc. The difficulty is that I do not know how to fill out the 'etcetera'. I am confident that there are conditions present which would result in the match's lighting if it were struck, and because of this I am confident that the match would light if struck, but I cannot enumerate those conditions. If I cannot enumerate them, they certainly cannot be part of my meaning when I say, 'If this match were struck it would light'. Consequently, membership in $C$ cannot be simply a matter of what one means when he asserts a conditional.

I have argued that subjunctive conditionals are not subject to the kind of pragmatic ambiguity philosophers have often supposed. Thus there must be precise conditions determining what is included in $C$ and what is not. Membership in $C$ does not consist merely of implicit premises enthymematically omitted from the antecedent in formulating the conditional. Rather, our problem is one of understanding a principle of detachment for subjunctive conditionals. When $\ulcorner(P \,\&\, R. \Rightarrow Q)\urcorner$ expresses a lawlike connection (obtained by instantiation in a law), under what conditions can we detach $R$ from the antecedent to infer the conditional $\ulcorner P > Q\urcorner$? It is not enough to require that $R$ be true, but it is not clear what more should be required.

Goodman (1955) has made a suggestion regarding the solution to

this problem. His observation is that in deciding what would be the case if some proposition $P$ were true, we cannot put just any true statement $R$ into $C$ – at the very least we must require that $R$ would not be false if $P$ were true. So Goodman defines $R$ to be *cotenable* with $P$ just in case $\ulcorner\sim(P > \sim R)\urcorner$ is true, and then his proposal is that $C$ consists of the set of true statements cotenable with $P$.

Goodman has certainly found a necessary condition for a statement to be included in $C$, but it is not a sufficient condition. It is not enough to require that $R$ would not be false if $P$ were true – we must require that $R$ would be true if $P$ were true. Perhaps Goodman was supposing that these are the same thing – that $\ulcorner\sim(P > \sim R)\urcorner$ is logically equivalent to $\ulcorner(P > R)\urcorner$. It is probably true that most philosophers have supposed the negation of a subjunctive conditional to be equivalent to the subjunctive conditional which results from negating the consequent. For example, at first it seems that to deny that if my car had a full tank of gas I could drive all the way to Boston is to affirm that if my car had a full tank of gas I still could not drive all the way to Boston. But I think that this is wrong. It is quite possible for both $\ulcorner(P > \sim R)\urcorner$ and $\ulcorner(P > R)\urcorner$ to be false. For example, suppose there were two kinds of gasoline, long-mileage-gasoline (LMG) and short-mileage-gasoline (SMG), and that a tankful of LMG would get me to Boston but a tankful of SMG would not. Then there are two ways in which the antecedent 'My car has a full tank of gas' could be true, and there need be no way to choose between them and say that if my car had a full tank, it would definitely be full of one of these kinds of gas rather than the other. Under these circumstances we can conclude neither that if my car had a full tank of gas I could drive all the way to Boston, nor that if my car had a full tank I still could not drive all the way to Boston. Rather, we are forced to say that if my car had a full tank I might be able to drive to Boston, and also if my car had a full tank I might not be able to drive to Boston.

This introduces a new kind of subjunctive conditional that I have not mentioned before – the 'might be' conditional. Let us symbolize $\ulcorner$It might be true that $Q$ if it were true that $P\urcorner$ as $\ulcorner QMP\urcorner$. Perhaps it is not obvious how 'might be' conditionals are related to 'would be' conditionals, but I think it is at least clear that $\ulcorner QMP\urcorner$ and $\ulcorner(P > \sim Q)\urcorner$ are incompatible. If $Q$ would be false if $P$ were true, it cannot be that $Q$

might be true if $P$ were true. Thus $\ulcorner QMP \urcorner$ entails $\ulcorner \sim(P > \sim Q) \urcorner$, and similarly $\ulcorner (\sim Q)MP \urcorner$ entails $\ulcorner \sim(P > Q) \urcorner$. Hence if we have both $\ulcorner QMP \urcorner$ and $\ulcorner (\sim Q)MP \urcorner$, it follows that $\ulcorner P > Q \urcorner$ and $\ulcorner P > \sim Q \urcorner$ are both false. And we have seen an example in which we do have both $\ulcorner QMP \urcorner$ and $\ulcorner (\sim Q)MP \urcorner$ true. Therefore there is a difference between requiring that $\ulcorner \sim(P > \sim Q) \urcorner$ be true and requiring that $\ulcorner P > Q \urcorner$ be true.

Given this difference, I think it is clear that Goodman's requirement of cotenability is too weak. This is demonstrated by seeing that it would lead us right back to a special case of the principle that if $\ulcorner \sim(P > \sim Q) \urcorner$ is true then $\ulcorner P > Q \urcorner$ is true. More precisely, Goodman's proposal implies that whenever $P$ is false and $\ulcorner \sim(P > \sim Q) \urcorner$ is true, then $\ulcorner P > Q \urcorner$ is true. This implication is established as follows. First, we need two obvious principles regarding subjunctive conditionals:

(2.1) If $\ulcorner P > Q \urcorner$ is true and $Q$ entails $R$, then $\ulcorner P > R \urcorner$ is true.

(2.2) If $\ulcorner P > (P \supset Q) \urcorner$ is true, then $\ulcorner P > Q \urcorner$ is true.

2.1 is so obvious as to need no defense. 2.2 holds because if $\ulcorner P \supset Q \urcorner$ would be true if $P$ were true, then both $P$ and $\ulcorner P \supset Q \urcorner$ would be true if $P$ were true, and hence $Q$ would have to be true if $P$ were true. Given these principles, let us suppose, with Goodman, that truth and cotenability are all that is required for inclusion in $C$. Suppose $P$ is false, and $\ulcorner \sim(P > \sim Q) \urcorner$ is true. Then by 2.1 $\ulcorner \sim(P > (P \ \& \sim Q)) \urcorner$ is true, and so $\ulcorner \sim(P > \sim(P \supset Q)) \urcorner$ is true. But as $P$ is false, $\ulcorner P \supset Q \urcorner$ is true[1], and it follows from Goodman's proposal that $\ulcorner P > (P \supset Q) \urcorner$ is true. Then from 2.2 it follows that $\ulcorner P > Q \urcorner$ is true. But the example regarding the two kinds of gasoline is a counter-example to this conclusion just as much as it is a counter-example to the more general principle which does not assume that $P$ is false. Thus Goodman's proposal is unacceptable.

What is required for inclusion in $C$ is not merely that $\ulcorner \sim(P > \sim R) \urcorner$ be true, but rather that $R$ would still be true even in $P$ were true, i.e., that $\ulcorner P > R \urcorner$ is true. If we do have $\ulcorner P > R \urcorner$, then it seems clear that we are safe in including $R$ in $C$. This results directly from the intuitive validity of the following principle regarding subjunctive conditionals:

(2.3) $[(P \ \& \ R) > Q] \ \& \ (P > R). \supset (P > Q).$

But it is equally clear that once we have made this our condition for

inclusion in *C*, we are no longer giving an analysis of subjunctive conditionals. This is because we are now *using* subjunctive conditionals to define the class *C*. Thus the linguistic theory runs into a rather severe stumbling block. This is not to say that the linguistic theory cannot be salvaged, but if salvage is to be accomplished, membership in *C* must be characterized in a very different way. In effect, this is the task that will be undertaken in Chapter IV.

There are two fundamental problems for the linguistic theory. The first is that of membership in *C*, which we have been discussing. But even if we could solve this first problem, there would remain another problem. The linguistic theory proposes to characterize subjunctive conditionals partly in terms of physical laws, but the latter concept seems just as problematic as the concept of a subjunctive conditional. Philosophers have sometimes supposed that any true universally quantified material conditional (henceforth: 'material generalization') is a law, but that is rather obviously unsatisfactory. There is a clear intuitive difference between what philosophers have often called accidental and non-accidental generalizations. The material generalization 'All three thousand pound human beings have wings' is true, because there are no three thousand pound human beings. But a generalization that is true simply because it is vacuous will certainly not support subjunctive conditionals. Nor will it suffice merely to rule out those material generalizations that have vacuous antecedents. For example, if there were just one eleven-toed English mathematician whose mother is from Dublin, and he were named 'Charlie', then it would be true that all eleven-toed English mathematicians whose mothers are from Dublin are named 'Charlie', but this generalization would be true 'purely by accident' and no one would want to count it as a law. To be contrasted with this are generalizations like 'All pulsars are neutron stars' or 'All creatures with hearts are creatures with kidneys' which, if true, are in some sense 'nomic' and would be considered laws. The problem is to say what distinguishes laws from accidental generalizations.

I think that a partial answer to this question is that it is a mistake to suppose that laws can be adequately formulated using material conditionals. On the contrary, laws are fundamentally subjunctive in nature. A law tells us not just that all *actual* pulsars are neutron stars, but also

that *any* pulsar there *could* be *would* be a neutron star. Similarly, the law is not just that all actual creatures with hearts are creatures with kidneys, but also that *any* creature with a heart *would be* a creature with a kidney. Let us call generalizations of the form ⌜Any $A$ would be a $B$⌝ *subjunctive generalizations.* Then it seems to be the case that laws must be formulated using subjunctive generalizations. This may suffice to distinguish laws from accidental generalizations, but it makes it all the more difficult to give an adequate analysis of the concept of a law. It now appears that even if we could analyze subjunctive conditionals in terms of laws, we would only have succeeded in analyzing one class of subjunctive statements in terms of another class of subjunctive statements, whereas the basic problem is to analyze subjunctive statements in terms of non-subjunctive statements.

This makes the prospects for the linguistic theory look rather bleak. Fortunately, I do not think they are as bleak as they look. Although laws are subjunctive, it will turn out that they are a much simpler variety of subjunctive statement than are subjunctive conditionals. In Chapter III I will argue that it is possible to give an analysis of subjunctive generalizations in terms of non-subjunctive statements. Once that is accomplished, there will be no circularity in using subjunctive generalizations in analyzing subjunctive conditionals. So the only remaining problem for the linguistic theory will be to characterize membership in the class $C$. I think that the key to the solution to this problem lies in the 'possible worlds' approach to the analysis of subjunctive conditionals, so let us turn now to that approach.

## 3. The 'possible worlds' approach

This approach is due originally to Robert Stalnaker (1968). Stalnaker constructed a formal semantics for subjunctive conditionals based upon the well known formal semantics for modal logic. The idea behind Stalnaker's semantics is that ⌜$(P > Q)$⌝ is true just in case, if we add $P$ to our stock of truths and modify them so as to make them consistent with $P$, but make the modification as small as possible, then the resulting sets of statements entails $Q$. Putting this in terms of possible worlds, we look at that world in which $P$ is true which is most like the

real world, and we see whether $Q$ is true in it. Formally, we suppose we have a two-place function $f$ such that for each sentence $P$ and possible world $\alpha$, $f(\alpha, P)$ is that world containing $P$ which is most like the world $\alpha$. Then $\ulcorner(P > Q)\urcorner$ is true in $\alpha$ iff $Q$ is true in $f(\alpha, P)$. In case $P$ is logically inconsistent, and so true in no possible world, $f(\alpha, P)$ would be undefined, so we pick an arbitrary object $\gamma$, which we designate 'the impossible world', to be the value of $f(\alpha, P)$ in this case. Generating a formal semantics, we can take a *model* to be an ordered quadruple $\langle \alpha, K, f, \gamma \rangle$ where $K$ represents the set of all possible worlds, $\alpha \in K$, and $f$ is the *selection function* which, to each $\beta$ in $K$ and sentence $P$ assigns a member of $K^2$. We will want to put some constraints on the selection function, but subject to those constraints we will say that a sentence involving conditionals is *valid* iff it is true in every model.

Intuitively, not just any function $f$ will constitute a reasonable selection function. It seems clear that the following three requirements should be placed upon $f$:

(3.1) For any $P$ and $\alpha$, $P$ is true in $f(\alpha, P)$.

(3.2) For any $P$ and $\alpha$, $f(\alpha, P) = \gamma$ only if there is no possible world in which $P$ is true.

(3.3) For all $P$ and $\alpha$, if $P$ is true in $\alpha$ then $f(\alpha, P) = \alpha$.

Principle 3.3 reflects the fact that if $P$ is already true, then we do not have to make any changes to the set of truths in order to accommodate $P$. Stalnaker proposes that $f$ is based upon a simple ordering of possible worlds with respect to their resemblance to $\alpha$. This leads him to add the following:

(3.4) For all $P$, $Q$, and $\alpha$, if $P$ is true in $f(\alpha, Q)$ and $Q$ is true in $f(\alpha, P)$, then $f(\alpha, P) = f(\alpha, Q)$.

It might be felt that if one is going to defend a possible worlds analysis of subjunctive conditionals, he should first defend an ontology which contains possible worlds. Stalnaker (1972) and David Lewis (1973, p. 84) both endorse platonistic views of possible worlds. I do not want to say that such a view is wrong, but I do find it somewhat mysterious and I would prefer not to commit myself to it. There are

other alternatives. I prefer to identify a possible world with the set of propositions true in it; or more precisely, to define a possible world to be a maximal consistent set of propositions. Lewis argues that the only way to make sense of 'consistent' here is in terms of possible worlds, but I would hope that he is wrong. I have suggested an alternative account in Pollock (1974), but even if that account is found wanting, I suspect that some account can eventually be given. My point here is not to defend any particular account of possible worlds, but merely to argue that ontological questions about possible worlds can be safely separated from the question how to analyse subjunctive conditionals. The philosopher who approaches the latter question in terms of possible worlds is not automatically subject to criticism for doing so.

The conditions that Stalnaker has placed upon $f$ are clearly not sufficient to pick it out uniquely. There will be many different functions satisfying these conditions. Stalnaker allows for the possibility that there might be additional formal constraints that could plausibly be imposed on $f$, but he denies that we could ever have conditions which would determine $f$ uniquely. He maintains that there are going to be many different equally good selection functions, and that the choice between them is not determined by semantical considerations, but rather by the pragmatics of language. Here Stalnaker is harking back to the traditional view that subjunctive conditionals are subject to some kind of thorough-going ambiguity. It is supposed that the meaning of the conditional cannot resolve the ambiguity; rather, it is to be resolved by pragmatic considerations having to do with the occasion of utterance. Thus Stalnaker calls this kind of ambiguity 'pragmatic ambiguity'.

It is noteworthy that Stalnaker gives no new examples of ambiguity. He bases his view that conditionals are pragmatically ambiguous on the discussions of Chisholm, Goodman, and others which were considered in Section 2. I have urged that those discussions were unconvincing. Thus Stalnaker's position on pragmatic ambiguity is subject to rebuttal on the same grounds. Those grounds do not constitute an argument to the effect that subjunctive conditionals are not pragmatically ambiguous, but rather consist of an attempt to undercut the reasons that have been given for thinking that they are pragmatically ambiguous. Claims of pragmatic ambiguity were based upon an appeal to certain kinds of

examples, and I have argued that those examples are more plausibly interpreted as illustrating an ambiguity in the antecedent of the conditional rather than a kind thoroughgoing ambiguity in the conditional connective itself.

Perhaps the main reason Stalnaker is led to his doctrine of pragmatic ambiguity is that, on his semantics, the following principle is valid: $\ulcorner[(P > Q) \vee (P > \sim Q)]\urcorner$. According to this principle, for any $P$ and $Q$, either $Q$ would be true if $P$ were true, or $Q$ would be false if $P$ were true. But there are many choices of $P$ and $Q$ for which we seem unable to decide which of these alternatives is correct. Stalnaker's answer is that the choice is to be determined by pragmatic considerations. I believe that this is manifestly false. Let $P$ be 'The temperature outside is not 30°' and let $Q$ be 'The temperature outside is 40°'. Is it true that if it weren't 30° out now then it would be 40°, or is it true that if it weren't 30° out now it would not be 40°? I should think that neither of these is true. Nor will it help to bring in pragmatic considerations. Pragmatic considerations are irrelevant in deciding whether it would be 40° out now. Rather than affirm either $\ulcorner(P > Q)\urcorner$ or $\ulcorner(P > \sim Q)\urcorner$, we should deny both and say that if it weren't 30° out now then it *might* be 40° out, but also it might be something else.

Stalnaker's analysis ignores 'might be's'. The principle $\ulcorner[(P > Q) \vee (P > \sim Q)]\urcorner$ should not be valid. I argued in Section 1 that $\ulcorner Q$ might be true is $P$ were true$\urcorner$ is incompatible with $\ulcorner(P > \sim Q)\urcorner$, and there are cases like the above in which we want to say that $Q$ might be true if $P$ were true, and also that $Q$ might be false if $P$ were true. Rather than an appeal to pragmatic considerations to resolve those cases where we cannot decide which of $\ulcorner(P > Q)\urcorner$ and $\ulcorner(P > \sim Q)\urcorner$ should be true, we should deny that one of them always has to be true. What is required is a modification of the semantics so that $\ulcorner[(P > Q) \vee (P > \sim Q)]\urcorner$ is no longer valid. As we will see, this is precisely what David Lewis has accomplished.

There are further difficulties for Stalnaker's semantics. Stalnaker suggests that his semantical rules (possibly augmented *slightly*) exhaust the semantical features of subjunctive conditionals, and that the choice of a selection function within these constraints is a matter of pragmatics. But most such selection functions would be totally preposterous. By judicious choice of a selection function, we could make virtually

any conditional either true or false as we wish. For example, we could choose a function which selects the closest world in which I drop a piece of chalk (which in fact I did not drop) to be one in which Richard Nixon jumps over the moon. Then it would be true that if I had dropped that piece of chalk, Richard Nixon would have jumped over the moon. But that conditional just isn't true. In general, given any sentences $P$ and $Q$, if $P$ is false and does not logically entail $\ulcorner\sim Q\urcorner$, there will be a selection function satisfying Stalnaker's formal constraints which will make $\ulcorner(P > Q)\urcorner$ true. But this is absurd. It might possibly be true that there is *some* ambiguity built into subjunctive conditionals, but no one can believe that they are subject to this kind of rampant ambiguity which would result from Stalnaker's position.

Perhaps part of the difficulty is that in talking about pragmatics, Stalnaker's words (but not Stalnaker himself) suggest that this is just a matter of "pragmatic considerations" – appeals to practicality or some such thing. But the examples Stalnaker gives of the pragmatics of language are not at all like that. Pragmatics is supposed to govern such things as the reference of indexical expressions. Pragmatics, so construed, has nothing to do with practicality. It functions according to definite rules of language having to do with, for example, pointing to fix the referent of 'he'. There is nothing hit or miss about this, and nothing 'to be decided' by considerations of practicality. It would not be unreasonable to include pragmatics as part of semantics. Whether we do that or not would seem to be nothing more than a matter of terminology. But however we choose to classify pragmatic rules, there is no reason to think that there are no precise linguistic rules which uniquely determine the selection function. The fact that most potential selection functions are obviously ruled out by our understanding of the conditional (which presumably means they are ruled out by the rules of language governing conditionals) suggests that all but one may be ruled out, and it certainly indicates that there are many more constraints on the selection function than indicated by Stalnaker.[3]

Regardless of whether there is just one selection function or many selection functions, Stalnaker's semantics requires revision to rid it of the offending theorem $\ulcorner[(P > Q) \vee (P > \sim Q)]\urcorner$. Such revision has been provided by Davis Lewis, (1972, 1973, 1973a). Lewis begins by proposing that the selection function be understood in terms of our

ordinary notion of comparative similarity. What is at issue is whether a world $\alpha$ is more similar to a given world than is a world $\beta$. Lewis writes that he means precisely the same relation of comparative similarity as when we judge the comparative similarity of cities or people or philosophies. Thus not just any old selection function will do. A proper selection function has to be built out of this notion which we take as basic for the sake of the analysis. This is certainly a step in the right direction, although as I will urge, it will not quite do.

Starting with this notion of comparative similarity, Lewis observes that Stalnaker's semantics assumes there will always be just one world making $P$ true which will be more like a given world than will any other world in which $P$ is true. But this is implausible. Why couldn't there be ties? This seems to be just what happens when we judge that if Bizet and Verdi were compatriots they might both be French, and they might both be Italian. A world in which they are both French is just as much like the real world as one in which they are both Italian. Thus, rather than having $f(\alpha, P)$ be a single world, it should be a whole set of worlds. This leads Lewis to the analysis he numbers 'Analysis 2' according to which, if $f(\alpha, P)$ is the set of all worlds in which $P$ is true which are at least as similar to $\alpha$ as are any other worlds in which $P$ is true, then $\ulcorner(P > Q)\urcorner$ is true iff $Q$ is true in *every* world in $f(\alpha, P)$. This analysis embodies the 'limit assumption' according to which there are worlds in which $P$ is true which are *maximally* similar to $\alpha$. This rules out the possibility that worlds might get indefinitely closer to $\alpha$ without limit.

Lewis rejects the limit assumption on the basis of examples like the following. He draws a line slightly less than an inch long, and then he says

> Suppose we entertain the counterfactual supposition that at this point there appears a line more than an inch long.... There are worlds with a line 2″ long; worlds presumably closer to ours with a line $1\frac{1}{2}$″ long; worlds presumably still closer to ours with a line $1\frac{1}{4}$″ long; worlds presumably still closer.... But how long is the line in the *closest* worlds with a line more than an inch long? If it is $1+x$″ for any $x$ however small, why are there not other worlds still closer to ours in which it is $1+\frac{1}{2}x$″, a length still closer to its actual length? The shorter we make the line (above 1″), the closer we come to the actual length; so the closer we come, presumably, to our actual world. Just as there is no shortest possible length above 1″, so there is no closest world to ours among the worlds with lines more than an inch long (Lewis, 1973, pp. 20–21).

This leads Lewis to propose the following analysis:

$\ulcorner(P > Q)\urcorner$ is true in $\alpha$ iff there is some world in which $\ulcorner(P \,\&\, Q)\urcorner$ is true which is more like $\alpha$ than is any world in which $\ulcorner(P \,\&\, \sim Q)\urcorner$ is true.

Lewis' argument seems quite persuasive, but it leads to a most peculiar result. On his diagnosis, for each $x$ greater than zero there are worlds in which the length of the line is between 1″ and $1 + x$″ which are closer to the actual world than are those worlds in which it is $1 + x$″. It follows that the conditional 'If the line were more than an inch long, it would not be $1 + x$ inches long' is true for each $x$ greater than zero. But if the line would not be $1 + x$″ long for any $x$, it follows that the line would not be more than an inch long (if it is more than an inch long, then it is more by some non-zero amount $x$). That is, if the line were more than an inch long, then it would not be more than an inch long. But that is certainly false. What has happened?

First it should be pointed out that the above informal argument employs a principle which is not valid on Lewis' semantics. Unfortunately, this does not help Lewis because the principle ought to be valid. Let us define, where $\Gamma$ is a set of sentences and $P$ and $Q$ are sentences:

(3.5) (a) $\Gamma \rightarrow P$ iff every possible world making all of the sentences in $\Gamma$ true makes $P$ true.

(b) $Q \rightarrow P$ iff $\{Q\} \rightarrow P$.

'$\rightarrow$' is the entailment relation. Entailment by a single sentence may or may not be represented by an operator in the object language, but of course the general concept of entailment by a set of sentences cannot be represented in that way. The following *Consequence Principle* is valid on Lewis' semantics:

(3.6) If $\ulcorner(P > Q)\urcorner$ is true and $Q \rightarrow R$, then $\ulcorner(P > R)\urcorner$ is true.

It seems clear intuitively that this principle ought to hold, so it is gratifying that it does. Because adjunctivity also holds (that is, if $\ulcorner(P > Q)\urcorner$ and $\ulcorner(P > R)\urcorner$ are true, so is $\ulcorner P > (Q \,\&\, R)\urcorner$), we can generalize the consequence principle to obtain the *Finite Consequence Principle*:

(3.7) If $\Gamma$ is a finite set of sentences, and for each $Q \in \Gamma$, $\ulcorner(P > Q)\urcorner$ is true, and $\Gamma \rightarrow R$, then $\ulcorner(P > R)\urcorner$ is true.

On intuitive grounds, it seems that the following *Generalized Consequence Principle* should also hold:

(3.8) If $\Gamma$ is a set of sentences, and for each $Q \in \Gamma$, $\ulcorner(P > Q)\urcorner$ is true, and $\Gamma \rightarrow R$, then $\ulcorner(P > R)\urcorner$ is true.

The generalized consequence principle seems just as obvious as either the original consequence principle or the finite consequence principle, and indeed should be true for exactly the same reason. It is surprising then that on Lewis' semantics this principle does not hold. The example about the length of the line can be turned into a proof of this fact.

As the generalized consequence principle ought to be valid, what happens if we impose on the similarity relation the condition that it is valid? Surprisingly enough, this is equivalent to the limit assumption.[4] Thus we are led back to Analysis 2. This might seem surprising. What are these closest worlds? They are the worlds that *might* be the actual world if $P$ were true. We can make this precise as follows. Where $\alpha$ is a world, let us define

(3.9) $\alpha \mathbf{M} P$ iff $\sim(\exists Q)[\ulcorner P > \sim Q\urcorner$ is true & $Q$ is true in $\alpha]$.

We then prove two theorems:

(3.10) If 3.8 holds and $\ulcorner\sim(P > \sim R)\urcorner$ is true, then $(\exists\alpha)(\alpha \mathbf{M} P$ & $R$ is true in $\alpha)$.

*Proof:* Let $\Gamma = \{R\} \cup \{S; \ulcorner(P > S)\urcorner$ is true$\}$. Suppose there is no possible world $\alpha$ in which all the sentences in $\Gamma$ are true. Then $\Gamma$ is inconsistent, i.e., $\{S; \ulcorner(P > S)\urcorner$ is true$\} \rightarrow \ulcorner\sim R\urcorner$. Then by 3.8, $\ulcorner(P > \sim R)\urcorner$ is true, contrary to supposition. Therefore there is a world $\alpha$ in which all the sentences in $\Gamma$ are true. For any sentence $Q$, if $\ulcorner(P > \sim Q)\urcorner$ is true then $\ulcorner\sim Q\urcorner \in \Gamma$, so $\ulcorner\sim Q\urcorner$ is true in $\alpha$, and hence $Q$ is not true in $\alpha$. Thus $\alpha \mathbf{M} P$. And as $R \in \Gamma$, $R$ is true in $\alpha$.

(3.11) If 3.8 holds, then $\ulcorner(P > Q)\urcorner$ is true iff $(\forall\alpha)[\alpha \mathbf{M} P \supset Q$ is true in $\alpha]$.

*Proof:* Suppose $\ulcorner(P > Q)\urcorner$ is true. Suppose $\alpha \mathbf{M} P$. Then $\ulcorner(P > \not\sim Q)\urcorner$ is true (by 3.8), so $\ulcorner\sim Q\urcorner$ is not true in $\alpha$, and hence $Q$ is true in $\alpha$. Conversely, suppose $\ulcorner(P > Q)\urcorner$ is not true. By theorem 3.10, $(\exists\alpha)(\alpha \mathbf{M} P$ & $Q$ is not true in $\alpha)$. Hence $\sim(\forall\alpha)(\alpha \mathbf{M} P \supset Q$ is true in $\alpha)$.

Theorem 3.11 tells that the closest worlds in which $P$ is true are those worlds that might be actual if $P$ were true.

Thus the generalized consequence principle leads us inexorably back to Analysis 2. What went wrong with Lewis' argument which led him to reject Analysis 2 and adopt his more complicated analysis instead? I think that the error lies in supposing that the relation we use in selecting possible worlds is our ordinary relation of comparative similarity. Lewis' own example about the line illustrates this nicely. He is surely right in supposing that a world in which the line is $1\frac{1}{2}''$ long is more like the real world than one in which the line is $2''$ long. If comparative similarity were the relation that is operative in generating counterfactuals, it would follow that if the line were more than an inch long then it would not be two inches long. But that is false. On the contrary, if the line were more than an inch long then it might be two inches long. It might also be three inches long, or four inches long, etc. None of these alternatives is ruled out, although they clearly do generate worlds that differ in their similarity to the real world.

Apparently comparative similarity is not the relation that is involved in determining which worlds we must look at in deciding whether $\ulcorner(P > Q)\urcorner$ is true. This can be seen more directly by considering what the operative relation is. What is involved in counterfactuals is the notion of a minimal change being made to the actual world in order to accommodate the counterfactual supposition. We must change the world in some way in order to make the antecedent true, but we are constrained to make the change as small as possible. The operative notion here is that of the minimality of change. Two different minimal changes can result in worlds that differ in their similarity to the real world. For example, a change yielding a world in which the line is $2''$ long is no greater than a change yielding a world in which the line is $1\frac{1}{2}''$ long. Each change is minimal, because each amounts merely to changing the length of the line to accommodate the counterfactual supposition and then making whatever additional changes are required for the sake of consistency. But the resulting worlds differ in their degree of similarity to the real world.

Apparently the notion of the magnitude of a change is not the same as that of the comparative similarity of worlds. Once this is recognized, we can build our semantics on the former notion rather than the latter.

Does this make any difference to the formal structure of the semantics? Lewis made the reasonable assumption that comparative similarity relative to any particular world constituted a simple ordering.[5] Unfortunately, if we consider the ordering of worlds in terms of the magnitude of change required to generate them from the real world, it seems that we do not have a simple ordering but only a partial ordering. That is, the ordering is not connected. Sometimes it makes perfectly good sense to say that one change is greater than another, i.e., when the one change contains the other. But it is not true in general that we can compare changes in worlds and say either that they are the same magnitude or that one is greater than the other. This can be illustrated as follows. Let $T$, $R$, and $S$ be three unrelated false sentences, e.g., the sentences 'My car is painted black', 'My maple tree died', 'My garbage can blew over'. Let $P$ be $\ulcorner T \vee (R \ \& \ S)\urcorner$, and let $Q$ be $\ulcorner T \vee R\urcorner$. Now consider what happens when we entertain $P$ and $Q$ as counterfactual suppositions. In general, if we have a disjunction whose disjuncts are unrelated, then if the disjunction were true, *either* disjunct might be true – neither disjunct would have to be false. In the case of $P$ and $Q$ we have (1) that $\ulcorner \sim (P > \sim T)\urcorner$ and $\ulcorner \sim (P > \sim (R \ \& \ S))\urcorner$ are true, and (2) that $\ulcorner \sim (Q > \sim T)\urcorner$ and $\ulcorner \sim (Q > \sim R)\urcorner$ are true. By (1) and Theorem 3.11, there are worlds $\alpha$ and $\beta$ such that $\alpha \mathbf{M} P$ and $\beta \mathbf{M} P$ and $T$ is true in $\alpha$ and $\ulcorner (R \ \& \ S)\urcorner$ is true in $\beta$. $\alpha$ results from making minimal changes to the real world so as to make $T$ true, and $\beta$ results from making minimal changes so as to make $\ulcorner (R \ \& \ S)\urcorner$ true. But then we will also have $\alpha \mathbf{M} Q$, because making $T$ true is also a way of making $Q$ true by making minimal changes. If connectedness holds, so that it always makes sense to compare the magnitudes of two changes, as we have both $\alpha \mathbf{M} P$ and $\beta \mathbf{M} P$, $\alpha$ and $\beta$ must result from changes of the same magnitude. But then as $\alpha \mathbf{M} Q$, the magnitude of change involved in $\alpha$ must be the minimal change possible in making $Q$ true. But (on the assumption of connectedness), $\beta$ involves the same amount of change, and $Q$ is true in $\beta$ (because $R$ is true in $\beta$), so we must also have $\beta \mathbf{M} Q$. But we *should not* have $\beta \mathbf{M} Q$. In constructing $\beta$ we have changed the real world more than necessary to make $Q$ true, because we have made *both* $R$ and $S$ true when we only needed to make $R$ true. Although $\alpha$ and $\beta$ both involved minimal changes relative to the counterfactual supposition $P$, only $\alpha$ and not $\beta$ involves a minimal

change relative to the counterfactual supposition $Q$. Thus connectedness fails. It does not make sense to ask whether $\alpha$ and $\beta$ involve 'the same amount of change'. It only makes sense to ask whether they involve minimal changes necessary to make a certain counterfactual supposition true.

The picture that emerges from this is that although it makes perfectly good sense to say that one change is larger than another when the first contains the second as part of it, it does not in general make sense to try to compare the magnitudes of changes. Changes form a lattice rather than a simple ordering. To say that a change is minimal relative to making a certain antecedent true is to say that we cannot make the antecedent true by making just part of that change. Two changes may each be minimal relative to making a certain antecedent true, but only one of them minimal with respect to making another antecedent true.

Thus we cannot assume that the ordering of worlds which is generated by our selection of minimally different worlds is a simple ordering. It is only a partial ordering. As I will argue in Chapter II, Lewis' assumption to the contrary leads him to embrace as valid certain theorems which should not be valid.

## 4. Conclusions

I have argued that there are some serious problems for the possible worlds approach as it has been developed to date. But, by Theorem 3.11, it follows that the basic insights are correct. As long as we agree that the Generalized Consequence Principle holds, it follows that the possible worlds approach can be made to work. But one may wonder whether the possible worlds approach really involves any advance over the linguistic theory. The possible worlds theory proceeds in terms of the notion of the minimal change to the real world necessary to make a certain sentence true. No one would hold this notion up as a model of clarity. Before we can regard the possible worlds theory as giving us anything that deserves to be called an 'analysis' of counterfactual conditionals, we must have an analysis of this notion of a minimal change. And upon reflection, the latter task seems to involve all of the

problems that arose for the linguistic theory. First, it seems that laws must play an important role in the notion of a minimal change. Most people's intuitions seem to agree that in making a counterfactual supposition true, we change laws only as a last resort. For purposes of counterfactuals, laws are more immutable than particular facts.[6] Thus a clear account of the notion of a minimal change must presuppose an account of what a law is. Second, the question which particular facts can be changed and which cannot in a minimal change seems to be the same as the question what truths can be put into the set $C$ in the linguistic theory. At this point, it may look as if the possible worlds approach has accomplished nothing. The same problems have arisen all over again in different guise. But this is misleading. As we will see in Chapter IV, the possible worlds approach provides a helpful new perspective on the problem which, in the end, facilitates its solution.

However, before we can attempt to provide that solution, we must return to some of the same old problems. In Chapter II we will undertake to sort out various logical facts about subjunctive conditionals which have obscured the solution to the general problem of providing an adequate analysis. In Chapter III we will examine the notion of a law, and we will attempt to provide an analysis of it. Then in Chapter IV, having laid the groundwork, we will return to the task of analyzing the notion of a minimal change and will provide what I hope is an adequate analysis, and thereby provide an analysis for counterfactual conditionals.

## NOTES

[1] If it is agreed that $\ulcorner P \,\&\, Q \urcorner$ entails $\ulcorner P > Q \urcorner$, a principle which will be defended later, then $\ulcorner \sim(P > \sim Q) \urcorner$ entails $\ulcorner P \supset Q \urcorner$ all by itself and we do not need the assumption that $P$ is false.

[2] Stalnaker also included an alternativeness relation, but I omit it for the sake of simplicity. It has no direct effect on the logic of conditionals.

[3] Of course, these constraints are most likely not 'formal' in the same sense as are those which Stalnaker lists.

[4] More precisely, it is equivalent *given* the assumption that comparative similarity relative to any world is at least a partial ordering.

[5] More precisely, he assumed a simple ordering of the equivalence classes of worlds equally similar to the given world.

[6] David Lewis denies this, but I think that his denial stems from his supposing that comparative similarity is the notion operative in generating counterfactuals.

CHAPTER II

# FOUR KINDS OF CONDITIONALS

## 1. Introduction

There are certain logical problems that must be discussed prior to undertaking the analysis of subjunctive conditionals. The discussion of these problems will clarify the nature of these conditionals and thereby facilitate their analysis. The problem of analyzing subjunctive conditionals has been exacerbated by the fact that there are actually several different kinds of subjunctive conditionals, with different properties and correspondingly different analyses, and philosophers have not been sufficiently careful in distinguishing between them. In this chapter I will distinguish between four main kinds of subjunctive conditionals and explore their interrelations. It will turn out that these conditionals are all interdefinable, so an analysis of one will yield analyses of all the others.

## 2. The four kinds

Let us say that a *simple subjunctive* is a conditional of the form ⌜If it were true that $P$ then it would be true that $Q$⌝. We will symbolize this as ⌜$P > Q$⌝. It is the simple subjunctive that has been discussed most frequently in the literature and which was under discussion in Chapter I. There is a more or less traditional assumption about subjunctive conditionals that has been uniformly rejected in the recent literature. This is the assumption that a subjunctive conditional asserts the existence of a *connection* between the antecedent and consequent. Certainly, some simple subjunctives are true because such a connection exists, but this is not invariably the case. The existence of such a connection is a sufficient, but not a necessary, condition for the truth of a simple subjunctive. This is easily seen by considering examples.

First, there are obvious examples of a simple subjunctive being true

because of the existence of such a connection. We say, 'If this match were struck, it would light', because we believe that, in some sense, striking the match would bring about its lighting. Similarly, we say 'If the bird you saw had been a raven, it would have been black', because a bird's being a raven in some sense requires (without logically entailing) its being black. In each of these examples, the truth of the antecedent in some sense *makes* the consequent true, or *requires* the consequent to be true. Let us say that in these cases the antecedent *necessitates* the consequent. Notice that necessitation is not always *causal* necessitation. We cannot say that the bird's being a raven causes it to be black. It might conceivably be true that all (contingent) necessitations are ultimately explicable in terms of causes, but it is clear that there are cases in which we cannot say simply that the truth of the antecedent causes the truth of the consequent.

No one is inclined to doubt that simple subjunctives are often true because the antecedent necessitates the consequent. But it has not always been recognized that a simple subjunctive can also be true when there is no such necessitation. For example, we might say of a witch doctor, 'It would not rain if he did not do a rain dance, but it would not rain if he did either.' This conjunction of two simple subjunctives expresses the *lack* of a connection rather than the presence of one.

Contrary to the traditional assumption, it seems clear that simple subjunctives do not express a relation of necessitation between their antecedent and consequent. Rather, the presence of such a connection is just one ground for asserting a simple subjunctive. It seems that there are basically two ways that a simple subjunctive can be true. On the one hand, there can be a connection between the antecedent and consequent so that the truth of the antecedent would bring it about, i.e., necessitate, that the consequent would be true. On the other hand, a simple subjunctive can be true because the consequent is already true and there is no connection between the antecedent and consequent such that the antecedent's being true would interfere with the consequent's being true. The latter case is typically expressed by 'even if' subjunctives: 'Even if the witch doctor were to do a rain dance, it would not rain'.

There is a special kind of English conditional – the 'even if' subjunctive – whose function is to express the second way in which the simple subjunctive can be true. It would be convenient if there were also a special subjunctive conditional whose function is to express the first way in which the simple subjunctive can be true – i.e., conditionals which express necessitation. If there weren't such conditionals, it would be worthwhile to introduce them. However, I believe that there are several locutions in English which can be used to express these conditionals. The most straightforward way of expressing necessitation is with the locution, ⌜If it were true that *P*, then it would be true that *Q since* it was true that *P*⌝. For example, we might say of the match, 'If it were struck, it would light since it was struck', but deny of the witch doctor, 'If he did a rain dance, it would fail to rain since he did the dance'. In using this locution, one must not fall into the common error of supposing that 'since' statements are always causal. For example, the bird is black since it is a raven. We might say that these statements are always 'necessitory', but there are kinds of necessitation which are at least not directly causal.

Another locution which can often be used to express necessitation consists of using 'couldn't' in place of 'wouldn't': ⌜If it were true that *P*, it couldn't be false that *Q*⌝. Thus, for example, the simple subjunctive 'If the witch doctor were to do a rain dance, it wouldn't rain' is true, but 'If the witch doctor were to do a rain dance, it couldn't rain' is false. The latter is false because it expresses necessitation, and there is no necessitation in this case. On the other hand, both 'If this match were struck, it wouldn't fail to light' and 'If this match were struck, it couldn't fail to light' are true, because here the antecedent does necessitate the consequent. Let us symbolize these necessitation conditionals as ⌜$P \gg Q$⌝.

It is not invariably the case that the English locution ⌜If *P* were true, *Q* couldn't be false⌝ expresses necessitation. For example, suppose we have a match which has been soaked in water. We could reasonably say of such a match, 'If this match were struck, it couldn't light', but certainly its being struck would not necessitate its not lighting. What is happening here is that the modal statement 'This match couldn't light' is true, and the above conditional really has the force of an 'even if'

conditional: 'Even if this match were struck, it couldn't light'. The 'couldn't' attaches to the consequent rather than, as in necessitation statements, expressing a relation between antecedent and consequent.[1]

To further confuse matters, we sometimes use the English locution ⌜If it were true that $P$ then it would be true that $Q$⌝ to express necessitation rather than the simple subjunctive. For example, suppose we have a broken-down old car whose engine is beyond repair. We might reasonably assert of this car (and equally of any car), 'Its engine would not run if it had a broken piston, but it is not the case that its engine would not run if the car had a flat tire'. But, shifting gears mentally, we may also affirm, 'The engine would not run if the car had a flat tire', because we know that the engine will not run no matter what, and hence would not run even if the car had a flat tire. We seem to be contradicting ourselves here. We are both affirming and denying 'The engine would not run if the car had a flat tire'. But we are not really contradicting ourselves. What we are saying is that (1) the engine would not run even if the car had a flat tire, but (2) the engine's not running would not be necessitated by the car's having a flat tire. Apparently the English words 'If it were true that $P$, then it would be true that $Q$' are used ambiguously, most often to express simple subjunctives, but sometimes to express necessitation.

These ambiguities in the English locutions used to express subjunctive conditionals have, I suspect, contributed significantly to the difficulties in analyzing subjunctive conditionals. They mean, in particular, that we cannot introduce '$>$' and '$\gg$' as simple paraphrases of the English locutions If it were true that $P$ then it would be true that $Q$⌝ and ⌜If it were true that $P$, it could not be false that $Q$⌝. This is no particular hardship, because the concepts themselves seem to make good intuitive sense. We have no difficulty recognizing that the English words do on occasion mean different things.

We have distinguished between simple subjunctives, 'even if' conditionals and necessitation conditionals. These are three of the four kinds of conditionals referred to in the title of the chapter. The remaining kind of conditional is the 'might be' conditional: ⌜If it were true that $P$, it might be true that $Q$⌝. In the following sections each of these conditionals will be discussed in turn, and the attempt will be made to make each clearer and explore its relation to the others.

## 3. 'EVEN IF' SUBJUNCTIVES

Let us begin with 'even if' subjunctive conditionals. Let us symbolize ⌜$Q$ even if $P$⌝ as ⌜$QEP$⌝. What does it mean to say that $Q$ would be true even if $P$ were true? At the very least, this implies that $Q$ is true now. One might think it also implies that $P$ is false, because we would not ordinarily say '$Q$ would still be true even if $P$ were true' if we knew that $P$ *is* true. But this is one of Grice's conversational implicatures rather than a logical implication. We frequently have occasion to judge that ⌜$QEP$⌝ is true in cases in which we do not know whether $P$ is true. If we later discover that $P$ was true, this does not show that we were wrong in thinking that $Q$ would still be true even if $P$ were true. On the contrary, this would seem to show that we were right. Consequently, ⌜$QEP$⌝ does not entail ⌜$\sim P$⌝.

⌜$QEP$⌝ entails $Q$, but obviously it entails much more besides. To get at what else it entails, consider some examples:

(1) My car would still be white even if the maple tree in my front yard died.

(2) Match $m$ would still be dry even if it were struck.

(3) $\sim$Match $m$ would not light even if it were struck.

(4) $\sim$This bird would still be black even if it were not a raven.

(5) $\sim$The Japanese current would still run alongside Japan even if Japan were only fifty miles from Alaska.

The reason my car would still be white even if my maple tree were to die is that the death of the tree does not enter into anything which would necessitate my car's not being white. Analogously, match $m$ would still be dry even if it were struck, because striking it does not enter into anything which would necessitate its not being dry. On the other hand, striking match $m$ does enter into something which would necessitate its not being the case that it does not light, so it is false that match $m$ would not light even if struck. These examples suggest, as a first approximation, that ⌜$QEP$⌝ is equivalent to

$$Q \ \& \ \sim(\exists R)[(P \ \& \ R) \gg \sim Q].$$

But this is surely too strong. It is always possible to find *some* $R$ which,

together with $P$, would necessitate $\ulcorner\sim Q\urcorner$. For example, $R$ could be $\ulcorner(P \supset \sim Q)\urcorner$. A natural suggestion to remedy this would be to require that $R$ is something that would be true if $P$ were true:

$$Q \;\&\; \sim(\exists R)[(P \;\&\; R. \gg \sim Q) \;\&\; (P > R)].$$

But this does not work for cases (4)–(6). If I point to a raven and say 'This bird would still be black even if it were not a raven', I am wrong. If it were not a raven, it might be a cardinal, or a bluejay, or all sorts of non-black birds. But there is nothing that would be true of the bird which, together with its not being a raven, would necessitate its not being black. On the contrary, it *would* be black if, for example, it were a crow. The reason it is false that the bird would still be black even if it were not a raven is that there are all sorts of things that *might* be true of it (e.g., being a cardinal) which, together with its not being a raven, would necessitate its being non-black. Analogously, it need not be the case that the Japanese current would still run alongside Japan even if Japan were only fifty miles from Alaska, because there are different ways in which Japan might be only fifty miles from Alaska. If this resulted from the Pacific Ocean's being only fifty miles wide, then perhaps the Japanese Current would still run alongside Japan (this would depend upon general oceanographic laws which are probably unknown at this time). But if Japan were only fifty miles from Alaska but a thousand miles from Asia, then the Japanese current would not run alongside Japan. In both of these cases, we infer $\ulcorner\sim(QEP)\urcorner$ because $\ulcorner(\exists R)[(P \;\&\; R. \gg \sim Q) \;\&\; (R$ might be true if $P$ were true)$]\urcorner$ is true. These examples support the following equivalence, which I think is correct:

(3.1) $\ulcorner QEP\urcorner$ is equivalent to $\ulcorner Q \;\&\; \sim(\exists R)[(P \;\&\; R. \gg \sim Q) \;\&\; (R$ might be true if $P$ were true)$]\urcorner$.

There are certain example which appear to be counter-examples to this analysis. For example, we might say of a person 'He would be fired even if he drank just a little' without meaning to imply that he will be fired. This appears to be an example in which $\ulcorner QEP\urcorner$ does not entail $Q$.[2] However, I do not think that this is a real counter-example. 'He would be fired even if he drank just a little' is a shortened form of 'If he drank, he would be fired even if he drank just a little'. The latter

can be symbolized as $\ulcorner D > (FEL) \urcorner$. Of course, one might reply that this doesn't make any difference. Whether 'He would be fired even if he drank just a little' is short for something else or not, it is not in accord with our analysis. If one wishes to take that line, then the reply is simply that we are not attempting to analyze all possible uses of 'even if'. We are merely analyzing what is in some sense 'the standard use' of 'even if'.

Now let us consider the logical properties of 'even if'. It seems clear that all of the following principles ought to hold:

(3.2) $QEP \ \& \ (Q \rightarrow R). \supset REP.$

(3.3) $QEP \ \& \ REP. \supset (Q \ \& \ R)EP.$

(3.4) $QEP \ \& \ RE(P \ \& \ Q). \supset REP.$

(3.5) $PEQ \ \& \ PER. \supset PE(Q \vee R).$

(3.6) $(P \ \& \ Q) \supset QEP.$

However, these cannot be proven until we know something about the logical properties of 'might be'.

## 4. 'Might be' conditionals

We were forced to use conditionals of the form $\ulcorner Q$ might be true if $P$ were true$\urcorner$ in analyzing $\ulcorner QEP \urcorner$, so let us turn to their analysis next. We can symbolize them as $\ulcorner QMP \urcorner$. To say that $Q$ might be true if $P$ were true seems to be the same as saying that it is not the case that $Q$ would definitely be false if $P$ were true, which suggests the following analysis:

(4.1) $(QMP) \equiv \sim(P > \sim Q).$

This would be unexceptionable were it not that many philosophers have thought that from the falsity of $\ulcorner Q$ would be true if $P$ were true$\urcorner$ it follows that $Q$ would be false if $P$ were true:

(4.2) $\sim(P > Q) \supset (P > \sim Q).$

Certainly $\ulcorner QMP \urcorner$ does not entail $\ulcorner (P > Q) \urcorner$, so either 4.1 fails, or 4.2 does not hold. Which is the culprit?

I think it is fairly easy to see that 4.2 is false. If 4.2 were true, then

$\ulcorner(P > Q) \vee (P > \sim Q)\urcorner$ would be a truth of logic. Clearly, $\ulcorner QMP\urcorner$ is incompatible with $\ulcorner(P > \sim Q)\urcorner$. If $Q$ would be false if $P$ were true, then it is not the case that $Q$ might be true if $P$ were true:

(4.3) $\quad (P > \sim Q) \supset \sim(QMP).$

Consequently:

(4.4) $\quad [(P > Q) \vee (P > \sim Q)] \supset \sim[QMP \ \& \ (\sim Q)MP].$

Thus if 4.2 were true, it would be impossible to have both $\ulcorner QMP\urcorner$ and $\ulcorner(\sim Q)MP\urcorner$ true. But as we have seen, this is not impossible. For example, it is true both that if the temperature outside now were not 30° then it might be 40°, and that it might be something other than 40°. Thus 4.2 cannot be true.

It is not difficult to see why 4.2 fails. What $\ulcorner P > Q\urcorner$ says is, roughly, that whatever *might* be the case if $P$ were true, $Q$ would be true.[3] Insofar as there are different things that might be the case if $P$ were true, it can happen that neither $Q$ nor $\ulcorner\sim Q\urcorner$ would be true in *all* the circumstances that might occur if $P$ were true.

Should we conclude then that 4.1 is true? 4.1 seems right except possibly for one case. It seems that the following entailment should hold:

(4.5) $\quad (P > Q) \supset QMP.$

This will result from 4.1 except in the case where we have both $\ulcorner P > Q\urcorner$ and $\ulcorner P > \sim Q\urcorner$ true. It will be seen later that this can happen only when $P$ is logically impossible. My intuitions fail me in this case. We could ensure the truth of 4.5 by analyzing $\ulcorner QMP\urcorner$ as:

(4.6) $\quad (QMP) \equiv [(P > Q) \vee \sim(P > \sim Q)]$

rather than as in 4.1, but I see no clear way to decide which analysis is correct. I suspect that our use of the English phrase 'might be' is simply undetermined in the case where both $\ulcorner P > Q\urcorner$ and $\ulcorner P > \sim Q\urcorner$ are true, in which case it makes no difference which way we analyze $\ulcorner QMP\urcorner$. Furthermore, it will result that it makes no difference to the analysis of 'even if' which analysis of 'might be' is adopted. Consequently, I propose to adopt the simpler analysis and take 4.1 as defining $\ulcorner QMP\urcorner$.

## 5. NECESSITATION CONDITIONALS

Now let us turn to conditionals expressing necessitation. How can we analyze $\ulcorner P \gg Q \urcorner$? Clearly, this entails $\ulcorner P > Q \urcorner$, but equally clearly, this is not a sufficient condition for the truth of $\ulcorner P \gg Q \urcorner$. This is due to the fact that $\ulcorner P > Q \urcorner$ may be true just because $Q$ is already true and $P$'s being true is irrelevant to the truth of $Q$ and hence would not interfere with the truth of $Q$. This suggests that what is required for the truth of $\ulcorner P \gg Q \urcorner$ is not just that $\ulcorner P > Q \urcorner$ is true, but that $\ulcorner P > Q \urcorner$ would still be true even if $Q$ weren't already true:

$$(P \gg Q) \equiv [(P > Q) E \sim Q].$$

However, this is too strong a requirement. The difficulty arises in the case in which $P$ and $Q$ are both true. Characteristically, $P$ only necessitates $Q$ because certain other propositions are true and satisfy whatever other conditions are required for the validity of detachment in necessitation conditionals. For example, consider a match which lit ($L$) since it was struck ($S$). So $\ulcorner S \gg L \urcorner$, $S$, and $L$ are all true. However, a necessary condition for the match to light when struck is (we can suppose) its being dry. Consequently, if the match had not lit, this could be either because it was not struck or because it was not dry. That is, $\ulcorner (\sim D) M \sim L \urcorner$ is true. But had the match not been dry, then it would not have lit if struck, i.e., $\ulcorner S > L \urcorner$ would have been false. Thus there is something ($\ulcorner \sim D \urcorner$) which might have been true if the match had lit and which is such that had it been true then $\ulcorner S > L \urcorner$ would have been false. So $\ulcorner (S > L) E \sim L \urcorner$ is false, and hence this is too strong a requirement for the truth of $\ulcorner S \gg L \urcorner$.

The reason that $\ulcorner (P > Q) E \sim Q \urcorner$ will not in general be true if $\ulcorner P \gg Q \urcorner$ is true is that $\ulcorner P > Q \urcorner$ is only true because certain other propositions are also true which, in collaboration with $P$, bring about the truth of $Q$. On the hypothesis that $Q$ is false, all that we can conclude is that either $P$ is false or one of these collateral truths is false, and in the latter case $\ulcorner P > Q \urcorner$ would not normally remain true. We can handle this difficulty by, in effect, building into our counterfactual hypothesis an explanation for *why* $Q$ is false. My proposal is that our analysis should be:

(5.1) $$(P \gg Q) \equiv [(P > Q) E (\sim P \ \& \ \sim Q)].$$

Whereas the hypothesis $\ulcorner\sim Q\urcorner$ leaves open whether it is $P$ or one of the collateral truths that is false, the hypothesis $\ulcorner\sim P \ \& \ \sim Q\urcorner$ tells us which is false. For example, if the match were not struck and did not light, then it would still have been dry, and so had it been struck it would have lit. Thus it seems to me that 5.1 captures precisely what is meant by the necessitation conditional. Notice that what 5.1 requires is just that $\ulcorner P > Q\urcorner$ would be true even if it were a counterfactual.

There appears to be a certain circularity in the analyses of necessitation conditionals and 'even if' conditionals. We have analyzed necessitation conditionals in terms of 'even if' conditionals and the simple subjunctive, but in analyzing 'even if' conditionals we used necessitation conditionals. Fortunately, this circularity can be untangled. It will be shown in Section 6 that the above analyses entail simpler analyses that use only the simple subjunctive, thus avoiding the circularity.

Next let us consider what logical inferences ought to be valid for necessitation conditionals. Logical entailment is a special case of necessitation. If $P$ entails $Q$, then certainly $P$'s being true would necessitate $Q$'s being true:

(5.2) $\quad (P \rightarrow Q) \supset (P \gg Q).$

Obviously a necessitation conditional entails a simple subjunctive:

(5.3) $\quad (P \gg Q) \supset (P > Q).$

It will be a consequence of this analysis that all counterfactual conditionals express necessitation:

(5.4) $\quad \sim Q \ \& \ (P > Q). \supset (P \gg Q).$

This is in accord with the intuition that there are just two ways for a simple subjunctive to be true: (1) the conclusion may be true already, and the antecedent's being true would not interfere with that; (2) the antecedent's being true would necessitate the conclusion's being true. If the conditional is counterfactual, so that the conclusion is false, then possibility (1) is eliminated and the only way the conditional can be true is if the antecedent necessitates the consequent.

A surprisingly wide variety of standard principles fail for necessitation conditionals. To begin with, transitivity fails:

(5.5) It is not true in general that if $\ulcorner(P \gg Q) \ \& \ (Q \gg R)\urcorner$ is true then $\ulcorner(P \gg R)\urcorner$ is true.

For example, suppose we have a stick of old dynamite which would explode ($E$) if dropped ($D$), but which would not explode if first soaked in water ($W$). Then we have $\ulcorner(W \ \& \ D. \gg D)\urcorner$ and $\ulcorner(D \gg E)\urcorner$, but we do not have $\ulcorner(W \ \& \ D. \gg E)\urcorner$.

Nor do we have the *Consequence Principle* that if $P$ necessitates $Q$, then $P$ necessitates anything entailed by $Q$:

(5.6) It is not true in general that if $\ulcorner(P \gg Q) \ \& \ (Q \rightarrow R)\urcorner$ is true then $\ulcorner(P \gg R)\urcorner$ is true.

For example, it is presumably true that my car would still be white even if the maple tree in my front yard were to die. Thus if that maple tree were to die, the conjunction 'My car is white and the maple tree in my front yard died' would be true. This is a counterfactual conditional, so it is an instance of necessitation. The consequent entails 'My car is white', but clearly this is not necessitated by the maple tree's dying. The problem is that $P$ may necessitate a conjunction (i.e., bring it about that the conjunction is true) by making one conjunct true, where the other conjunct is already true and would still be true even if $P$ were true; but just because $P$ necessitates the conjunction, it does not follow that $P$ necessitates each conjunct.

In light of 5.6, it might be objected that our analysis of necessitation is defective. It may be felt that insofar as $P$ necessitates a conclusion, it must necessitate *all* of that conclusion. One can certainly define such a notion of 'strong necessitation' in terms of our present notion of necessitation:

(5.7) $(P \ggg Q) \equiv (R)[(Q \rightarrow R) \supset (P \gg R)]$.

However, I suspect that strong necessitation is not a useful concept. For example, it is not the case that if the truth of $P$ would *cause* the truth of $Q$, then $P$ strongly necessitates $Q$. This is because $Q$ will characteristically entail all sorts of things to which $P$ is irrelevant. For example, my pushing a button may cause a certain stick of dynamite to explode. The statement 'Stick $d$ of dynamite explodes' entails 'There is dynamite in the world', but the latter is certainly not necessitated by my pushing the button.

The notion of necessitation that I am trying to analyse here is that of the truth of one statement 'bringing it about' that another statement is

true. Characteristically, the truth of the second statement will be brought about by making true whatever parts of it are not already true and leaving unchanged those parts that are already true. Then the latter parts will not be necessitated by the first statement.

In connection with 5.6, one should realize that necessitations are not 'necessary connections', except perhaps in a very weak sense. It can be completely accidental that one statement necessitates another, because it is accidental that the unnecessitated part of the second statement is true. Perhaps the term 'necessitation' is inappropriate for the notion I have in mind here, but I have been unable to find a better term.

Contraposition fails for the same sort of reason that the Consequence Principle fails:

(5.8) It is not true in general that if $\ulcorner(P \gg Q)\urcorner$ is true then $\ulcorner(\sim Q \gg \sim P)\urcorner$ is true.

For example, although my tree's dying ($D$) would necessitate the conjunction that my tree died and my car is white ($\ulcorner D \ \& \ W\urcorner$), its being false that both my tree died and my car is white would not necessitate that my tree did not die, i.e., $\ulcorner\sim(D \ \& \ W) \gg \sim D\urcorner$ is not true. This is because if $\ulcorner D \ \& \ W\urcorner$ were true, then if the conjunction 'My tree died and my car is white' were false, it might be false in either of two ways: it might be false that my tree died, and it might be false that my car is white. Only the first of these requires that my tree did not die, so the falsity of the conjunction does not necessitate that my tree did not die.

Surprisingly, adjunctivity fails:

(5.9) It is not true in general that if $\ulcorner(P \gg Q) \ \& \ (P \gg R)\urcorner$ is true, then $\ulcorner P \gg (Q \ \& \ R)\urcorner$ is true.

The best way to muster intuitions against the principle of adjunctivity is to concentrate on those instances of necessitation which are causal. Our intuitions seem to be clearest in those cases. Those cases occur when $P$'s being true would be *causally sufficient* for $Q$ to be true. Making this precise, $P$ is causally sufficient for $Q$ iff $P$'s being true would cause $Q$ to be true if $P$ and $Q$ weren't already true. There is a wide variety of cases in which $P$ would necessitate $Q$ iff $P$ would be causally sufficient for $Q$. The following counterexample to the principle of adjunctivity is such a case. Suppose we have a button and two lights

$A$ and $B$. If light $A$ is off, pushing the button results in light $A$'s being on and light $B$'s being off; if light $A$ is on, pushing the button results in light $A$'s remaining on and light $B$'s being on. Suppose both lights are on. The button's being pushed is causally sufficient for light $A$ to be on. Furthermore, because light $A$ is on, pushing the button is also causally sufficient for light $B$ to be on – that is, if light $B$ were not on, pushing the button would cause it to come on. But pushing the button is not causally sufficient for both lights to be on; if they were not both on, then light $A$ might be off, in which case pushing the button would cause only light $A$ to be on.

Although adjunctivity fails, it will turn out later that a weakened version of adjunctivity does hold:

(5.10) $[(P \gg Q)\ \&\ (P \gg R)\ \&\ (\sim P > \sim Q)\ \&\ (\sim P > \sim R)] \supset [P \gg (Q\ \&\ R)]$.

The principle of dilemma also fails for necessitation conditionals:

(5.11) It is not true in general that if $\ulcorner (P \gg R)\ \&\ (Q \gg R)\urcorner$ is true then $\ulcorner (P \vee Q) \gg R\urcorner$ is true.

For example, consider a complicated circuit consisting of a light and three switches $A$, $B$, and $C$. If switches $A$ and $B$ are both closed then the light is on, and if switch $C$ is closed then the light is on. Switches $A$ and $B$ can be operated independently of one another, however there is an interlock system which prevents switch $C$ from being closed unless switch $A$ is already closed. Suppose in fact that all three switches are closed and the light is on. Switch $C$'s being closed necessitates that the light is on, i.e., $\ulcorner C \gg L\urcorner$ is true. Because switch $A$ is closed, switch $B$'s being closed also necessitates that the light is on, i.e., $\ulcorner B \gg L\urcorner$ is true. But $\ulcorner (B \vee C) \gg L\urcorner$ is false. This is because if the light were off and switches $B$ and $C$ were both open (i.e., if $\ulcorner \sim (B \vee C)\ \&\ \sim L\urcorner$ were true), then as switch $C$ would be open, switch $A$ would also have to be open, making $\ulcorner B > L\urcorner$ false, and hence $\ulcorner (B \vee C) > L\urcorner$ would be false.

Once again, it will turn out that a weakened version of the principle of dilemma is true:

(5.12) $[(P \gg R)\ \&\ (Q \gg R)\ \&\ (\sim R > \sim P)\ \&\ (\sim R > \sim Q)] \supset [(P \vee Q) \gg R]$.

It is apparent that '≫' is a peculiar kind of conditional. We must be quite careful about assuming that it satisfies standard logical laws when it does not.

## 6. Simple subjunctives

We analyzed necessitation in terms of 'even if' and simple subjunctives, we analyzed 'even if' in terms of necessitation and 'might be', and we analyzed 'might be' in terms of the simple subjunctive. I have promised to untangle the circularity between necessitation and 'even if', so what remains is to give an analysis of simple subjunctives. That task will be relegated to the next three chapters. In the meantime, and preparatory to giving a full-fledged analysis, we can state some principles regarding when simple subjunctives are true and when they are false, but these principles merely relate simple subjunctives to one another rather than defining them in terms of something new. Thus, in effect, all we are doing is providing some axioms for simple subjunctives.[4] Nevertheless, I think that these axioms will throw considerable light on simple subjunctives, and will yield some interesting results concerning the relations between our four kinds of conditionals. Furthermore, the axioms will provide a condition of adequacy against which a putative analysis of simple subjunctives can be tested. If such an analysis does not make the axioms true, this is a reason for doubting the analysis.

It is possible to say a great deal about the circumstances under which simple subjunctives are true. First, suppose someone asserts $\ulcorner(P > Q)\urcorner$, and upon investigation we find that $P$ and $Q$ are both true. This verifies the assertion that if $P$ were true, $Q$ would be true; because $P$ is true, and sure enough, $Q$ is true too. For example, we might affirm of a political candidate, 'If he were elected, he would end the war'. If the candidate is subsequently elected, and does end the war, this shows that we were right. So we have:

(6.1) $\quad (P \ \& \ Q) \supset (P > Q).$

Admittedly, there is something odd about asserting a subjunctive conditional when we already know that the antecedent and consequent

are true. If we already know that they are true, there is no good reason to use the subjunctive mood. But, to paraphrase Grice, this is a remark about conversation and not about logic. And I suspect that a further source of reluctance to accept 6.1 results from confusing the simple subjunctive with necessitation conditionals. 6.1 clearly fails for the latter.

It seems evident that entailments generate subjunctive conditionals:

(6.2) $(P \rightarrow Q) \supset (P > Q)$.

And it seems clear that adjunctivity ought to hold for simple subjunctives:

(6.3) $(P > Q)\ \&\ (P > R).\supset [P > (Q\ \&\ R)]$.

Dilemma ought to hold too: if $R$ would be true if $P$ were true, and $R$ would be true if $Q$ were true, then $R$ would be true if either $P$ or $Q$ were true:

(6.4) $(P > R)\ \&\ (Q > R).\supset [(P \vee Q) > R]$.

If $Q$ would be true if $P$ were true, then anything logically entailed by $Q$ would be true if $P$ were true:

(6.5) $(P > Q)\ \&\ (Q \rightarrow R).\supset (P > R)$.

Obviously:

(6.6) $(P > Q) \supset (P \supset Q)$.

Taking '$\leftrightarrow$' to symbolize logical equivalence:

(6.7) $(P \leftrightarrow Q)\ \&\ (Q > R).\supset (P > R)$.

Simple subjunctives are 'defeasible' in the sense that we can have $\ulcorner(P > Q)\urcorner$ true, but for some $R$, $\ulcorner(P\ \&\ R. > Q)\urcorner$ false. However, there is at least one case in which conjoining $R$ with $P$ cannot defeat the conditional. If $R$ would be true if $P$ were true, then, in some sense, $\ulcorner(P\ \&\ R)\urcorner$ being true is not a different circumstance from $P$ being true, so if $Q$ would be true if $P$ were true, then $Q$ would also be true if $\ulcorner(P\ \&\ R)\urcorner$ were true:

(6.8) $(P > Q)\ \&\ (P > R).\supset [(P\ \&\ R) > Q]$.

There are some natural laws that do not hold for simple subjunctives. Transitivity fails. Consider our stick of old dynamite again. If it were dropped, it would explode; and if it were soaked in water and then dropped, it would be dropped; but it is not the case that if it were soaked in water and then dropped, it would explode. Similarly, contraposition fails. It might be true that it would not rain if the witch doctor did a rain dance, but false that if it were to rain he would not have done a rain dance.

From 6.1–6.8 we can derive a series of interesting theorems. First, we obtain theorems which enable us to disentangle the circularity in the definitions of '$E$' and '$\gg$':

(6.9) $(P \gg Q) \supset (P > Q)$.

*Proof:* Suppose $\ulcorner P \gg Q \urcorner$ is true. Then $\ulcorner (P > Q) E (\sim P \ \& \ \sim Q) \urcorner$ is true, so $\ulcorner P > Q \urcorner$ is true.

(6.10) $(P \ \& \ Q. > R) \supset (P > . Q \supset R)$.

*Proof:* Suppose $\ulcorner P \ \& \ Q. > R \urcorner$ is true. $R$ entails $\ulcorner Q \supset R \urcorner$, so by 6.5, $\ulcorner (P \ \& \ Q) > (Q \supset R) \urcorner$ is true. $\ulcorner (P \ \& \ \sim Q) \urcorner$ entails $\ulcorner Q \supset R \urcorner$, so by 6.2, $\ulcorner (P \ \& \ \sim Q) > (Q \supset R) \urcorner$ is true. Thus by 6.4, $\ulcorner (P \ \& \ \sim Q. \vee . P \ \& \ Q) > (Q \supset R) \urcorner$ is true. So by 6.7, $\ulcorner P > . Q \supset R \urcorner$ is true.

Using '$\Box$' to symbolize logical necessity:

(6.11) $\Box Q \supset QEP$.

*Proof:* Suppose $\ulcorner \Box Q \urcorner$ is true. Then $Q$ is true. Suppose $\ulcorner (P \ \& \ R) \gg \sim Q \urcorner$ is true. Then by 6.9, $\ulcorner (P \ \& \ R) > \sim Q \urcorner$ is true. As $\ulcorner \Box Q \urcorner$ is true, $\ulcorner \sim Q \urcorner$ entails $\ulcorner \sim R \urcorner$, so by 6.5, $\ulcorner (P \ \& \ R) > \sim R \urcorner$ is true. By 6.10, $\ulcorner P > . R \supset \sim R \urcorner$ is true, so by 6.5, $\ulcorner P > \sim R \urcorner$ is true. Therefore, $\ulcorner QEP \urcorner$ is true.

(6.12) $(P \rightarrow Q) \supset (P \gg Q)$.

*Proof:* $\ulcorner P \rightarrow Q \urcorner$ entails $\ulcorner P > Q \urcorner$, so $\ulcorner (P \rightarrow Q) \supset \Box (P > Q) \urcorner$ is true (I assume that logical necessity satisfies the axioms of S4), By 6.11, $\ulcorner \Box (P > Q) \supset (P > Q) E \sim Q \urcorner$ is true, so $\ulcorner (P \rightarrow Q) \supset (P \gg Q) \urcorner$ is true.

(6.13) $(QEP) \equiv [Q \ \& \ (P > Q)]$.

*Proof:* Suppose $\ulcorner QEP \urcorner$ is true. Then $Q$ is true. and $(\forall R)\ulcorner[(P\ \&\ R. \gg \sim Q) \supset (P > \sim R)]\urcorner$ is true. But $\ulcorner P\ \&\ \sim Q\urcorner$ entails $\ulcorner \sim Q\urcorner$, so by 6.12, $\ulcorner (P\ \&\ \sim Q) \gg \sim Q\urcorner$ is true. Thus $\ulcorner P > \sim\sim Q\urcorner$ is true, and hence by 6.5, $\ulcorner P > Q\urcorner$ is true.

Conversely, suppose $\ulcorner Q\ \&\ (P > Q)\urcorner$ is true. Suppose $\ulcorner (P\ \&\ R) \gg \sim Q\urcorner$ is true. By 6.9, $\ulcorner (P\ \&\ R) > \sim Q\urcorner$ is true. So by 6.10, $\ulcorner P > .R \supset \sim Q\urcorner$ is true. By 6.3, $\ulcorner P > [Q\ \&\ (R \supset \sim Q)]\urcorner$ is true, so by 6.5, $\ulcorner P > \sim R\urcorner$ is true. Thus $\ulcorner QEP\urcorner$ is true.

Theorem 6.13 allows us to break the circle between necessitation and 'even if' and define both in terms of the simple subjunctive. 6.13 gives us the definition of 'even if', and consequently we have:

(6.14) $\quad (P \gg Q) \equiv .(P > Q)\ \&\ [(\sim P\ \&\ \sim Q) > (P > Q)].$

It is now a simple matter to derive principles 5.2–5.5, 5.10 and 5.12 which were listed above for necessitation conditionals, and principles 3.2–3.6 for 'even if'.

Although we introduced the simple subjunctive in terms of axioms rather than by giving a definition of it, we can now prove that it is definable in terms of our other kinds of conditionals:

(6.15) $\quad (P > Q) \equiv (\exists R)[REP\ \&\ (P\ \&\ R. \gg Q)].$

*Proof:* Suppose $\ulcorner P > Q\urcorner$ is true. Then $\ulcorner P \supset Q\urcorner$ is true, and as $Q$ entails $\ulcorner P \supset Q\urcorner$, $\ulcorner P > (P \supset Q)\urcorner$ is true. Thus by 6.13, $\ulcorner (P \supset Q)EP\urcorner$ is true. And $\ulcorner P\ \&\ (P \supset Q)\urcorner$ entails $Q$, so by theorem 6.12, $\ulcorner [P\ \&\ (P \supset Q)] \gg Q\urcorner$ is true.

Conversely, suppose $\ulcorner REP\ \&\ (P\ \&\ R. \gg Q)\urcorner$ is true. By (6.13), $\ulcorner P > R\urcorner$ is true. By 6.9 and 6.10, $\ulcorner P > (R \supset Q)\urcorner$ is true, so by 6.3 and 6.5, $\ulcorner P > Q\urcorner$ is true.

From 6.14 we immediately obtain:

(6.16) $\quad \sim Q \supset [(P > Q) \equiv (P \gg Q)].$

This means that a counterfactual always expresses necessitation. This, together with the fact that counterfactuals are those subjunctive conditionals that philosophers have most often thought about, explains the pervasive view that subjunctive conditionals always express necessitation.

From 6.13 and 6.16 we can see that the simple subjunctive is just the disjunction of necessitation and 'even if':

(6.17) $(P > Q) \equiv [QEP \vee (P \gg Q)]$.

Thus there are just these two ways that the simple subjunctive can be true. Either $Q$ is *made true* by $P$, or $Q$ is already true and $P$ would not disrupt this. This is a very illuminating theorem. It explains why the logic of '>' is so peculiar, being, as it is, a mixture of two such different concepts.

Each of our four kinds of conditionals is explicitly definable in terms of each of the others. This follows from the fact, already established, that each kind of conditional is definable in terms of the simple subjunctive, together with the following theorem according to which the simple subjunctive is definable in terms of each of the other kinds of conditionals:

(6.18) $\ulcorner P > Q \urcorner$ is equivalent to each of the following:

(i) $\ulcorner \sim[(\sim Q)MP] \urcorner$;

(ii) $\ulcorner (P \supset Q)EP \urcorner$;

(iii) $\ulcorner P \gg (P \supset Q) \urcorner$.

Consequently, if we can provide an analysis of any of these kinds of conditionals, analyses of the others will follow.

## 7. The axiomatization of simple subjunctives

Principles 6.1–6.8, in effect, constitute an axiomatization of simple subjunctives. However, principles 6.2, 6.5, and 6.7 employ the concept of entailment, and thus require a modal logic for the underlying logic rather than just the propositional calculus. We can instead replace those principles by rules of inference in a more restrictive language in which entailment cannot be expressed. Let *SS* be the formal theory whose axioms and rules are as follows:

A1 All tautologies.

A2 $(P > Q)\ \&\ (P > R).\supset [P > (Q\ \&\ R)]$.

A3 $(P > R)\ \&\ (Q > R).\supset [(P \vee Q) > R]$.

A4 $(P > Q)\ \&\ (P > R).\supset[(P\ \&\ Q) > R]$.

A5 $(P\ \&\ Q)\supset(P > Q)$.

A6 $(P > Q)\supset(P \supset Q)$.

R1 If $P$ and $\ulcorner(P \supset Q)\urcorner$ are theorems, so is $Q$.

R2 If $\ulcorner(P \supset Q)\urcorner$ is a theorem, so is $\ulcorner(P > Q)\urcorner$.

R3 If $\ulcorner(Q \supset R)\urcorner$ is a theorem, so is $\ulcorner(P > Q)\supset(P > R)\urcorner$.

R4 If $\ulcorner(P \equiv Q)\urcorner$ is a theorem, so is $\ulcorner(P > R)\supset(Q > R)\urcorner$.

I conjecture that *SS* contains as theorems all true principles regarding simple subjunctives that can be formulated in this language.

It is of interest to compare *SS* with other well known theories of subjunctive conditionals. The best known such theories are *C*1 of Lewis (1972) and *CQ* of Stalnaker (1968). *SS* is contained in *C*1 which is contained in *CQ*. *CQ* contains the theorem $\ulcorner(P > Q)\vee(P > \sim Q)\urcorner$, which we have rejected. *SS* is weaker than *C*1. *C*1 can be obtained from *SS* by adding the following axiom:

(7.1) $(P > Q)\ \&\ RMP.\supset[(P\ \&\ R) > Q]$.

Unfortunately, this axiom is false. This axiom would be valid only if the ordering of possible worlds according to magnitude of change were connected, and we saw in Chapter I that it is not. We can construct counterexamples to 7.1 using the same constructions that showed the ordering not to be connected. Let *S*, *T*, and *U* be any three unrelated false statements, e.g., 'My car is painted black', 'My garbage can blew over', and 'My maple tree died'. The following is a substitution instance of 7.1:

(7.2) $[(S \vee T) > \sim U]\ \&\ (U \vee T)M(S \vee T).$
$\supset[(S \vee T)\ \&\ (U \vee T).> \sim U]$.

From 7.2 we readily obtain the principle:

(7.3) $\sim[(S \vee T) > S]\ \&\ \sim\{[(S\ \&\ U) \vee T) > T\}.\supset UM(S \vee T)$.

The color of my car and the state of my garbage can are irrelevant (we can suppose) to the state of my tree, so my tree would not die even if either my car were painted black or my garbage can blew over; hence $\ulcorner UM(S \vee T)\urcorner$ is false. But the antecedent of 7.3 is true. Disjunctions

whose disjuncts are unrelated to one another cannot necessitate either disjunct. If we know that the disjunction is true, that leaves open which disjunct is true. In particular, it is not true that if either my car were painted black or my garbage can blew over then my car would be painted black; and it is not true that if either my car were painted black and my maple tree died, or my garbage can blew over, then my garbage can would have blown over. Thus 7.3 is false. Consequently, as we would expect from its semantics, *C*1 is too strong.

We have seen that the stronger brethren of *SS* contain invalid axioms. Is it true then that *SS* is complete? We are not really in a position to answer this question, because we do not have a satisfactory semantics for subjunctive conditionals. However, it is possible to construct a semantics for *SS* based upon the Stalnaker-Lewis semantics. Stalnaker has pointed out[5] that *SS* is complete on a semantics which is identical to Lewis' 'Analysis 2' except that the ordering of equivalence classes of worlds is only a partial ordering rather than a simple ordering. In light of the discussion of the ordering of worlds in Chapter I, this is at least suggestive of the 'real completeness' of *SS*.

## 8. Conclusions

We have distinguished between four different kinds of subjunctive conditionals and explored their logical properties. Of these, it is particularly important to distinguish between the simple subjunctive and the necessitation conditional. Philosophers have often rejected theses about the simple subjunctive by appealing to intuitions only appropriate to necessitation. It has been shown that our other kinds of conditionals are definable in terms of the simple subjunctive, so if we can provide an analysis of the simple subjunctive we will have an analysis of them all. And we have a set of axioms for the simple subjunctive. These will be of importance in testing any putative analysis. If such an analysis does not make these axioms valid, this is at least a reason for being suspicious of the analysis.

## NOTES

[1] The modality expressed here by 'couldn't', as in 'The match couldn't light', is an intriguing one. Although, as is well known, a modality can be defined in terms of

subjunctive conditionals (as $\ulcorner P > (Q \ \& \sim Q)\urcorner$), this modality is not the same as 'couldn't'. It will result from the discussion of Chapter IV that if $\ulcorner P > (Q \ \& \sim Q)\urcorner$ is true, then $P$ is necessarily false. Thus this cannot be a correct analysis of 'couldn't' as it occurs in this context. A different analysis will be proposed in Chapter III in terms of 'actual necessity'.

[2] I am indebted to David Lewis for this example.

[3] Or more precisely, $Q$ is true in every world that might be actual if $P$ were true.

[4] In the choice of these axioms, I have been influenced by the work of David Lewis.

[5] In correspondence.

CHAPTER III

# SUBJUNCTIVE GENERALIZATIONS

## 1. Introduction

An understanding of laws is fundamental to an understanding of subjunctive conditionals. I argued in Chapter I that laws are themselves subjunctive. A law says not just that all actual $A$'s are $B$'s, but that *any* $A$ *would be* a $B$. Statements of this latter form will be called subjunctive generalizations. They are to be contrasted with *material generalizations* which have the form $\ulcorner(x)(Ax \supset Bx)\urcorner$. Our problem in this chapter is to understand subjunctive generalizations.

Subjunctive generalizations look initially like universally quantified subjunctive conditionals of some sort. There is more than one kind of subjunctive conditional, but the most obvious candidate is the simple subjunctive. According to this proposal, the subjunctive generalization $\ulcorner$Any $A$ would be a $B\urcorner$ is equivalent to $\ulcorner(x)(Ax > Bx)\urcorner$.[1]

There are several difficulties for this proposal. The first difficulty is that the simple subjunctive is automatically true if both antecedent and consequent are true, i.e., $\ulcorner(P \,\&\, Q)\urcorner$ entails $\ulcorner(P > Q)\urcorner$. It would seem to follow that $\ulcorner(x)(Ax \,\&\, Bx)\urcorner$ entails $\ulcorner(x)(Ax > Bx)\urcorner$. But it is quite clear that $\ulcorner(x)(Ax \,\&\, Bx)\urcorner$ does not entail $\ulcorner$Any $A$ would be a $B\urcorner$. For example, consider Black Bart who dislikes everything except redheaded women who like him. Regrettably, there are no redheaded women who like Black Bart, although were there any he would like them. If we let $A$ be a tautological predicate and $B$ be the predicate 'is disliked by Black Bart', then $\ulcorner(x)(Ax \,\&\, Bx)\urcorner$ is true. But it is certainly not true that anything at all *would* be disliked by Black Bart. On the contrary, we know of something that would not be disliked by Black Bart – namely, redheaded women who like him. Thus it appears that this analysis of subjunctive generalization fails.

Apparently a subjunctive generalization cannot be expressed as $\ulcorner(x)(Ax > Bx)\urcorner$. We have a genuinely different subjunctive here. It is still natural to symbolize it as a universally quantified conditional, but

we must use some conditional other than the simple subjunctive. Let us use '$\Rrightarrow$' to symbolize ⌜Any $A$ would be a $B$⌝ as ⌜$(x)(Ax \Rrightarrow Bx)$⌝. But this leads to unexpected difficulties. What do the quantifiers range over? They cannot be taken as ranging over just existent objects, because ⌜Any $A$ would be a $B$⌝ entails the simple subjunctive ⌜$(Aa > Ba)$⌝ even for non-existent objects. For example, if I know that any raven would be black, I can conclude that if there were a raven in the next room, it would be black. This is connected with the problem just discussed for symbolizing the subjunctive generalization as ⌜$(x)(Ax > Bx)$⌝. The problem seems to be in part that the subjunctive generalization is not just about all actual objects, and hence it cannot be symbolized as ⌜$(x)(Ax \Rrightarrow Bx)$⌝ because the quantifier in this formula does range only over actual objects.

A natural suggestion is that the quantifier in a subjunctive generalization ranges over all *physically possible objects*. For example, when we say that any raven would be black, we are claiming not just that any actual object would be black if it were a raven, but that any *possible raven* would be black. Putting the suggestion this way makes it sound metaphysically outrageous. What is this domain of possible objects over which the quantifiers are supposed to range? However, the suggestion can be rephrased in a way that makes it appear much more reasonable. Let us say that a proposition is *physically possible*, and symbolize this as ⌜$\underset{p}{\Diamond} P$⌝, just in case there is no physical law that is inconsistent with it. We then define physical necessity, ⌜$\underset{p}{\Box} P$⌝, as meaning ⌜$\sim \underset{p}{\Diamond} \sim P$⌝. Talk about quantification over physically possible objects can then be replaced by talk about the physical necessity of the result of quantifying over actual objects. In other words, the proposal is that ⌜Any $A$ would be a $B$⌝ can be analyzed as ⌜$\underset{p}{\Box}(x)(Ax \Rrightarrow Bx)$⌝.

A serious difficulty for this proposal is that we have defined physical necessity in terms of the notion of a physical law, but physical laws are subjunctive generalizations, and so it may (and in fact will – see below) turn out that this proposal is ultimately circular. However, there is another more serious difficulty – even if we could ignore the circularity, the proposed analysis is false. This can be seen by considering subjunctive generalizations whose antecedents are physical impossible. Let us

call these *counter-legals*, following Nelson Goodman (1955). For example, suppose it is true, as has been conjectured, that any pulsar would be a neutron star. Then it is true (non-vacuously) that any pulsar which is not a neutron star would emit periodic bursts of radiation, but it is false that any pulsar which is not a neutron star would be a neutron star. In other words, we distinguish between true and false counter-legals. They are not all vacuously true merely because the antecedents are false of all physically possible objects. Furthermore, such counter-legals entail subjunctive conditionals about *physically impossible* objects. For example, the above generalization entails 'If there were a pulsar just to the left of Alpha Centauri which was not a neutron star, it would emit periodic bursts of radiation'. This seems to indicate that these subjunctive generalizations are not just about physically possible objects after all. What they are about seems to be a function partly of the nature of the antecedent.

I think that the attempt to analyze subjunctive generalizations in this way as universally quantified subjunctive conditionals is essentially bankrupt, for the reasons just outlined. I believe it will prove more fruitful to turn directly to an examination of the way subjunctive generalizations work rather than trying to reduce them to something else. Once we have an account of the way subjunctive generalizations work, it will prove possible to characterize them in terms of physical necessity and quantification in ways more complicated than those considered above. However, such characterizations will still not constitute analyses because of the circularity involved in appealing to physical necessity.

## 2. Rudiments of an analysis

To simplify our writing tasks, let us symbolize ⌜Any $A$ would be a $B$⌝ as ⌜$Ax \Rrightarrow Bx$⌝. This is intended merely as a symbolization, and in no way reflects an analysis of these subjunctive generalizations. Notice that in this symbolization, '$\Rrightarrow$' is not a conditional – it connects predicates, or more generally open formulas, rather than sentences. It is actually a binary variable-binding operator, binding, as it does, the free variables in ⌜$Ax$⌝ and ⌜$Bx$⌝. More generally, whenever $\varphi$ and $\psi$

are open formulas containing the same $n$ free variables, $\ulcorner \varphi \Rrightarrow \psi \urcorner$ says that any $n$-tuple satisfying $\varphi$ would satisfy $\psi$. In $\ulcorner \varphi \Rrightarrow \psi \urcorner$, '$\Rrightarrow$' binds the $n$ free variables. As an additional piece of jargon, we will read $\ulcorner \varphi \Rrightarrow \psi \urcorner$ as saying that something's satisfying $\varphi$ *implicates* its satisfying $\psi$. Our problem is now to clarify the way this operator works.

It is simple to give a rather vague general picture of how subjunctive generalizations work. We start by confirming a number of generalizations inductively. Let us call the set of these basic generalizations $N$. Then we derive new generalizations from those in $N$ using various logical inferences. Obviously, we do want to be able to derive generalizations from those we confirm directly, but there is a more important reason than simple expediency for having this two step procedure. Many generalizations (in fact, I think most) fail to be "projectible" in the sense of Goodman (1955).[2] The only way to confirm non-projectible generalizations is 'indirectly' by deriving them logically from projectible conditionals that have been confirmed directly. Consequently, this two step pattern of confirmation for subjunctive generalizations is fundamental to the whole concept of inductive confirmation.

In explaining how subjunctive generalizations work, it suffices to say (1) how projectible generalizations are confirmed directly by induction, and (2) how we derive new generalizations from those we have confirmed already. (1) is a familiar problem. I discussed it at length in Pollock (1974), where I claimed to give a complete account of such confirmation. I have nothing new to add to that account, so let us proceed to (2). I believe that once we have an account of both (1) and (2), we can claim to have a complete analysis of subjunctive generalizations. This will not be an analysis of the familiar truth-condition sort, but I seriously doubt whether that sort of analysis is possible here. Instead, what we will have is a complete account of how to operate with the concept of a subjunctive generalization, and that suffices to characterize the concept just as much as a truth-condition analysis would. In the jargon of Pollock (1974), I am proposing that we give an analysis of subjunctive generalizations in terms of their justification conditions.[3]

Before turning to the task of elucidating the inferences that allow us to derive new generalizations from those in $N$, we must make a distinction which has generally been overlooked. There are really two

different kinds of subjunctive generalizations. When we think of subjunctive generalizations, we think most naturally of laws. Those laws that we confirm inductively must have physically possible antecedents (otherwise there would be no confirming instances for them), so the problem of interpreting the apparent quantifiers in them is greatly simplified. Insofar as $\ulcorner \underset{p}{\Diamond}(\exists x)Ax \urcorner$ is true, we can think of $\ulcorner A \Rrightarrow B \urcorner$ as saying simply that any physically possible $A$ would be a $B$. These law statements, which are in some sense about all physically possible objects, constitute one kind of subjunctive generalization. But philosophers have often overlooked the fact that there is another kind of subjunctive generalization. For example, Chisholm (1955) gives as an example of a law, 'Anyone who drank from that bottle would be poisoned', said of a certain bottle containing poison. But this is obviously not a law. It is not the least bit plausible to regard Chisholm's generalization as saying that laws of nature dictate that any person who drank from that bottle would be poisoned. It would be quite possible to have someone drink from the bottle without being poisoned – all we would have to do would be to wash the bottle out first and fill it with water. And although we are well aware of this, we still regard Chisholm's generalization as true.

We have distinguished between the subjunctive generalization $\ulcorner (Ax \Rrightarrow Bx) \urcorner$ and the universally quantified subjunctive conditional $\ulcorner (x)(Ax > Bx) \urcorner$. It might be supposed that Chisholm's generalization actually has the latter form rather than being a true subjunctive generalization. That it does not can be seen as follows. Suppose there were no living creatures. Under these circumstances, there would be nothing which logically *could* be a person who drinks from the bottle. I am supposing here that it is a necessary feature of a non-living thing that it not be a person. The basic sortals under which non-living things fall are such as to logically preclude their being persons. This is not to say that a non-living thing could not *turn into* a person, but this would be a substantial change and would involve the original object ceasing to exist and the person coming into existence. If a person is created out of clay, the clay ceases to exist. We do not say that the clay 'is now a person'. Assuming this is correct, in a world in which there are no living creatures, $\ulcorner (x)\Box \sim Ax \urcorner$ is true (where $\ulcorner Ax \urcorner$ is $\ulcorner x$ is a person who

drinks from this bottle$^{\neg}$). It is a consequence of our previous account of subjunctive conditionals that the following is valid: $\ulcorner\Box\sim P \supset (P > Q)\urcorner$. Presumably then, we also have $\ulcorner(x)\Box\sim Ax \supset (x)(Ax > Bx)\urcorner$ for any predicate $B$ at all. But this would make both 'Anyone who drank from this bottle would be poisoned' and 'Anyone who drank from this bottle would fail to be poisoned' vacuously true in a world in which there are no living things. However, they would not be vacuously true; the first would be true, and the second false. Consequently, Chisholm's generalization does not have the form $\ulcorner(x)(Ax > Bx)\urcorner$. It is a genuine subjunctive generalization. Therefore, there are true subjunctive generalizations that are not laws.

Although the generalization 'Anyone who drank from this bottle would be poisoned' is not about all physically possible persons who might drink from this bottle, it is in some sense about all 'actually possible' persons who might drink from that bottle. The reason we regard the generalization as true is that although we know a way to enable a person to drink from the bottle without being poisoned – namely, wash the bottle out – we also know that that is not going to happen. The bottle is our container for rat poison, and it is where we are going to continue to keep the rat poison. It is *because* of these facts that anyone who drank from the bottle would be poisoned. And it is because of these facts that, although it is physically possible for a person to drink from the bottle without being poisoned, it is not 'actually possible' for a person to do so. This notion of actual possibility is not a modality with which philosophers are very familiar. It has been largely overlooked, although I suspect that it may be of some importance in explaining various locutions that philosophers have always found puzzling. And I think that, although it is less familiar, it is no more suspect than the notion of physical possibility. As we will see in section four, it can be clearly defined in a manner which should make it philosophically respectable.

Apparently there are these two distinct kinds of subjunctive generalizations, one kind being, in some sense, about all physically possible objects, and the other being only about all actually possible objects. Let us reserve '$\Rrightarrow$' for symbolizing the first kind, and let us call them *strong subjunctive generalizations*; let us use '$\Rightarrow$' to symbolize the second kind, and call them *weak subjunctive generalizations.* In

analyzing subjunctive generalizations, we would like to analyze both of these kinds of subjunctive generalizations. The basic scheme for analyzing them would seem to be the same. In each case we begin by confirming a basic set of generalizations inductively, and then others are derived logically from the basic ones. It will be apparent below that the inferences employed in deriving new generalizations from the basic ones are the same for both kinds of generalizations. Thus the difference between them must arise ultimately from the way the basic ones are inductively confirmed. We do not have to look far to find the source of that difference.

Inductive confirmation is *defeasible.* A statement of evidence $E$ may confirm a generalization $H$, and yet when an additional statement $E'$ (e.g., one containing a counter-example to $H$) is conjoined with $E$, the resulting conjunction may no longer confirm $H$. We say that $E'$ is a *defeater*, and that it *defeats* the confirmation of $H$ by $E$. Explaining how confirmation works is a matter of explaining both what statements constitute evidence and what statements constitute defeaters.[4]. I think that the difference between strong and weak generalizations lies in what statements are defeaters for their inductive confirmation. There is a simple defeater for the confirmation of $\ulcorner A \Rrightarrow B \urcorner$ which is not a defeater for $\ulcorner A \Rightarrow B \urcorner$. The former is about all physically possible objects, but the latter is not. Consequently, a defeater for $\ulcorner A \Rrightarrow B \urcorner$ which is not a defeater $\ulcorner A \Rightarrow B \urcorner$ is ⌜It is physically possible for there to be an $A$ which is not a $B$⌝.

To define the difference between '$\Rrightarrow$' (law statements) and '$\Rightarrow$' in terms of this defeater as I have just stated it is circular, because we defined physical possibility in terms of laws. But the circularity can be avoided by restating this defeater in terms of the grounds we have for thinking that it is true. These grounds are basically inductive. We discover inductively that certain kinds of things are always possible. For example, we might confirm inductively that the 9:14 train from Boston is always late. This is a weak subjunctive generalization rather than a strong subjunctive generalization. It is certainly not physically impossible for a 9:14 train from Boston to arrive on time. For example, suppose the reason the train is always late is that the railroad has only old broken-down equipment. If the equipment were replaced by new equipment, the train would be on time. These observations

constitute a defeater for the strong generalization according to which it would be a law that the 9:14 train from Boston is always late, and it is because of this defeater that we only affirm the weak generalization. The way in which this defeater works is the following. We begin by ascertaining that it is always physically possible to replace old broken-down equipment with new equipment. We can ascertain this because it follows from our definition of physical possibility that whatever is true is possible. Thus if we know of a number of cases in which broken-down equipment has been replaced, we also know of a number of cases in which it is physically possible to replace broken-down equipment. So we have a number of confirming instances for the generalization that it is always physically possible to replace broken-down equipment. Of course, we also know of many cases in which broken-down equipment has not been replaced, but this is irrelevant to the induction. The only thing that would be relevant to the induction as a counter-example to what is being confirmed (that it is always physically possible to replace broken-down equipment) would be a case in which we know it is physically impossible to replace broken-down equipment. To have such a case, we would have to have confirmed a law which entails that the equipment could not be replaced in some particular case. We do not know of such a law, so we have confirmation for the generalization that it is always physically possible to replace broken-down equipment. This entails that it is physically possible to replace the broken-down equipment used on the 9:14 train from Boston. We are supposing we know that if the equipment were replaced, the train would be on time. And I assume the following principle regarding simple subjunctives:

$$\underset{p}{\Diamond} P \;\&\; (P > Q). \supset \underset{p}{\Diamond} Q.$$

Thus we can conclude that it is physically possible for the 9:14 train from Boston to be on time. Hence we have a defeater for the strong generalization that the 9:14 train is always late, and are left with only the weak generalization.

This illustrates the basic scheme by which we defeat strong generalizations by having physically possible counter-examples. I will not attempt to give a general characterization of this scheme here, but that should not be too difficult to do. What has been said should be

sufficient to indicate the source of the difference between strong and weak subjunctive generalizations.

## 3. Strong generalizations

Now let us turn to the task of elucidating the inferences that allow us to derive new generalizations from those that we confirm directly by induction. Let us begin with strong generalizations. Those confirmed directly constitute the set $N$. Then additional generalizations are derived from these. There are inferences valid in the predicate calculus that are not valid here. For example, although $\ulcorner(x)(Ax \ \& \ Bx)\urcorner$ entails $\ulcorner(x)(Ax \supset Bx)\urcorner$, we have seen that it does not entail $\ulcorner(A \Rrightarrow B)\urcorner$. Upon reflection, however, it may seem that all those inferences in the predicate calculus which proceed exclusively from universally quantified conditionals to universally quantified conditions are valid. In other words, letting $\ulcorner\forall\varphi\urcorner$ stand for the universal closure of the open formula $\varphi$, it might be supposed that the following principle holds:

(3.1) If the universally quantified material conditionals $\ulcorner\forall(P_1 \supset Q_n)\urcorner, \ldots, \ulcorner\forall(P_n \supset Q_n)\urcorner$ imply $\ulcorner\forall(R \supset S)\urcorner$ in the predicate calculus, then the strong generalizations $\ulcorner(P_1 \Rrightarrow Q_1)\urcorner, \ldots, \ulcorner(P_n \Rrightarrow Q_n)\urcorner$ entail $\ulcorner(R \Rrightarrow S)\urcorner$.

It can be shown that a necessary and sufficient condition for 3.1 to hold is for the following to hold:

(3.2) $[P(x_1, \ldots, x_n) \Rrightarrow Q(x_1, \ldots, x_n)]$
$\supset [P(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n)$
$\Rrightarrow Q(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n)]$.

(3.3) $[(P \Rrightarrow Q) \ \& \ (P \Rrightarrow R)] \supset [P \Rrightarrow (Q \ \& \ R)]$.

(3.4) $(P \rightarrow Q) \supset (P \Rrightarrow Q)$.

(3.5) $(P \Rrightarrow Q) \supset (\sim Q \Rrightarrow \sim P)$.

(3.6) $[(P \Rrightarrow Q) \ \& \ (Q \Rrightarrow R)] \supset (P \Rrightarrow R)$.

(3.7) $(P \Rrightarrow Q) \supset [(x)P \Rrightarrow (x)Q]$.

At first, it may seem that 3.2–3.7 are unexceptionable. But they

cannot all hold. Jointly they entail:

(3.8) $(P \Rightarrow Q) \supset (P \ \& \ {\sim}Q . \Rightarrow R)$.

Principle 3.8 says that a physically impossible antecedent implicates anything at all. There are clear counterexamples to 3.8. For example, suppose it is a law of nature that no human being can live for more than 200 years, and also (probably contrary to fact) that a human being acquires more and more knowledge the longer he lives. From this we can conclude:

> A three hundred year old human being would know more than a 100 year old human being.

On the other hand, we cannot conclude:

> A three hundred year old human being would know less than a 100 year old human being.

And yet both of these inferences would be licensed by 3.8. Thus 3.8 must be false.

It is rather obvious what is wrong with 3.8. Principles 3.2–3.7 all hold as long as the antecedents of the generalizations are physically possible, i.e., are compatible with all physical laws. But if the antecedent of a generalization is incompatible with some physical law, then we cannot use *that* law in deciding what would happen if the antecedent were true. For example, if $\ulcorner(P \Rightarrow Q)\urcorner$ is true, we cannot assume this law in deciding what characteristics would be possessed by something satisfying $\ulcorner(P \ \& \ {\sim}Q)\urcorner$. In particular, we cannot conclude that $\ulcorner[(P \ \& \ {\sim}Q) \Rightarrow Q]\urcorner$ is true. In deciding what an antecedent implicates, we must rule out all laws incompatible with that antecedent. For example, let us suppose that it is a law that all white dwarf stars must have radii smaller than that of the sun. We cannot conclude from this that any white dwarf larger than the sun would be smaller than the sun. On the other hand, we can non-vacuously conclude that a white dwarf having a radius larger than that of the sun would have a mass greater than that of the sun. This is because the laws used in deriving the latter conclusion are compatible with the supposition of a white dwarf having a radius larger than that of the sun.

More generally, an antecedent may conflict with several laws conjointly but none individually. To illustrate, suppose we have the law

statements

$$(P \Rrightarrow Q), \quad (Q \Rrightarrow R), \quad (R \Rrightarrow S), \quad (P \Rrightarrow T).$$

For example (to compound our tale of science fiction) these might be the laws, 'Any white dwarf is a stellar object of radius smaller than the sun', 'Any stellar object of radius smaller than the sun is of luminosity less than $L_0$' (where $L_0$ is some fixed luminosity), 'Anything of luminosity less than $L_0$ cannot be seen at a distance greater than one billion light years', and 'Any white dwarf would be more massive than the sun'. Then $\ulcorner(P \ \& \ \sim R)\urcorner$ does not conflict with any of these statements individually. Assuming these laws, we can conclude $\ulcorner(P \ \& \ \sim R. \Rrightarrow T)\urcorner$ (i.e., 'Any white dwarf of luminosity no less than $L_0$ would be more massive than the sun'). This comes directly out of the law $\ulcorner(P \Rrightarrow T)\urcorner$, which is compatible with the antecedent $\ulcorner(P \ \& \ \sim R)\urcorner$. But we cannot conclude $\ulcorner(P \ \& \ \sim R. \Rrightarrow S)\urcorner$ (i.e., 'Any white dwarf of luminosity no less than $L_0$ could not be seen at a distance greater than one billion light years'). This is because, in order to get this, we must use all three of $\ulcorner(P \Rrightarrow Q)\urcorner$, $\ulcorner(Q \Rrightarrow R)\urcorner$, and $\ulcorner(R \Rrightarrow S)\urcorner$, and although $\ulcorner(P \ \& \ \sim R)\urcorner$ is compatible with each of these individually, it is incompatible with the set of them.

How do we characterize which of these inferences are valid? We begin with the set $N$ of basic laws that we confirm directly by induction. Then we want to obtain derivative laws from those in $N$. The above picture suggests that $\ulcorner(P \Rrightarrow Q)\urcorner$ follows from $N$ just in case there is some subset $N_P$ of $N$ which is consistent with $P$ and which, by some reasonable principles of inference, implies $\ulcorner(P \Rrightarrow Q)\urcorner$. However, this is still not quite right. The difficulty is that there may be more than one such set $N_P$, and different such sets might lead to different conclusions. For example, suppose we have in $N$ the two laws $\ulcorner(A \Rrightarrow B)\urcorner$ and $\ulcorner(C \Rrightarrow \sim B)\urcorner$. The above principle would allow us to derive both $\ulcorner(A \ \& \ C. \Rrightarrow B)\urcorner$ and $\ulcorner(A \ \& \ C. \Rrightarrow \sim B)\urcorner$ from $N$. But this is clearly wrong. Under these circumstances, we should not be able to derive either conclusion from $N$. Or to take another example, suppose $N$ is the set

$$(P \Rrightarrow Q), \quad (Q \Rrightarrow R), \quad (R \Rrightarrow S), \quad (T \Rrightarrow \sim R).$$

We want to know whether $\ulcorner(P \ \& \ T. \Rrightarrow S)\urcorner$ follows from this. For

example, the first three laws might be as before, and the fourth might be 'Any pulsar is of luminosity at least $L_0$'. Intuitively, we cannot conclude that any white dwarf which is a pulsar could not be seen from a distance of one billion light years. This is because, although its being a white dwarf favors that conclusion, its being a pulsar blocks the normal way of inferring this conclusion from the fact that the object is a white dwarf. What is required for a conditional $\ulcorner(P \Rrightarrow Q)\urcorner$ to be inferrable from $N$ is not for there to be *some* subset $N_P$ consistent with $P$ from which we can infer $\ulcorner(P \Rrightarrow Q)\urcorner$, but for *every maximal* subset of $N$ consistent with $P$ to be such that we can infer $\ulcorner(P \Rrightarrow Q)\urcorner$ from it. In other words, there may be different ways to render $N$ consistent with $P$ by omitting minimally many things from $N$. The supposition that $P$ is true amounts to supposing that one of those minimally altered subsets of $N$ constitutes the actual set of basic laws, but no particular such subset is favored over any other. Any of them might constitute the actual set of laws, and so it is only when every such set allows us to infer $\ulcorner(P \Rrightarrow Q)\urcorner$ that we can conclude that $\ulcorner(P \Rrightarrow Q)\urcorner$ is true given $N$ as it actually is.

Let us say that a set is *P-consistent* just in case it is logically consistent with $\ulcorner(\exists x_1), \ldots (\exists x_n)P\urcorner$ (supposing $P$ to have $n$ free variables). If $X$ is a set of sentences, and $Y$ is a subset of $X$, let us say that $Y$ is a *maximal-P-consistent* subset of $X$ just in case (i) $Y$ is $P$-consistent, and (ii) there is no $Z$ such that $Y \subset Z \subseteq X$ and $Z$ is $P$-consistent. Thus the maximal-$P$-consistent subsets of $N$ are those sets that result from making minimal deletions in $N$ so as to render it consistent with $P$. Then the above account of strong generalizations amounts to saying that $\ulcorner(P \Rrightarrow Q)\urcorner$ is true iff $\ulcorner(P \Rrightarrow Q)\urcorner$ can be inferred from every maximal-$P$-consistent subset of $N$.

But we have not yet explained what sorts of inferences are allowable in inferring $\ulcorner(P \Rrightarrow Q)\urcorner$ from the maximal-$P$-consistent subsets. Suppose $N^*$ is such a subset. As $P$ is consistent with $N^*$, there would seem to be nothing wrong with using 3.2–3.7 in making such inferences. These would seem to constitute a minimal set of rules of inference allowable here. But there is also a rather obvious upper bound to what inferences are allowable. It seems clear that no inference from strong generalizations to strong generalizations is valid when the corresponding inference from material generalizations to material generalizations

is invalid in the predicate calculus. The actual set of allowable inferences must lie somewhat between what is derivable from 3.2–3.7 and this upper bound. But, as was remarked earlier, 3.2–3.7 are conjointly equivalent to principle 3.1, which is the strongest principle that does not transgress this upper bound. Consequently, the upper bound limits us exactly to 3.2–3.7, or equivalently, to 3.1. The inferences that are allowable in getting from $N^*$ to $\ulcorner(P \Rrightarrow Q)\urcorner$ are precisely those inferences that are allowable in the predicate calculus in going from material generalizations to material generalizations.

This means that we can give a very simple characterization of what strong generalizations are true in terms of the set $N$. Let $\forall N$ be the set of material generalizations corresponding to the strong generalizations in $N$. Then we have:

(3.9) $\ulcorner(P \Rrightarrow Q)\urcorner$ is true iff every maximal-$P$-consistent subset of $\forall N$ entails $\ulcorner\forall(P \supset Q)\urcorner$.[5]

Before proceeding further, we must be a bit more precise about the basic generalizations that make up the set $N$. So far all we have said about them is that they must be projectible, but some additional restrictions are required if our account of the way in which laws are derived from $N$ is to work. The difficulty is that projectible laws need not be, in the appropriate sense, 'basic'. For example, if $\ulcorner(A \Rrightarrow B)\urcorner$, $\ulcorner(B \Rrightarrow C)\urcorner \in N$, then $\ulcorner(A \Rrightarrow C)\urcorner$ which is derivable from them is also projectible.[6] However, we should not count $\ulcorner(A \Rrightarrow C)\urcorner$ as basic for the purposes of deriving additional generalizations. If we did treat $\ulcorner(A \Rrightarrow C)\urcorner$ as basic, then we could infer the truth of $\ulcorner(A \,\&\, {\sim}B. \Rrightarrow C)\urcorner$. But we should not be able to infer this because part of the *reason* $\ulcorner(A \Rrightarrow C)\urcorner$ is true is that $\ulcorner(A \Rrightarrow B)\urcorner$ is true, and the antecedent $\ulcorner(A \,\&\, {\sim}B)\urcorner$ conflicts with the latter law. Thus we should require that in order for $\ulcorner(A \Rrightarrow C)\urcorner$ to be in $N$, there cannot be open formulas $B_0, \ldots, B_n$ such that $\ulcorner(A \Rrightarrow B_0)\urcorner$, $\ulcorner(B_0 \Rrightarrow B_1)\urcorner, \ldots, \ulcorner(B_n \Rrightarrow C)\urcorner \in N$. Similarly, we should preclude the case in which $\ulcorner(A \Rrightarrow B_0)\urcorner$, $\ulcorner(A \Rrightarrow B_1)\urcorner, \ldots, \ulcorner(A \Rrightarrow B_n)\urcorner$, $\ulcorner(B_0 \,\&\, \ldots \,\&\, B_n. \Rrightarrow C)\urcorner \in N$. Combining these restrictions we obtain the general constraint:

(3.10) If $\ulcorner(A \Rrightarrow C)\urcorner \in N$, then there cannot be finite sets $\Phi_0, \ldots, \Phi_n$ of predicates such that:
(i) for each $\varphi \in \Phi_0$, $\ulcorner(A \Rrightarrow \varphi)\urcorner \in N$;

(ii) for each $i < n$ and $\varphi \in \Phi_{i+1}$ there is a $\Gamma \subseteq \bigcup_{j \leq i} \Phi_j$ such that $\ulcorner(\Pi\Gamma \Rrightarrow \varphi)\urcorner \in N$;[7]

(iii) there is a $\Gamma \subseteq \bigcup_{i \leq n} \Phi_i$ such that either $\ulcorner(\Pi\Gamma \Rrightarrow C)\urcorner \in N$ or $\Gamma$ entails $C$.

Let $N^+$ be the set of all true projectible strong subjunctive generalizations. Then $N$ is a subset of $N^+$ which satisfies constraint 3.10 and which also satisfies the condition that if $\ulcorner(A \Rrightarrow B)\urcorner \in N^+$ then $\forall N$ entails $\ulcorner\forall(A \supset B)\urcorner$. Can we take this as defining $N$? Almost, but not quite. In general, there may be more than one subset of $N^+$ satisfying these conditions. For example, suppose $N^+$ contains the three subjunctive biconditionals $\ulcorner(A \Lleftarrow\Rrightarrow B)\urcorner$, $\ulcorner(B \Lleftarrow\Rrightarrow C)\urcorner$, and $\ulcorner(C \Lleftarrow\Rrightarrow A)\urcorner$.[8] Any one of these is derivable from the other two, so (if these are not derivable from other generalizations in $N^+$) there will be three different subsets of $N^+$ satisfying the above conditions, each one containing a different pair of these subjunctive biconditionals. Which of these subsets of $N^+$ is $N$? In fact, there is no way to choose between these subsets, because there is no basis for choosing between the three biconditionals. They all have an equal claim to be regarded as basic, and so they should all be regarded as basic. This indicates that $N$ should be the union of all the subsets of $N^+$ satisfying these two constraints:

(3.11) $N = \bigcup \{X;\ X \subseteq N^+$ and $X$ satisfies constraint 3.10 and for any generalization $\ulcorner(A \Rrightarrow B)\urcorner$ in $N^+$, $\forall X$ entails $\ulcorner\forall(A \supset B)\urcorner\}$

Given this definition of $N$, I believe that our account of derived subjunctive generalizations, and hence our definition of $\ulcorner(P \Rrightarrow Q)\urcorner$, works correctly.

Given our definitions, we can explore the logical properties of strong generalizations. We have been saying that a sentence is physically possible just in case it is consistent with all physical laws. But by 3.9, this is the same thing as being consistent with $\forall N$. Thus:

(3.12) $\ulcorner\underset{p}{\Diamond} P\urcorner$ is true iff $P$ is consistent with $\forall N$.

Let us also define, in case $P$ is an open formula:

(3.13) $\ulcorner\underset{p}{\Diamond} P\urcorner$ is true iff $\ulcorner\underset{p}{\Diamond}(\exists x_1) \ldots (\exists x_n) P\urcorner$ is true.

We define in the conventional way:

(3.14) $\ulcorner\underset{p}{\Box}P\urcorner$ is true iff $\forall N$ entails $P$.

Our definitions of physical possibility and physical necessity make reference to $N$. It would be nice to be able to define these modalities simply in terms of strong generalizations without referring to $N$. It follows from 3.9 that most of the natural ways of defining modalities in terms of conditionals yield logical necessity rather than physical necessity:

(3.15) $(P \Rrightarrow .Q \;\&\; {\sim}Q) \equiv \Box{\sim}P.$

(3.16) $(P \Rrightarrow {\sim}P) \equiv \Box{\sim}P.$

However, there is one normal way of defining a modality which does give us physical necessity:

(3.17) $(Q \vee {\sim}Q. \Rrightarrow P) \equiv \underset{p}{\Box}P.$[9]

We can list a number of theorems that result from our definition of '$\Rrightarrow$' and the physical modalities:

(3.18) $(P \Rrightarrow Q) \;\&\; (Q \rightarrow R). \supset (P \Rrightarrow R).$

(3.19) $(P \rightarrow Q) \supset (P \Rrightarrow Q).$

(3.20) $(P \Rrightarrow Q) \;\&\; (P \Rrightarrow R). \supset [P \Rrightarrow (Q \;\&\; R)].$

(3.21) $(P \leftrightarrow Q) \;\&\; (Q \Rrightarrow R). \supset (P \Rrightarrow R).$

(3.22) $(P \Rrightarrow R) \;\&\; (Q \Rrightarrow R). \supset [(P \vee Q) \Rrightarrow R].$

(3.23) $(P \;\&\; {\sim}Q. \Rrightarrow R) \supset [P \Rrightarrow (Q \vee R)].$

(3.24) $(P \Rrightarrow Q) \;\&\; (P \Rrightarrow R). \supset [(P \;\&\; R) \Rrightarrow Q].$

(3.25) $(P \Rrightarrow Q) \;\&\; [(P \;\&\; Q) \Rrightarrow R]. \supset (P \Rrightarrow R).$

Certain normal principles hold only with the additional assumption that antecedents of generalizations are physically possible:

(3.26) $\underset{p}{\Diamond}P \;\&\; (P \Rrightarrow Q) \;\&\; (Q \Rrightarrow R). \supset (P \Rrightarrow R).$

(3.27) $\underset{p}{\Diamond}{\sim}Q \;\&\; (P \Rrightarrow Q). \supset ({\sim}Q \Rrightarrow {\sim}P).$

Now that we have an account of the way in which strong subjunctive generalizations work, we can return to the task of characterizing them in terms of physical necessity and quantification. A natural proposal was that $\ulcorner(A \Rrightarrow B)\urcorner$ is equivalent to $\ulcorner\underset{p}{\Box}\forall(A \supset B)\urcorner$. This seemed to be correct for non-counter-legal generalizations, but it failed for counter-legals. It is now easy to see both that it is correct for non-counter-legals and why it fails for counter-legals. In the non-counter-legal case, $\forall N$ itself is the only maximal-$A$-consistent subset of $\forall N$, so 3.9 gives us the result that $\ulcorner(A \Rrightarrow B)\urcorner$ is true iff $\ulcorner\forall(A \supset B)\urcorner$ is entailed by $\forall N$, i.e., iff $\ulcorner\forall(A \supset B)\urcorner$ is physically necessary:

$$(3.28) \quad \underset{p}{\Diamond}A \supset [(A \Rrightarrow B) \equiv \underset{p}{\Box}\forall(A \supset B)].$$

But if $\ulcorner(A \Rrightarrow B)\urcorner$ is counter-legal, then what is required is not that $\ulcorner\forall(A \supset B)\urcorner$ be entailed by $\forall N$, but rather that $\ulcorner\forall(A \supset B)\urcorner$ be entailed by every maximal-$A$-consistent subset of $\forall N$. Thus $\ulcorner\underset{p}{\Box}\forall(A \supset B)\urcorner$ does not entail $\ulcorner(A \Rrightarrow B)\urcorner$. However, there is a natural way of thinking of the different maximal-$A$-consistent subsets of $\forall N$. If, contrary to fact, it were physically possible for there to be an $A$, then $N$ could not be the actual set of basic strong generalizations. Instead, the set of basic strong generalizations would have to be consistent with there being an $A$. The different maximal-$A$-consistent subsets of $\forall N$ represents the the different ways of minimally modifying $N$ in order to make it consistent with there being an $A$, and as such it is reasonable to think of them as representing the sets of basic strong generalizations in the different worlds that might be actual if it were physically possible for there to be an $A$. Then to say that every such maximal-$A$-consistent subset of $\forall N$ entails $\ulcorner\forall(A \supset B)\urcorner$ is the same as saying that $\ulcorner\forall(A \supset B)\urcorner$ is physically necessary in every world that might be actual if it were physically possible for there to be an $A$, i.e.,

$$(3.29) \quad (A \Rrightarrow B) \equiv [\underset{p}{\Diamond}A > \underset{p}{\Box}\forall(A \supset B)].$$

The above reasoning is, I think, persuasive, but we cannot assert 3.29 as a theorem because we do not yet have an analysis of subjunctive conditionals. However, when we finally do get a complete analysis of

subjunctive conditionals in Chapter VI, we will be able to prove 3.29. Thus this is a correct characterization of strong subjunctive generalizations. Of course, it is not an analysis because it employs the notions of physical possibility, physical necessity, and the simple subjunctive, all three of which are ultimately defined in terms of strong subjunctive generalizations.

## 4. Weak generalizations

Now let us turn to weak generalizations. Their characterization is rather obvious given our treatment of strong generalizations. Once again, we begin with a set $W$ of basic weak generalizations that have been confirmed directly by induction, and then we derive new generalizations from those in $W$. The set $W$ is constructed from the set $W^+$ of true projectible weak subjunctive generalizations in just the way $N$ was constructed from $N^+$. If we had no strong generalizations to contend with, we could define $\ulcorner(P \Rightarrow Q)\urcorner$ in a way completely analogous to principle 3.9:

> $\ulcorner(P \Rightarrow Q)\urcorner$ is true iff every maximal-$P$-consistent subset of $W$ entails $\ulcorner\forall(P \supset Q)\urcorner$.

However, our characterization is made more complicated by the presence of strong generalizations. The difficulty is that it is not sufficient to render $\forall N$ and $\forall W$ each consistent with $P$ individually. We must render them *jointly* consistent with $P$. For example, if $\forall N$ contains the single generalization $\ulcorner\forall(P \supset R)\urcorner$, and $\forall W$ contains the single generalization $\ulcorner\forall(Q \supset \sim R)\urcorner$, then both $\forall N$ and $\forall W$ are consistent with $\ulcorner(P \,\&\, Q)\urcorner$. But $\forall(N \cup W)$ is not consistent with $\ulcorner(P \,\&\, Q)\urcorner$. If we ignored this inconsistency, we would be able to infer both $\ulcorner(P \,\&\, Q. \Rightarrow R)\urcorner$ and $\ulcorner(P \,\&\, Q. \Rightarrow \sim R)\urcorner$. This should obviously be prohibited. However, it is not sufficient to just require that $\forall(N \cup W)$ be consistent with $P$. Laws take precedence over weak generalizations. For example, it is a law that any object released in a vacuum at the earth's surface would fall towards the center of the earth. We also have the weak generalization that any helium filled balloon released at the surface of the earth would rise. Given these two generalizations, we can conclude that any helium

filled balloon released in a vacuum at the surface of the earth would fall towards the center of the earth; and we cannot conclude that such a balloon would rise. When a strong generalization conflicts with a weak generalization, we make our inferences on the basis of the strong generalization.

To say that laws take precedence over weak generalizations is to say that in rendering $\forall(N \cup W)$ $P$-consistent through deletion, we first render $\forall N$ $P$-consistent by making as few deletions as possible, and then we delete whatever we must from $\forall W$ to render $\forall(N \cup W)$ $P$-consistent. Or, more precisely, we first find a maximal-$P$-consistent subset $N_P$ of $\forall N$, and then we find a maximal-$P$-consistent subset $W_P$ of $\forall(N \cup W)$, *subject to the restriction that* $N_P \subseteq W_P$. This yields the following analysis:

(4.1) $\ulcorner(P \Rightarrow Q)\urcorner$ is true iff for every maximal-$P$-consistent subset $N_P$ of $\forall N$, and every maximal-$P$-consistent subset $W_P$ of $\forall(N \cup W)$ such that $N_P \subseteq W_P$, $W_P$ entails $\ulcorner\forall(P \supset Q)\urcorner$.

Having explicated '$\Rightarrow$', it seems that we can now define 'actual possibility' in a manner completely analogous to the definition of physical possibility:

(4.2) $\ulcorner\underset{a}{\Diamond}P\urcorner$ is true iff $P$ is consistent with $\forall(N \cup W)$.

If $P$ is an open formula:

(4.3) $\ulcorner\underset{a}{\Diamond}P\urcorner$ is true iff $\ulcorner\underset{a}{\Diamond}(\exists x_1) \ldots (\exists x_n)P\urcorner$ is true.

'Actual necessity' is defined in the normal way:

(4.4) $\ulcorner\underset{a}{\Box}P\urcorner$ is true iff $\ulcorner\sim\underset{a}{\Diamond}\sim P\urcorner$ is true.

We can easily obtain a number of theorems about weak generalizations, most of them analogous to theorems about strong generalizations:

(4.5) $\underset{a}{\Box}P \equiv (Q \vee \sim Q. \Rightarrow P)$.

(4.6) $(P \Rrightarrow Q) \supset (P \Rightarrow Q)$.

(4.7) $(P \Rightarrow Q) \supset \forall(P \supset Q)$.

(4.8) $(P \Rightarrow Q) \ \& \ (P \Rightarrow R). \supset [P \Rightarrow (Q \ \& \ R)]$.

(4.9) $(P \leftrightarrow Q) \ \& \ (Q \Rightarrow R). \supset (P \Rightarrow R)$.

(4.10) $(P \Rightarrow R) \ \& \ (Q \Rightarrow R). \supset [(P \vee Q) \Rightarrow R]$.

(4.11) $[(P \ \& \ \sim Q) \Rightarrow R] \supset [P \Rightarrow (Q \vee R)]$.

(4.12) $(P \Rightarrow Q) \ \& \ (P \Rightarrow R). \supset [(P \ \& \ R) \Rightarrow Q]$.

(4.13) $(P \Rightarrow Q) \ \& \ [(P \ \& \ Q) \Rightarrow R]. \supset (P \Rightarrow R)$.

(4.14) $(P \Rightarrow Q) \ \& \ (Q \rightarrow R). \supset (P \Rightarrow R)$.

(4.15) $\underset{a}{\Diamond} P \ \& \ (P \Rightarrow Q) \ \& \ (Q \Rightarrow R). \supset (P \Rightarrow R)$.

(4.16) $\underset{a}{\Diamond} \sim Q \ \& \ (P \Rightarrow Q). \supset (\sim Q \Rightarrow \sim P)$.

We have explained weak subjunctive generalizations in a way precisely analogous to the way in which we explained strong subjunctive generalizations, and we have defined actual possibility and actual necessity in a way precisely analogous to the way in which we defined the better known modalities of physical possibility and physical necessity. Despite all this, one is apt to have the lingering feeling that he does not really understand actual possibility and actual necessity. I think that the lack of understanding here is more of a psychological lack than a logical lack. Given our examples, it becomes completely obvious that there are weak subjunctive generalizations, and in fact that most of the subjunctive generalizations we affirm are weak generalizations rather than strong generalizations. Furthermore, the characterization of weak subjunctive generalizations in terms of the way in which they are confirmed seems to me entirely adequate, and given that, the formal definitions of actual possibility and actual necessity cannot be faulted. What is lacking is not a logical characterization of these concepts, but rather an intuitive feel for how they fit into our ordinary thinking and reasoning and how they are related to other concepts.

We can provide a bit of intuitive feel for actual possibility and actual necessity by showing that these concepts really are employed in our ordinary reasoning despite the fact that logicians appear to have completely overlooked them. We often have occasion to assert that

something might be the case, or that something else just wouldn't be the case. For example, in talking about the 9:14 train from Boston, I could assert that, on the one hand, it might show up with all the cars painted fluorescent pink (they do strange things like that on this railroad), but on the other hand, it just wouldn't show up on time. The English language puts a bit of a strain on us here. In talking about an actual 9:14 train (e.g., the one this morning), it is not entirely appropriate to use the subjunctive 'wouldn't' and say 'It just wouldn't show up on time'. Instead, we say 'It just won't show up on time'. But notice that this is intended to be stronger than a prediction. Part of the purpose of the 'just' in the sentence is to indicate that 'won't' is functioning as a modality rather than in its purely indicative sense. In saying that the train just won't show up on time, we are saying that it is in some sense necessary that it won't show up on time. In these sentences 'might', 'wouldn't', and 'won't' are functioning as singulary modal operators, and not as parts of conditionals. We use these modalities all the time, and they are precisely the operators of actual possibility and actual necessity.

Additional feel for weak subjunctive generalizations can be provided by seeing how they are related to other concepts. As in the case of strong subjunctive generalizations, it will turn out that they can be expressed in terms of actual possibility, quantification, and the simple subjunctive:

$$(4.17) \quad (P \Rightarrow Q) \equiv (\underset{a}{\Diamond} P > \underset{a}{\Box} \forall (P \supset Q)].$$

This principle should seem plausible for the same reasons principle 3.28 did, and we will be able to prove it in Chapter VI.

But perhaps 4.17 is not very helpful, because it expresses '$\Rightarrow$' in terms of actual possibility, which if anything is a less familiar concept than '$\Rightarrow$' itself. A much more helpful characterization of weak subjunctive generalizations can be obtained in another way. Weak subjunctive generalizations are true *because of* physically contingent facts about the world. The 9:14 train from Boston is always late *because* the railroad has only old broken-down equipment; anyone who drank from Chisholm's bottle would be poisoned *because* the bottle contains rat poison. In general, a weak generalization $\ulcorner(Ax \Rightarrow Bx)\urcorner$ is true only

when there is a true statement $P$ such that $\ulcorner(Ax \ \& \ P. \Rrightarrow Bx)\urcorner$ is true. However, it is quite clear that $P$'s merely being true is not enough to ensure the truth of the weak generalization. What other constraints are required? In the case of Chisholm's bottle, $P$ is the statement that the bottle contains rat poison. It is at least required that $P$ would still be true even if someone were to drink from the bottle:

(4.18) $(Ax \Rightarrow Bx) \supset (\exists P)[(Ax \ \& \ P. \Rrightarrow Bx) \ \& \ PE(\exists x)Ax]$.

However, 4.18 cannot be turned into a biconditional. The requirement $\ulcorner PE(\exists x)Ax\urcorner$ is not yet strong enough to ensure the truth of $\ulcorner(Ax \Rightarrow Bx)\urcorner$. This is because if the bottle does contain poison and someone has in fact drunk from it, then $\ulcorner PE(\exists x)Ax\urcorner$ is automatically true. This follows from our analysis of 'even if' in Chapter II. But this is not enough to make $\ulcorner(Ax \Rightarrow Bx)\urcorner$ true. We must require that $P$ would still be true even if someone else drank from the bottle. It seems that the general requirement should be that $P$ would still be true even if there were a new $A$, i.e., an $A$ which does not now exist:

(4.19) $(Ax \Rightarrow Bx) \equiv (\exists P)[(Ax \ \& \ P. \Rrightarrow Bx)$ & $P$ would be true even if there were an $A$ which does not now exist].

How can we make the final clause of 4.19 precise? The only way I can see to do this is by talking about sets of objects. There are no generally accepted conventions regarding the behavior of sets across possible worlds, so we are free to make up our own. It seems reasonable to regard a set in one world as the same set as a set in another world just in case they contain the same objects. In other words, we identify sets across possible worlds in terms of their members, just as we identify sets within a world. This will be our convention. It will be discussed in more detail in Chapter VI. Then to talk about there being an $A$ which does not now exist is to talk about there being an $A$ which is not a member of the set of all objects existing in this world:

(4.20) $(Ax \Rightarrow Bx) \equiv (\exists X)(\exists P)[(y)(y \in X \equiv y = y)$ &
$(Ax \ \& \ P. \Rrightarrow Bx) \ \& \ PE(\exists x)(Ax \ \& \ x \notin X)]$.

This seems to me to capture exactly what we mean by the weak subjunctive generalization. Given our full analysis of subjunctive conditionals in Chapter VI, 4.20 will be equivalent to the simpler characterization:

(4.21) $(Ax \Rightarrow Bx) \equiv (\exists X)[(y)(y \in X \equiv y = y)\ \&$
$(x)(Ax \supset Bx)E(\exists x)(Ax\ \&\ x \notin X)].$

This also seems to be a very intuitive characterization of the weak subjunctive generalizations. $\ulcorner(Ax \Rightarrow Bx)\urcorner$ certainly entails $\ulcorner(x)(Ax \supset Bx)\urcorner$ and it is plausible to suppose that what more is required for the truth of $\ulcorner(Ax \Rightarrow Bx)\urcorner$ is that if there were a new $A$ (one that does not now exist), it would be a $B$ too. This is just what 4.21 requires.

However, if 4.21 is to be acceptable, we must somehow ensure that a certain kind of counter-example cannot occur. Let $X$ be the set of all the objects existing in this world. If the generalization $\ulcorner(Ax \Rightarrow x \in X)\urcorner$ were true and logically contingent (i.e., $\ulcorner(x)(Ax \rightarrow x \in X)\urcorner$ is false), this would constitute an immediate counter-example to 4.21. Can a subjunctive generalization of this form ever be true? In order for it to be true, our generalizations would somehow have to dictate that there could not be any $A$'s that do not already exist. This would be very odd. Certainly our generalizations might dictate the total number of $A$'s there are in the universe (in particular, a physical law might dictate that there is just one $A$), and hence they might preclude there being *additional* $A$'s, but this is not what $\ulcorner(\exists x)(Ax\ \&\ x \notin X)\urcorner$ requires. The latter does not require that there be more $A$'s than there are now, but just that there be a *different* $A$.

In fact, I think it is logically impossible to have a true subjunctive generalization of the form $\ulcorner(Ax \Rightarrow x \in Y)\urcorner$ for any set $Y$ (except in the trivial case where $\ulcorner Ax\urcorner$ entails $\ulcorner x \in Y\urcorner$). This results from the fact that subjunctive generalizations must either be projectible or else entailed by projectible generalizations. In order for any generalization of the form $\ulcorner(Ax \Rightarrow x \in Y)\urcorner$ to be true, some generalization of that form would have to be projectible. But it is quite obvious that no such generalization could be projectible. If $\ulcorner x \in Y\urcorner$ were a projectible predicate, then we would find ourselves immediately confirming, for every projectible predicate $\ulcorner Ax\urcorner$, that nothing could be an $A$ other than what is already an $A$. Clearly, such conclusions are not automatically confirmed, so $\ulcorner x \in Y\urcorner$ cannot be a projectible predicate. Consequently, no contingent generalization of the form $\ulcorner(Ax \Rightarrow x \in Y)\urcorner$ can ever be true, and by virtue of principle 4.17, this implies the validity of the following principle:

(4.22) $\underset{a}{\Box}(x)(Ax \supset x \in X)\ \&\ \underset{a}{\Diamond}(\exists x)Ax. \supset (x)(Ax \rightarrow x \in X).$

For the same reason, we should have

(4.23) $\underset{a}{\Box}(x)(Ax \supset .x = a_1 \vee \ldots \vee x = a_n)\ \&\ \underset{a}{\Diamond}(\exists x)Ax. \supset$

$\Box(x)(Ax \supset .x = a_1 \vee \ldots \vee x = a_n).$

Given these principles (whose validity will be built into the semantics of Chapter VI), it will become possible to prove the correctness of 4.20 and 4.21. Thus the latter principles do constitute a correct characterization of weak subjunctive generalizations, and derivatively of actual possibility and actual necessity. Of course, principles 4.20 and 4.21 cannot be regarded as providing analyses of these concepts, because 4.20 and 4.21 employ subjunctive conditionals which are themselves analyzed in terms of weak subjunctive generalizations. However, the purpose of 4.20 and 4.21 was not to provide an analysis anyway, but to provide more of an intuitive feel for the way weak subjunctive generalizations are related to other concepts. We already have a logically adequate analysis of weak subjunctive generalizations in terms of the way they are confirmed.

## 5. Conclusions

In this chapter I have proposed what I believe to be a logically adequate analysis of both strong and weak subjunctive generalizations in terms of non-subjunctive statements. This should free us to use the notion of a subjunctive generalization in the analysis of subjunctive conditionals without feat of circularity. We have, in fact, got a tentative solution to one of the two major problems facing the traditional linguistic theory of subjunctive conditionals.

## NOTES

[1] This is suggested by Stalnaker (1968).

[2] See Pollock (1974) for a precise definition of projectibility and an argument to the effect that most subjunctive generalizations fail to be projectible.

[3] A complete defense of the position that an account of the justification conditions of a concept constitutes an analysis of that concept is provided by Pollock (1974).

[4] A fuller discussion of this can be found in Pollock (1974).
[5] As will be seen in Chapters IV and VI, this analysis must be made a bit more complicated. However, the present version of the analysis is sufficient for the time being.
[6] A defense of this can be found in Pollock (1974).
[7] $\ulcorner\Pi\Gamma\urcorner$ is the conjunction of $\Gamma$.
[8] $\ulcorner A \Leftrightarrow B\urcorner$ is defined to be $\ulcorner(A \Rrightarrow B)\ \&\ (B \Rrightarrow A)\urcorner$.
[9] I am indebted to Rolf Eberle for this observation.

CHAPTER IV

# THE BASIC ANALYSIS OF SUBJUNCTIVE CONDITIONALS

## 1. Introduction

Having laid the groundwork, we can now attempt to construct an analysis of subjunctive conditionals. The basic tool for this analysis is provided by Theorem 3.11 of Chapter I. According to that theorem, a subjunctive conditional $\ulcorner(P > Q)\urcorner$ is true iff $Q$ is true in every possible world that might be actual if $P$ were true. That is, assuming the Generalized Consequence Principle, we have:

(1.1) $\ulcorner(P > Q)\urcorner$ is true in the actual world iff for every possible world $\alpha$, if $\alpha \mathbf{M} P$ then $Q$ is true in $\alpha$; $\ulcorner QMP\urcorner$ is true iff for some $\alpha$ such that $\alpha \mathbf{M} P$, $Q$ is true in $\alpha$

This is not yet a philosophically satisfactory definition of the simple subjunctive, because the relation **M** was defined in terms of '>', but if we can provide an alternative analysis of **M**, principle 1.1 will constitute an analysis of '>'. That will be our strategy here. Let us say that $\alpha$ is a *P-world* when $\alpha \mathbf{M} P$.[1] Thus our task is to analyze the notion of a $P$-world.

In this chapter we will restrict our attention to subjunctive conditionals whose antecedents and consequents are indicative, and we will identify a possible world with the set of its indicative truths. In the next two chapters we will generalize the analysis to handle non-indicative antecedents and consequents.

## 2. The analysis of **M**

The basic idea underlying my analysis of **M** was introduced in Chapter I. This is that a $P$-world is one that is obtained from the real world by making minimal changes which suffice to make $P$ true. In constructing $P$-worlds, we make changes in the real world only insofar as we are

forced to do so in order to accommodate $P$'s being true. Truths of the actual world that are in the appropriate sense 'irrelevant' to $P$ must also be truths of any $P$-world. Gratuitous changes are disallowed. The problem is to say just what constitutes a non-gratuitous change.

Let us begin with the case in which $P$ is an indicative statement which is consistent with all true subjunctive generalizations. In this case, it seems clear that any $P$-world must preserve all of those subjunctive generalizations. If $P$ is, for example, 'I dropped this piece of chalk five minutes ago', then if in constructing a world, we alter the law of gravity instead of concluding that the chalk would have fallen to the ground, that would be a gratuitous change. Such a world would not be a $P$-world. Subjunctive generalizations take precedence over other truths in the construction of $P$-worlds. Let $G$ be the set of universally quantified material conditionals corresponding to those subjunctive generalizations that are true in the actual world. Then if $P$ is consistent with $G$ and $\alpha$ is a $P$-world, we must have $G \subseteq \alpha$.

Most changes are ruled out as gratuitous. What changes are not gratuitous? The most obvious case occurs when $Q$ is true in the actual world, but $G$ entails $\ulcorner(P \supset \sim Q)\urcorner$. Then we *must* make $Q$ false in any $P$-world. On our assumption that $P$ is consistent with $G$ (i.e., $P$ is 'actually possible'), it follows that $G$ entails $\ulcorner(P \supset R)\urcorner$ iff $P$ weakly implicates $R$, i.e., iff $\ulcorner(P \Rightarrow R)\urcorner$ is true. Let us say that $P$ *counter-implicates* $R$ when $\ulcorner P \Rightarrow \sim R\urcorner$ is true. So if $P$ counter-implicates $Q$, $Q$ must be false in any $P$-world.

But this cannot be the only time that a change is allowed. For example, we may have two true propositions $Q$ and $R$ such that the conjunction $\ulcorner(Q \,\&\, R)\urcorner$ is counter-implicated by $P$, but neither $Q$ nor $R$ by itself is counter-implicated. We cannot have both $Q$ and $R$ true in a $P$-world, so some change is required. For example, let $P$ be 'Bizet and Verdi were compatriots', and let $Q$ be 'Verdi was Italian' and $R$ be 'Bizet was French'. Then $\ulcorner P \Rightarrow \sim(Q \,\&\, R)\urcorner$ is true, but $P$ does not counter-implicate either $Q$ or $R$. How do we decide whether to change the truth value of $Q$ or that of $R$? There is no way to decide. There is no basis for giving preference to one over the other. The situation is rather that *either* $Q$ or $R$ *might* be false – there is a $P$-world in which $Q$ is false (but $R$ true), and another in which $R$ is false (but $Q$ true) – but we cannot conclude of either $Q$ or $R$ that it *would* be false. If

Bizet and Verdi were compatriots, they might both be French, and they might both be Italian, but it is not true that they *would* both be French, or that they *would* both be Italian.

This suggests that in constructing $P$-worlds, we simply take the set of actual truths and then make minimal changes so as to render it $P$-consistent. The result of any such set of minimal changes describes a $P$-world. As there may be more than one way of minimally changing the actual world in order to achieve $P$-consistency, there may be more than one $P$-world. Making this precise, if $\alpha_0$ is the set of actual truths (i.e., $\alpha_0$ is the real world), and $\alpha$ is a possible world, the proposal is:

(2.1) ANALYSIS I: $\alpha \mathbf{M} P$ iff $G \subseteq \alpha$ and $\alpha \cap \alpha_0$ is a maximal $P$-consistent subset of $\alpha_0$.

In other words, it is required that $\alpha$ preserve as many of the actual truths as possible while still making $P$ true.

Although I believe that Analysis I is on the right track, there are immediate difficulties for it. As it stands, it would lead to the result that if $P$ is false then absolutely no truth is preserved in every $P$-world: that is, given any truth $Q$, there would be a $P$-world in which $Q$ is false. This is because if $P$ is false, then both $Q$ and $\ulcorner(P \supset \sim Q)\urcorner$ are true. Thus in making minimal changes to $\alpha_0$ so as to accommodate the truth of $P$, we have a choice between preserving $Q$ and preserving $\ulcorner(P \supset \sim Q)\urcorner$, and in the latter case we will be forced to make $Q$ false. For example, if $P$ is 'My maple tree died' and $Q$ is 'My car is painted white', $Q$ should be preserved in every $P$-world. But according to Analysis I, we have a choice between preserving 'If my maple tree died, my car is not painted white' and 'My car is painted white', and in those $P$-worlds in which we preserve the former, it is not true that my car is painted white.

Evidently not *all* truths are equally good candidates for being preserved in $P$-worlds. In the above example, $Q$ is a candidate for preservation, but $\ulcorner(P \supset \sim Q)\urcorner$ is not a candidate. Let us call those propositions which are candidates for preservation *stable* propositions. It appears that there is a class $S$ of stable propositions, and in deciding whether a change is minimal or gratuitous, we only look at what happens to the members of $S$. In making minimal changes, we seek to make minimal changes in the stable propositions, and we ignore what

happens to other propositions. This leads to:

(2.2) ANALYSIS II: If $S$ is the class of all stable propositions, then $\alpha \mathbf{M} P$ iff $P \in \alpha$ and $G \subseteq \alpha$ and $S \cap \alpha \cap \alpha_0$ is a maximal-$P$-consistent subset of $S \cap \alpha_0$.

We have seen that not all propositions are stable. Which ones are? As we have just seen, conditionals, and hence disjunctions, generally fail to be stable. On the other hand, a conjunction of stable propositions is automatically stable: If $Q$ and $R$ are both stable, they will both be preserved (and hence their conjunction will be preserved) in any $P$-world unless there is some set $\Lambda$ of stable propositions true in the actual world whose conjunction with $P$ is counter-implicated by either $Q$ or $R$; but then the conjunction of $\Lambda$ with $P$ would also be counter-implicated by $\ulcorner(Q \,\&\, R)\urcorner$. Thus a conjunction of stable propositions is automatically preserved unless it comes into conflict with some other true stable propositions, which is to say that the conjunction is stable.

The fact that conjunctions of stable propositions are stable provides some insight into just which propositions are stable. Conjunctions of stable propositions are stable, but only derivatively so *because* their conjuncts are stable. Those conjuncts might themselves be conjunctions which are stable because their conjuncts are stable, and so on. However, this cannot go on indefinitely. Eventually we must reach conjuncts that are stable in their own right. These conjuncts are still (as a matter of logic) equivalent to conjunctions,[2] but those conjunctions are conjunctions of disjunctions, and hence there is no reason to expect their conjuncts to be stable.

In examining stable propositions, we may in the above manner be able to take them apart into simpler and simpler components from which they inherit their stability, but eventually this can no longer be done and at that point the stability is basic. The stable propositions which are in this way basic are in a certain sense 'simple'. They are those which cannot, except artificially, be regarded as conjunctions, disjunctions, or in general as compounds or logical constructions out of simpler propositions.

These simple propositions would seem to be propositions ascribing 'simple states' to objects. These are non-contrived and not explicitly

compound states like colors, dispositions, shapes, etc. Other propositions, e.g., those describing events, are not simple. For example, an event always involves a change of state, and hence is compound (saying that the state is one way at one time and another way at another time).

I feel that these notions of a simple proposition and a simple state make good intuitive sense. However, any endorsement of simple propositions is bound to meet with suspicion in light of the recent history of philosophy. The logical atomists built outlandish theories around simple propositions, and the notion of a simple proposition thereby acquired a stigma. However, this is guilt by association. Just because logical atomism was a bad theory which made use of the notion of a simple proposition is no reason to think that there is something wrong with simple propositions themselves. Talk about simple propositions is just a matter of taking logical form seriously. We *do* distinguish, for example, between those propositions which really are conjunctions, and those which are only equivalent to conjunctions. The proposition that my car is white is not a conjunction, although it is, as a matter of logic, equivalent to a conjunction.

All of this is going to be repugnant to many recent philosophers who have supposed that logical form really does not make sense as applied to propositions. They have supposed that although sentences have a structure, with the result that one can be contained in another, one can be a conjunction while another is only equivalent to a conjunction, and so forth, propositions have no such structure. This comes from looking only at the properties of propositions which can be constructed out of the purely logical relations of entailment and equivalence. Using only these relations, there is no way to order propositions in terms of logical complexity, and no sense can be made out of a putative difference between, for example, a proposition being equivalent to a conjunction and really being a conjunction. But propositions do have a structure of a different sort – they have an epistemological structure. Certain propositions are epistemologically basic. These basic propositions provide grounds for believing others, which in turn provide grounds for believing still others, and so on. There is a kind of natural epistemological order here.[3] It is on this basis that we can make sense of simple propositions. We must not make the mistake of supposing that they are to be identified with the epistemologically basic ones. Rather, they are

those which are not logically complex in the sense of being conjunctions, disjunctions, etc. The notion of a proposition being a conjunction or a disjunction is reflected in (and, I would propose, reducible to) what possible grounds one can have for believing it. For example, what distinguishes a conjunction is that the two conjuncts provide a conclusive reason for believing the conjunction, and the denial of either conjunct provides a conclusive reason for rejecting the conjunction. One might suppose that this is also true of a proposition which is merely equivalent to a conjunction, but that would be a mistake. If $P$ is merely equivalent to $\ulcorner Q \,\&\, R \urcorner$, then $\ulcorner \sim Q \urcorner$ does not by itself constitute a reason for rejecting $P$. One must also have reason to believe that $P$ *is* equivalent to $\ulcorner Q \,\&\, R \urcorner$. If one does not know that such an equivalence holds, then clearly, knowing things about $Q$ and $R$ gives him no reason for believing or disbelieving $P$. I think that by appealing to epistemological considerations of this sort, it will prove possible to give a precise definition of the notion of a simple proposition. This will be undertaken in section three. However, given the epistemological way of thinking of logical form, I think it must be admitted that even without a precise definition, the notion of a simple proposition makes sense and that we do talk about the logical forms of propositions despite our not having an adequate theory of what logical form is all about. Thus for now I will make free use of the notion of a simple proposition without attempting further clarification.

Stable propositions form a broader class than just the simple propositions. For example, conjunctions of simple propositions are stable. However, this results directly from the notion of stability and does not have to be built into our analysis. Can we then just replace the class of stable propositions in Analysis II by the class of simple propositions? Almost, but not quite. Although this does not follow from the definition of stability, it is rather obvious that what we may call the 'internal negations' of simple propositions are also stable. In constructing $P$-worlds, we do not just try to preserve those simple states that an object has – we attempt to preserve *whether or not* it has a given simple state. For example, if an object is insoluble, this is something that would be preserved just as much as its solubility were it soluble. An unnecessary change in either direction would be judged gratuitous and disallowed in the construction of $P$-worlds. The proposition '$x$ is insoluble' is not

what would ordinarily be called the negation of '$x$ is soluble'. This is because both '$x$ is soluble' and '$x$ is insoluble' entail the existence of $x$. Logicians have traditionally distinguished here between 'internal' and 'external' negation. The external negation is ordinary negation. The internal negation of a proposition of the form $\ulcorner x$ is $F\urcorner$ can be defined in terms of external negation as $\ulcorner x$ exists but $\sim Fx\urcorner$. Thus to preserve whether or not an object has a particular simple state is to preserve the truth value of the corresponding simple proposition and its internal negation.

Apparently, then, what we seek to preserve are those truths in the class consisting of all simple propositions and internal negations of simple propositions. Thus we are led to:

(2.3) ANALYSIS III: If $S$ is the class of all simple propositions and their internal negations, then $\alpha \mathbf{M} P$ iff $P \in \alpha$ and $G \subseteq \alpha$, and $S \cap \alpha \cap \alpha_0$ is a maximal-$P$-consistent subset of $S \cap \alpha_0$.

However, a difficulty arises for Analysis III. The only changes countenanced by this analysis consist of deleting truths from $\alpha$ (replacing them by their negations). However, a subjunctive hypothesis may force us to add new propositions too. For example, suppose $\ulcorner(\exists x)Fx\urcorner$ is false in $\alpha_0$. Taking this as our subjunctive hypothesis may force us to introduce a new object not existing in $\alpha_0$. This may constitute a smaller change than making one of the objects already in $\alpha_0$ an $F$ when it is not now an $F$. Then if $\alpha \mathbf{M} P$, $\alpha$ will contain simple propositions or the internal negations of simple propositions which were not true in $\alpha_0$. Thus the total change in going from $\alpha_0$ to $\alpha$ may consist of the deletion of some members of $\alpha_0$ and the addition of some new propositions. This change has two parts: the deletion of some propositions in $\alpha_0$, represented by $(\alpha_0 - \alpha)$; and the addition of some new propositions, represented by $(\alpha - \alpha_0)$. It will be algebraically convenient to represent this change as a set of propositions indexed by 0 and 1. The propositions indexed by 0 will be those deleted, and the propositions indexed by 1 will be those added. To this end, let us define:

$$(\alpha \triangle \beta) = [(\alpha - \beta) \times \{0\}] \cup [(\beta - \alpha) \times \{1\}].$$

The algebraic point of defining '$\alpha \triangle \beta$' in this way is that the inclusion of one change in another is now represented by ordinary class-inclusion.

In minimizing the change from $\alpha_0$ to $\alpha$, what we want to minimize is $[(S \cap \alpha_0) \triangle (S \cap \alpha)]$. Letting $S_\alpha = S \cap \alpha$, we are led to:

(2.4) ANALYSIS IV: If $S$ is the class of all simple propositions and their internal negations, then $\alpha \mathbf{M} P$ iff $P \in \alpha$ and $G \subseteq \alpha$, and there is no world $\beta$ such that $P \in \beta$ and $G \subseteq \beta$ and $(S_\beta \triangle S_{\alpha_0}) \subset (S_\alpha \triangle S_{\alpha_0})$.

Analysis IV still suffers from some rather grave shortcomings. Thus far I have been pretending that when $\ulcorner P \Rightarrow \sim(Q \,\&\, R)\urcorner$ is true, but neither $\ulcorner P \Rightarrow \sim Q\urcorner$ nor $\ulcorner P \Rightarrow \sim R\urcorner$ is true, then there is no basis for choosing between $Q$ and $R$, and hence $\ulcorner(\sim Q)MP\urcorner$ and $\ulcorner(\sim R)MP\urcorner$ are both true. But this is not always the case. For example, suppose for the sake of simplicity that dry matches always light when struck. Then consider a particular match I had five minutes ago which was dry ($D$) and in normal surroundings. If I had struck it ($S$), it would have lit ($L$). But this cannot be accommodated on our present account of subjunctive conditionals. We have $\ulcorner(S \,\&\, D) \Rightarrow L\urcorner$ true, and hence $\ulcorner S \Rightarrow \sim(D \,\&\, \sim L)\urcorner$ is true. But neither $\ulcorner S \Rightarrow \sim D\urcorner$ nor $\ulcorner S \Rightarrow \sim\sim L\urcorner$ is true. Thus our account would lead us to conclude that the match might not have been dry if it had been struck ($\ulcorner(\sim D)MS\urcorner$ is true), which is clearly mistaken. On the contrary, we would actually conclude that the match would still have been dry even if it were struck, and hence that $\ulcorner S > L\urcorner$ is true. What is our basis for concluding this?

We are looking for a basis for choosing between two propositions $Q$ and $R$ when their conjunction is counter-implicated by $P$ but neither proposition individually is counter-implicated by $P$. When do we preserve $Q$ in preference to $R$? Sometimes the answer is, 'When part of the reason $R$ is now true is that $P$ is false'. To illustrate, consider the match case again. We preserve $D$ ('The match is dry') in preference to $\ulcorner\sim L\urcorner$ ('The match did not light'), because part of the reason the match did not light is that it was not struck, but this is not part of the reason it was dry. More precisely, there was a chain of antecedent circumstances implicating that the match was dry, and another chain implicating that it did not light. These chains of circumstances represent the way in which it actually came about that the match was dry and did not light. But among the circumstances implicating the match's not lighting was the fact that it was not struck. If we remove that from the

antecedent circumstances, the remaining circumstances are no longer sufficient to implicate that the match did not light.

Consider another case. Suppose we have a metal cylinder of volume $v_0$ filled with an ideal gas (one obeying the Boyle–Charles law) under pressure $p_0$ and temperature $t_0$. If we had heated the cylinder, the pressure of the gas would have increased. We conclude this because we know the Boyle–Charles law (the pressure equals a constant times the ratio of the temperature and volume), and we believe that the volume would remain unchanged even if the cylinder were heated (let us pretend there is no thermal expansion in the metal). Why, in constructing $P$-worlds (where $P$ = 'The cylinder was heated'), do we preserve the volume of the gas at the expense of the pressure of the gas? Because part of the reason the pressure was $p_0$ is that the cylinder was not heated, but that is not part of the reason the volume was $v_0$. More precisely, the historically antecedent circumstances implicating that the pressure was $p_0$ contained in an essential way the fact that the cylinder was not heated, but this is not true of the circumstances implicating that the volume was $v_0$.

The above two examples have the characteristic that there are circumstances historically antecedent to $R$ which implicate $R$, and then others preceding those circumstances and implicating them, and so on. This chain of circumstances represents the way in which it actually came about that $R$ was true. If we trace out this chain of historically antecedent circumstances, we find that they contain $\ulcorner\sim P\urcorner$, so that by adopting the counterfactual hypothesis that $P$ is true, we 'undercut' the reason $R$ is true. Then given a conflict between one proposition $R$ which is undercut and another $Q$ which is not, we preserve $Q$ at the expense of $R$. In order for $P$ to undercut $R$, it is not necessary for the historically antecedent circumstances implicating $R$ to actually contain $\ulcorner\sim P\urcorner$; more generally, $P$ might counter-implicate some part of those circumstances. To illustrate this, consider an oak tree standing in a field. If a tree of this structure were subjected to a 200 mph wind for a period of five minutes, it would break: $(W \,\&\, S. \Rightarrow B)$. Furthermore, let us suppose that if certain meteorological conditions $M$ were to occur over terrain of this structure, there would be 200 mph winds for a period of at least five minutes: $(M \,\&\, T. \Rightarrow W)$. As $\ulcorner W \,\&\, S. \Rightarrow B\urcorner$ is true, we have $\ulcorner W \Rightarrow \sim(S \,\&\, \sim B)\urcorner$ true. $W$ does not undercut $S$, but $W$

does undercut $\ulcorner\sim B\urcorner$: the historically antecedent circumstances implicating $\ulcorner\sim B\urcorner$ include that the tree has not been subjected to certain forces, and the circumstances implicating the latter include that there have not been 200 mph winds. Thus we are led to conclude that $\ulcorner W > B\urcorner$ is true. Furthermore, as $\ulcorner M \,\&\, T. \Rightarrow W\urcorner$ is true, we have that $\ulcorner(M \,\&\, T) \Rightarrow \sim(S \,\&\, \sim B)\urcorner$ is true. $\ulcorner M \,\&\, T\urcorner$ does not undercut $S$, but it does undercut $\ulcorner\sim B\urcorner$ because it implicates $W$ which undercuts $\ulcorner\sim B\urcorner$. Thus we should conclude that $\ulcorner M \,\&\, T. > B\urcorner$ is true, and sure enough, that is precisely what we would conclude.

As a further illustration of my diagnosis of how these conditionals arise out of undercutting, it is worth noting that in the above example we would actually go further and draw the stronger conclusion that $\ulcorner M > B\urcorner$ is true. Intuitively, this is because we are confident that the terrain would retain its structure even if the meteorological conditions were to occur. Formally, this results once more from undercutting. As we have both $\ulcorner W \,\&\, S. \Rightarrow B\urcorner$ and $\ulcorner M \,\&\, T. \Rightarrow W\urcorner$, we have $\ulcorner M \,\&\, T \,\&\, S. \Rightarrow B\urcorner$, and hence $\ulcorner M \Rightarrow \sim(T \,\&\, S \,\&\, \sim B)\urcorner$. $M$ does not undercut either $T$ or $S$: no matter how far we trace out their historical antecedents, we find nothing counter-implicated by $M$. But $M$ does undercut $\ulcorner\sim B\urcorner$, because we have seen that the historical antecedents of $\ulcorner\sim B\urcorner$ include that the tree has not been subjected to 200 mph winds, and the circumstances implicating the latter include that conditions $M$ have not obtained. So once again, we can explain the truth of the counterfactual in terms of undercutting.

Let us contrast these examples with the Bizet/Verdi case. If Bizet and Verdi had been compatriots, then Bizet might have been Italian, and Verdi might have been French. Here we do not preserve either $Q$ ('Bizet was French') or $R$ ('Verdi was Italian') at the expense of the other. This is because their not being compatriots does not undercut the reason that either man had the nationality he did. For each of $Q$ and $R$, there is a chain of historical antecedents going back indefinitely far in time which at no point is counter-implicated by $P$ (although $P$ does, of course, counter-implicate the *combined* historical antecedents of $Q$ and $R$).

I would urge that in this notion of undercutting lies the solution to understanding how counterfactuals work. But how do we capture in a precise way this notion of $P$ undercutting $R$? I believe that this can be

done by preserving simple propositions with early dates in preference to those with later dates. I argued that simple propositions are propositions ascribing 'simple states' to objects. As such, simple propositions are *dated*. That is, they ascribe a state to an object *at a time*. For a possible world $\alpha$ and time $t$, let us define $S_\alpha(t)$ to be the set of simple propositions or internal negations of simple propositions true in $\alpha$ with date no later than $t$. Then I suggest that we impose the following requirement on $\alpha$ in order to have $\alpha \mathbf{M} P$:

(2.5) REQUIREMENT OF TEMPORAL PRIORITY: If $\alpha \mathbf{M} P$, then for each time $t$, $S_\alpha(t) \triangle S_{\alpha_0}(t)$ is minimal, i.e., there is no $\beta$ such that $P \in \beta$ and $G \subseteq \beta$ and $(S_\beta(t) \triangle S_{\alpha_0}(t)) \subset (S_\alpha(t) \triangle S_{\alpha_0}(t))$.

This requirement has the effect that in deciding what is true in $\alpha$ at a time $t$, we must first decide what is true for all prior times. We cannot modify a truth with a late date $t$ and then set about modifying propositions with earlier dates just to accommodate it, because that would lead to gratuitous changes in $S_\alpha(t)$, the set of truths preceding the one in question. We can only modify $S_\alpha(t)$ in response to 'internal stresses', and not just to bring about some change later than $t$.

It is not obvious how this requirement relates to the notion of one proposition undercutting another, so let me try to establish the connection. We ordinarily suppose that states have 'historical antecedents', i.e., combinations of earlier states of objects which implicate them. This is, essentially, the traditional assumption that every event has a cause. Regardless of whether this is always true, consider two simple propositions, $Q$ and $R$ which do have historical antecedents, and whose historical antecedents have historical antecedents, and so on indefinitely. Suppose $\ulcorner P \Rightarrow \sim(Q \,\&\, R)\urcorner$ is true, but neither $\ulcorner P \Rightarrow \sim Q\urcorner$ nor $\ulcorner P \Rightarrow \sim R\urcorner$ is true. What does the requirement of temporal priority tell us to preserve. We cannot just alter one or the other of $Q$ and $R$ without also altering its historical antecedents, and the requirement of temporal priority tells us that we must take things in temporal order, so we must decide which of the historical antecedents to preserve *before* we decide which of $Q$ and $R$ to preserve. Two possible cases can arise:

First, it may happen that at no point in the past are the historical antecedents of either $Q$ or $R$ counter-implicated by $P$. This is like the

Bizet/Verdi case. In constructing a $P$-world, we have a choice between including the historical antecedents of $Q$ and those of $R$. Either choice is possible, because in either case we can construct the set of truths so that for each $t$, $(S_\alpha(t) \triangle S_{\alpha_0}(t))$ is minimal and contains the historical antecedents to that point of $Q$ or $R$ (whichever is chosen to be preserved in $\alpha$). Thus there will be $P$-worlds containing $Q$ and $P$-worlds containing $R$, and so we conclude of either that it *might* be true if $P$ were true, but not of either that it would be true to the exclusion of the other.

Second, it may happen that at some point $t$ in time, the historical antecedents of one of $Q$ or $R$ (let us suppose it is $R$) are counter-implicated by $P$. In other words, $P$ undercuts $R$. As those historical antecedents are counter-implicated, they *cannot* be included in $S_\alpha(t)$. But given that the historical antecedents of $R$ are precluded from $S_\alpha(t)$, there is nothing in $S_\alpha(t)$ with which the historical antecedents of $Q$ conflict, and hence the historical antecedents of $Q$ *must* be included in $S_\alpha(t)$ – to omit them would be a gratuitous change and is ruled out by the requirement of minimal change. The historical antecedents of $Q$ implicate $Q$, so $Q$ must be included in $\alpha$, and hence $R$ must be precluded. Thus we are led to conclude that $Q$ would be true if $P$ were true, and $R$ would be false. Thus the requirement of temporal precedence seems to correctly capture the idea that if $P$ counter-implicates a conjunction and undercuts one of the conjuncts but not the other, then the conjunct that is undercut is sacrificed and the other preserved.

This can be illustrated by returning to the match example. The historical antecedents of the match's being dry include such things as that it has not recently been rained upon or dunked in a bucket of water. These in turn have historical antecedents having to do with the location of the match, local climatic conditions, the actions of nearby agents, etc. These historical antecedents themselves have historical antecedents (presumably), but at no point in tracing out this chain of historical antecedents do we encounter anything counter-implicated by the match's being struck. On the other hand, the historical antecedents of the match's not having lit include such things as its not having been heated to a certain temperature, or exposed to certain chemicals, *or struck*. This is immediately counter-implicated by the match's being struck. Hence we conclude that the historical antecedents of the

match's being dry would still have been true even if the match had been struck (as they conflict with nothing not counter-implicated by the match's being struck); and hence that the match would still have been dry if it had been struck; and thereby we are forced to conclude that the match would have lit if it had been struck.

The requirement of temporal priority seems to give the right answer when applied to states all of which have historical antecedents. But it is not obviously a necessary truth that all states have historical antecedents, and it has actually been proposed at various times that certain kinds of states (quantum mechanical states, miracles, etc.) do not have historical antecedents. How does the requirement of temporal priority fare in connection with these states, if indeed there are such states? First, suppose that neither $Q$ nor $R$ have historical antecedents. Suppose further that the date of $Q$ is earlier than that of $R$. Then the requirement of temporal priority tells us to preserve $Q$ at the expense of $R$. This seems reasonable: if $P$ were true, then at the time $Q$ occurred, nothing would yet have happened to preclude $Q$'s being true, so to make $Q$ false would be a gratuitous change; but then once $Q$ is true, when the date of $R$ comes up, something has occurred – namely $Q$ in conjunction with $P$ – to preclude $R$'s being true. Thus it seems correct to give preference to the earlier state. On the other hand, if $Q$ and $R$ have the same date, there is no reason to give preference to either – either *might* be true – and this is just what the requirement of temporal priority tells us. Finally, let us suppose $R$ has no historical antecedents, but $Q$ does, and the historical antecedents of $Q$ extend back in time prior to the date of $R$. Once more, if $P$ were true, then prior to the date of $R$ there would be nothing to preclude the occurrence of the historical antecedents of $Q$, and hence they would still occur even if $P$ were true. But those historical antecedents implicate $Q$, and hence together with $P$ they implicate $\sim R$. Therefore, there is something to preclude $R$'s being true when its date comes up. Once again, the requirement of temporal priority gives what seems to be the correct answer.

I believe that in the requirement of temporal priority lies virtually the entire solution to the problem of relating the truth conditions of subjunctive conditionals to the notion of a minimal change. At least in the case in which $P$ is consistent with $G$, this requirement seems to be sufficient to account for all of the judgments we make regarding

subjunctive conditionals. This leads to the analysis:

(2.6) ANALYSIS V: $\alpha \mathbf{M} P$ iff $P \in \alpha$ and $G \subseteq \alpha$ and for every time $t$, there is no world $\beta$ such that $P \in \beta$ and $G \subseteq \beta$ and $(S_{\alpha_0}(t) \triangle S_\beta(t)) \subset (S_{\alpha_0}(t) \triangle S_\alpha(t))$.

There remains a surprising difficulty for Analysis V. In order to appreciate the nature of this difficulty, it must be recognized that Analysis V incorporates two distinct elements – an account of how minimal changes enter into the truth conditions of subjunctive conditionals, and an analysis of minimal change. Up to this point all of our torturous twistings and turnings have been concerned with the first element and we have just assumed that a quite simple characterization of minimal change is adequate. That assumption must now be questioned. The two elements of our analysis can be separated as follows. Where $X$ is a set of propositions and $Y$ is a set of propositions indexed by 0 and 1, let us define:

(2.7) $X + Y = (X \cup \{Q; \langle Q, 1\rangle \in Y\}) - \{Q; \langle Q, 0\rangle \in Y\}$.

Thus $X + Y$ is the result of deleting from $X$ those propositions in $Y$ indexed by 0, and adding those propositions indexed by 1. In constructing $P$-worlds, we want to make minimal changes which will result in $P$ being true along with all of the members of $G$. Thus we are interested in changes to worlds which result in certain sets of propositions (in particular, $G \cup \{P\}$) being true. So let us define:

(2.8) If $\Gamma$ is a set of propositions, a set $X$ of indexed propositions is a *$\Gamma$-change* to $\alpha_0$ at time $t$ iff there is a world $\alpha$ such that $S_\alpha(t) = S_{\alpha_0}(t) + X$, and $\Gamma \subseteq \alpha$.

For simplicity, let us also say that $X$ is a $P$-change (relative to $G$) iff $X$ is a $G \cup \{P\}$ change.

Definition 2.8 defines the notion of a $\Gamma$-change, but does not tell us what is required for a $\Gamma$-change to be minimal. However, ignoring this oversight for the moment, we can restate that part of Analysis V which relates subjunctive conditionals to minimal changes as follows:

(2.9) $\alpha \mathbf{M} P$ iff $P \in \alpha$ and $G \subseteq \alpha$ and for every time $t$, $(S_{\alpha_0}(t) \triangle S_\alpha(t))$ is a minimal $P$-change (relative to $G$) to $\alpha_0$ at time $t$.

I believe that this much of Analysis V is correct, at least for the non-counter-legal case in which $P$ is consistent with $G$.

Turning next to the notion of a minimal change, we can define:

(2.10) $X$ is a *strictly minimal* $\Gamma$-change to $\alpha_0$ at time $t$ iff there is no $\Gamma$-change $Y$ to $\alpha_0$ at time $t$ such that $Y \subset X$.

Analysis V embodies the apparently reasonable identification of minimal $\Gamma$-changes with strictly minimal $\Gamma$-changes. However, this initially reasonable identification is ultimately indefensible. The difficulty arises from the fact that there can be $P$-changes which do not contain strictly minimal $P$-changes. For example, suppose there are finitely many $F$'s in $\alpha_0$, and let $P$ be the counterfactual hypothesis that there are infinitely many $F$'s. Any $P$-change must result in our adding infinitely many $F$'s. But given any such change, there is always a smaller $P$-change – one adding one fewer $F$. Thus there are no strictly minimal $P$-changes in this case. If we identify the minimal $P$-changes with the strictly minimal $P$-changes, this leads to the conclusion that all subjunctive conditionals having $P$ as their antecedent are vacuously true. But this conclusion is clearly incorrect. It is not true, for example, that if there were infinitely many $F$'s then there would be finitely many $F$'s. Or to take a concrete example, suppose the universe contains only finitely many stars. It is certainly not true that if there were infinitely many stars in the universe then my car would be painted black. Thus there being no strictly minimal $P$-changes cannot be adequate to make vacuously true all subjunctive conditionals having $P$ as their antecedent.

If a $P$-change $X$ does not contain a strictly minimal $P$-change, then there must be an infinite sequence $X_1, X_2, \ldots, X_n, \ldots$ of progressively smaller $P$-changes such that $X_1 \subset X$ and for each $i$, $X_{i+1} \subset X_i$. Set-theoretically, these $P$-changes constitute a nest whose lower limit (the intersection of all the $P$-changes in the nest) is not itself a $P$-change. In such a case, how do we evaluate a counterfactual whose antecedent is $P$? A natural suggestion would be that $\ulcorner P > Q \urcorner$ is true iff for every such sequence there is an $i$ such that for every $j \geq i$, the $P$-change $X_j$ makes $Q$ true. This would be analogous to David Lewis' analysis which we discussed in Chapter I.[4] Unfortunately, it is subject to similar difficulties. Most important, it would lead us to affirm subjunctive

conditionals which it seems ought not to be affirmed. For example, if there are infinitely many $F$'s in the real world, then for each natural number $n$, this proposal would make true the conditional ⌜If there were only finitely many $F$'s, then there would still be more than $n$ $F$'s⌝. But this cannot be correct. For at least some natural numbers $n$ it must be true that if there were only finitely many $F$'s, then there might be $n$ $F$'s. This example also shows that the Generalized Consequence Principle would fail on the present proposal, and it still seems that that principle should be true. Thus I think that this cannot be the correct way to analyze the truth conditions of subjunctive conditionals.

When we have such an infinite descending sequence of $P$-changes not containing any strictly minimal $P$-change, then at least some of the $P$-changes in the sequence must constitute minimal changes themselves. For example, if $P$ is the proposition that there are finitely many $F$'s, then any change which consists merely of changing the set of $F$'s so as to make it finite and altering other truths minimally to accommodate this change must be considered a minimal change, although as we have seen, such a change will not be a strictly minimal change. In order to avoid the above sorts of difficulties, our conception of a minimal change must be such that every $P$-change contains a minimal $P$-change. This can be considered a criterion of adequacy for any analysis of the notion of a minimal $P$-change.

How can we construct an analysis of the notion of a minimal $P$-change which will satisfy our criterion of adequacy? Let us say that a descending sequence (a nest) of $P$-changes is *unbounded* when there is no $P$-change which is contained in every member of the sequence. Then we might be tempted to suppose that every member of any unbounded sequence of $P$-changes is minimal. This would accommodate our observation that any change which 'just' makes the set of $F$'s finite would be considered a minimal $P$-change when $P$ is the proposition that there are finitely many $F$'s. Unfortunately, this proposal errs in the direction of being too inclusive. It does not allow us to discriminate between gratuitous and non-gratuitous changes. Given any unbounded sequence of $P$-changes, we can construct a new unbounded sequence of $P$-changes by adding a new first member which results from making all sorts of gratuitous changes to the first member of the original sequence. On the present proposal, this new

first member would have to be considered minimal, but it clearly should not be.

In order to sort this out we must consider how the necessity for unbounded sequences of $P$-changes arises. Let us say that $P$ itself is unbounded (in a world $\alpha$ at time $t$) when there are unbounded sequences of $P$-changes. Unbounded propositions frequently arise as limits of sequences of progressively stronger or progressively weaker bounded propositions. For example, consider the unbounded proposition ⌜There are infinitely many $F$'s⌝. This can be regarded as the limit of the sequence of progressively stronger propositions of the form ⌜There are more than $n$ $F$'s⌝. This limit can be thought of as the infinite conjunction of the propositions in the sequence.

This can be made precise as follows. I do not really want to commit myself to there being propositions which are infinite conjunctions, but the effect of infinite conjunctions can be captured unobjectionably as follows. Let us define:

(2.11) $P \leftrightarrow \Pi\Gamma$ iff for every possible world $\alpha$, $P$ is true in $\alpha$ iff every member of $\Gamma$ is true in $\alpha$.

Intuitively, '$P \leftrightarrow \Pi\Gamma$' means that $P$ is equivalent to the conjunction of all the propositions in $\Gamma$. However, what is being defined is the entire relation '$\leftrightarrow \Pi$', and we are not employing the expression '$\Pi\Gamma$' as a term purportedly denoting an infinite conjunction.

If $P$ is unbounded in $\alpha$, then there is often a sequence $\{Q_i; i \in \omega\}$ of bounded propositions such that (i) $P \leftrightarrow \Pi\{Q_i; i \in \omega\}$, and (ii) for each $i \in \omega$, $Q_{i+1} \rightarrow Q_i$. In this kind of case, a minimal $P$-change should be one that 'just' makes all of the $Q_i$'s true. Such minimal $P$-changes can be regarded as the upper limits of sequences of minimal changes making the $Q_i$'s true for progressively larger $i$. Precisely:

(2.12) If $\{Q_i; i \in \omega\}$ is a sequence of bounded propositions such that for each $i \in \omega$, $Q_{i+1} \rightarrow Q_i$, and $P \leftrightarrow \Pi\{Q_i; i \in \omega\}$, and $\{X_i; i \in \omega\}$ is a sequence such that for each $i \in \omega$, $X_i$ is a strictly minimal $Q_i$-change and $X_i \subseteq X_{i+1}$, then $\bigcup\{X_i; i \in \omega\}$ is a minimal $P$-change.

There is a second way that a proposition can be unbounded by virtue of being the limit of a sequence of bounded propositions. We have

seen that an unbounded proposition may be the upper limit (conjunction) of a sequence of progressively stronger propositions. It can also be the lower limit (disjunction) of a sequence of progressively weaker propositions. For example, if there are infinitely many $F$'s to begin with, then ⌜There are finitely many $F$'s⌝ is unbounded and can be regarded as the lower limit of the sequence of propositions of the form ⌜There are no more than $n$ $F$'s⌝. Let us define:

(2.13) $P \leftrightarrow \Sigma\Gamma$ iff for every possible world $\alpha$, $P$ is true in $\alpha$ iff some member of $\Gamma$ is true in $\alpha$.

Then $P$ can be unbounded by virtue of there being a sequence $\{Q_i; i \bullet \omega\}$ of bounded propositions such that $P \leftrightarrow \Sigma\{Q_i; i \in \omega\}$, and for each $i \in \omega$, $Q_i \rightarrow Q_{i+1}$.

If $P$ is the limit (i.e., disjunction) of such an infinite sequence of progressively weaker bounded false propositions $Q_i$, then it seems initially that for any $i \bullet \omega$, a strictly minimal $Q_i$-change should be considered a minimal $P$-change. However, this will not work. The difficulty is that we could gratuitously strengthen the first member of the sequence by conjoining it with an unrelated proposition $R$. $P$ would still be the disjunction of the resulting sequence of propositions, but a minimal ⌜$(Q_0$ & $R)$⌝-change should not be considered a minimal $P$-change. The solution to this difficulty seems to be to turn things around. If $P \leftrightarrow \Sigma\{Q_i; i \in \omega\}$, then ⌜$\sim P$⌝$\leftrightarrow \Pi\{$⌜$\sim Q_i$⌝$; i \in \omega\}$. Then we can rule out gratuitous changes by saying that a minimal $P$-change is a $P$-change which results in a world $\alpha$ which is such that we can get back to our original world $\alpha_o$ by making a minimal ⌜$\sim P$⌝-change to $\alpha$. Precisely:

(2.14) If $\{Q_i; i \in \omega\}$ is a sequence of false bounded propositions such that for each $i \in \omega$, $Q_i \rightarrow Q_{i+1}$, and $P \leftrightarrow \Sigma\{Q_i; i \in \omega\}$, then $X$ is a minimal $P$-change (relative to $G$) to $\alpha_o$ at time $t$ iff there is a world $\alpha$ such that $S_\alpha(t) = S_{\alpha_o}(t) + X$ and $G \cup \{P\} \subseteq \alpha$, and there is a $Y$ which is a minimal ⌜$\sim P$⌝-change (relative to $G$) to $\alpha$ at time $t$ such that $S_{\alpha_o}(t) = S_\alpha(t) + Y$.

Unbounded propositions can be either upper limits (conjunctions) or lower limits (disjunctions) of sequences of progressively stronger or

weaker bounded propositions. However, we cannot stop there. We can also have propositions which are unbounded by virtue of being limits of other unbounded propositions which are themselves limits of bounded propositions. For example, if there are now finitely many $F$'s but infinitely many $G$'s, then the proposition ⌜There are infinitely many $F$'s and finitely many $G$'s⌝ is unbounded, but it is not a limit of a sequence of bounded propositions. However, it can be regarded as the upper limit of the sequence of propositions ⌜There are at least $n$ $F$'s, and there are finitely many $G$'s⌝, and the latter propositions are themselves lower limits of the sequences of propositions of the form ⌜There are at least $n$ $F$'s, and there are no more than $m$ $G$'s⌝. This indicates that we must generalize our account and make it recursive. We first define the notion of a minimal ⌜There are at least $n$ $F$'s, and there are finitely many $G$'s⌝-change as above, and then by repeating our constructions we define the notion of a minimal ⌜There are infinitely many $F$'s and finitely many $G$'s⌝-change.

But before making the above precise, we must note a further difficulty. This is that $P$ may be unbounded not because of its own logical character, but because of the limit nature of propositions which follow from the combination of $P$ with $G$. Thus we must generalize all of the above to talk about the case where the whole set $G \cup \{P\}$ is the upper or lower limit of a sequence of progressively stronger or weaker sets of propositions. To this end we first define:

(2.15) If $\Gamma$ is a set of propositions and $\Psi$ is a collection of sets of propositions, then:
- (a) $\Gamma \leftrightarrow \Sigma\Psi$ iff for every possible world $\alpha$, all of the members of $\Gamma$ are true in $\alpha$ iff all of the members of some element of $\Psi$ are true in $\alpha$;
- (b) $\Gamma \leftrightarrow \Pi\Psi$ iff for every possible world $\alpha$, all of the members of $\Gamma$ are true in $\alpha$ iff all the members of every element of $\Psi$ are true in $\alpha$.

Let us also generalize our notion of entailment to hold between sets of propositions:

(2.16) If $\Gamma$ and $\Lambda$ are sets of propositions, then $\Gamma \rightarrow \Lambda$ iff for every possible world $\alpha$, if all the members of $\Gamma$ are true in $\alpha$ then all the members of $\Lambda$ are true in $\alpha$.

We can now formulate three principles regarding minimal changes:

(2.17) If $X$ is a strictly minimal $\Gamma$-change then $X$ is a minimal $\Gamma$-change.

(2.18) If $\Psi$ is a sequence $\{\Psi_i; i \in \omega\}$ of sets of propositions, and $\Gamma \leftrightarrow \Pi\Psi$, and for each $i \in \omega$, $\Psi_{i+1} \rightarrow \Psi_i$, and $\chi$ is a sequence $\{\chi_i; i \in \omega\}$ such that for each $i \in \omega$, $\chi_i$ is a minimal $\Psi_i$-change and $\chi_i \subseteq \chi_{i+1}$, then $\bigcup \chi$ is a minimal $\Gamma$-change.

(2.19) If $\Psi$ is a sequence $\{\Psi_i; i \in \omega\}$ of sets of propositions, and $\Gamma \leftrightarrow \Sigma\Psi$, and for each $i \in \omega$, $\Psi_i \rightarrow \Psi_{i+1}$, and there is a $Y$ such that if Neg $\Gamma$ is the set of negations of members of $\Gamma$ then $Y$ is a minimal Neg $\Gamma$-change to $\alpha$ at $t$ and $S_{\alpha_o}(t) = S_\alpha(t) + Y$, then $X$ is a minimal $\Gamma$-change to $\alpha_o$ at time $t$.

My conjecture is now that these three principles completely characterize the notion of a minimal change. In other words, proceeding recursively, for each natural number $k$ we define the notion of a minimal$_k$ change as follows:

(2.20) $X$ is a minimal$_k$ $\Gamma$-change to $\alpha_o$ at time $t$ iff either:
(a) $X$ is a strictly minimal $\Gamma$-change; or
(b) there is a sequence $\Psi$ such that $\Gamma \leftrightarrow \Pi\Psi$, and for each $i \in \omega$, $\Psi_{i+1} \rightarrow \Psi_i$, and there is a sequence $\chi$ such that for each $i \in \omega$, $\chi_i \subseteq \chi_{i+1}$ and for some $j < k$, $\chi_i$ is a minimal$_j$ $\Psi_i$-change, and $X = \bigcup \chi$; or
(c) there is a sequence $\Psi$ such that $\Gamma \leftrightarrow \Sigma\Psi$ and for each $i \in \omega$, $\Psi_i \rightarrow \Psi_{i+1}$, and there is a world $\alpha$ such that $S_\alpha(t) = S_{\alpha_o}(t) + X$ and $\Gamma \subseteq \alpha$, and for some $j < k$, there is a minimal Neg $\Gamma$-change $Y$ to $\alpha$ at time $t$ such that $S_{\alpha_o}(t) = S_\alpha(t) + Y$.

Then my conjecture is:

(2.21) $X$ is a minimal $\Gamma$-change iff for some natural number $k$, $X$ is a minimal$_k$ $\Gamma$-change.

Principle 2.21 constitutes my proposed analysis of the concept of a minimal change. Several remarks must be made about this analysis.

Two undefended assumptions are built into principle 2.21 and definition 2.20. First, it is assumed that limit propositions are always limits of $\omega$-sequences rather than limits of sequences of longer transfinite length. Second, it is assumed that a minimal change is always a $\text{minimal}_k$ change for some finite $k$. The form of definition 2.20 is such that it could trivially be extended to a definition by transfinite recursion of '$\text{minimal}_\beta$' for all ordinals $\beta$. My reason for adopting these 'finiteness' assumptions is simply that appeal to examples has not convinced me of the necessity of making the analysis more complicated than it already is. However, should we become convinced of that necessity, the required amendment to definition 2.20 is of a trivial nature, and it is quite obvious how to make the amendment.

I have defended a criterion of adequacy for any analysis of the notion of a minimal change. This is that every $\Gamma$-change ought to contain a minimal $\Gamma$-change. Is this criterion satisfied by the above analysis? Unfortunately, I do not know. My only evidence on this point is the negative evidence of not having found any counter-examples to the satisfaction of the criterion of adequacy. I do not know how one *could* give a general proof that the criterion of adequacy is satisfied. This is because of the extreme generality of the question (it is about *all* propositions) and uncertainties regarding the notion of a proposition. In Chapter VI, where we restrict our attention to propositions formulable in a certain kind of formal language, the question becomes, at least in principle, more readily answerable, although even there I am unable to supply a proof of the sort desired. Nevertheless, the lack of counter-examples to the satisfaction of the criterion of adequacy leads me to conjecture that it is satisfied.

Principles 2.9 and 2.21 constitute what will be called 'Analysis VI'. This is my analysis for non-counter-legal conditionals, and I believe that it is correct at least for the restricted case it is intended to capture. However, before we can claim to have a full fledged analysis of subjunctive conditionals, we must accomplish three more things. First, our analysis turns upon the notion of a simple proposition, so we must examine that notion more carefully and, hopefully, provide an analysis of it. This will be undertaken in the next section. Second, we must extend the analysis of subjunctive conditionals to include the case in which the antecedent is inconsistent with the set of true subjunctive

generalizations. This is the case of counter-legal conditionals, and it is the topic of section four. Third, we must extend the analysis still further to encompass the case in which the antecedent and consequent of the conditional are not indicative. We want to be able to deal with conditionals whose antecedents and consequents are themselves subjunctive conditionals or subjunctive generalizations. This task will be undertaken in the next two chapters.

## 3. Simple propositions

Our analysis of subjunctive conditionals turns rather heavily upon the notion of a simple proposition. Simple propositions will also play important roles when we turn to the analysis of causes and probabilities. The important role which simple propositions play in these analyses is unfortunate. Contemporary philosophers do not like simple propositions, and even if we can succeed in giving an adequate analysis of the notion of a simple proposition, their use here is going to be repugnant to many philosophers. Perhaps, then, we should review how we got into this fix. The basic idea lying behind our analysis of subjunctive conditionals is that they have to do with making minimal changes to accommodate the truth of the counterfactual hypothesis. But upon inspection, it seems that the notion of a minimal change only makes sense given a notion of a simple proposition. In making minimal changes we cannot treat all propositions on an equal footing, because that would have the result that virtually all changes would be minimal. Given any finite set of false proposition which we want to make true, we could always form their conjunction and then regard making that conjunction true as a single change. The only way to rule this out is to somehow rule out conjunctions and disjunctions from the propositions we look at in deciding whether a change is minimal. But supposing that this can be done is just to suppose that there is a class of propositions which are not conjunctions, disjunctions, etc., i.e., it is to suppose that there is a class of simple propositions. Thus it seems that the only way to make sense of minimal changes is in terms of simple propositions. It appears we have no option but to try to make sense of simple propositions.

In fact, I think that the notion of a simple proposition makes perfectly good sense when viewed from an epistemological point of view. There is an intuitive distinction between those propositions which one can know 'in one fell swoop' and those propositions which one knows by first knowing pieces of them. More precisely, we want to rule out conjunctions, disjunctions, and in general logical compounds as not being simple. It is clearly not satisfactory to rule out all propositions *equivalent* to conjunctions, disjunctions, and the like, because that would rule out *all* propositions. We want to rule out those propositions that are, in some sense, *really* conjunctions or disjunctions. I suggest that what characterizes such propositions is that the only non-inductive way of coming to know their truth is by ascertaining the truth of their conjuncts or disjuncts and then performing the appropriate logical inference. For example, the only non-inductive way of knowing the truth of 'Either the car is white or it has four wheels' is to know the truth of one of the disjuncts.[5] But although 'The car is white' is equivalent to the disjunction 'Either the car is white and has four wheels, or the car is white and does not have four wheels', we do not have to know the truth of either of these disjuncts before we can know that the car is white.

The case of conjunctions is analogous to the case of disjunctions. The only non-inductive way of knowing the truth of the conjunction 'The car is white and has four wheels' is to know the truth of both conjuncts.

In contrast to the situation for conjunctions and disjunctions, I would propose that a simple proposition is one whose truth can be known non-inductively without first coming to know the truth of some proposition or propositions which entail it. Thus disjunctions and conjunctions are not simple, because to (non-inductively) know a disjunction you must infer it from one of its disjuncts, and to (non-inductively) know a conjunction you must infer it from its conjuncts. But a proposition like 'The car is white' would seem to be simple, because it is possible to come to know its truth by proceeding directly in ways determined by its meaning and without inferring it logically from other propositions that entail it.[6] This makes our notion of a simple proposition, at base, an epistemological notion. A simple proposition is one whose truth can be known non-inductively by proceeding

in some direct manner which is dictated by the meaning of the proposition and which does not involve deducing the proposition logically from some simpler (in the order of epistemological complexity) propositions it is possible for us to know first. Thus my proposal is:

(3.1) $P$ is simple iff it is logically possible for one to know the truth of $P$ non-inductively without first knowing the truth of each proposition in some set $\Gamma$ which entails $P$.[7]

It is unfortunate that the success of this analysis turns upon the truth of certain epistemological theories. Although I have defended those theories elsewhere, a person who does not accept them cannot be expected to accept the above analysis of a simple proposition either. But I do not think that anything can be done about this. The notion of a simple proposition is basically epistemological, and hence its characterization must turn upon one's epistemological theories.

## 4. Counter-legal conditionals

A subjunctive conditional whose antecedent is inconsistent with the set $G$ of all true subjunctive generalizations is called a *counter-legal* conditional. Thus far we have been avoiding counter-legals. Now the analysis must be extended to include them. We can do this by making use of certain conclusions defended in Chapter II. We still restrict our attention to subjunctive conditionals having indicative antecedents and consequents.

To facilitate our discussion, let us now adopt a more complicated view of possible worlds. Thus far we have been identifying possible worlds with the sets of propositions true in those worlds. It is now more convenient to identify a possible world $\alpha$ with the ordered triple $\langle T_\alpha, N_\alpha, W_\alpha \rangle$ where $T_\alpha$ is the set of indicative propositions true in $\alpha$, $N_\alpha$ is the set of basic strong generalizations true in $\alpha$, and $W_\alpha$ is the set of basic weak generalizations true in $\alpha$. In Chapters V and VI we will adopt even more complex representations of possible worlds, but for now such ordered triples are sufficient for our purposes.

Now let us begin by considering strong subjunctive generalizations. There is a basic class $N$ whose members are confirmed inductively, and

then other strong generalizations are derived logically from those in $N$. Let $\forall N$ be the set of material generalizations corresponding to the subjunctive generalizations in $N$. Given two predicates $F$ and $G$, if the supposition $\ulcorner(\exists x)Fx\urcorner$ is consistent with $\forall N$, then $\ulcorner(Fx \Rrightarrow Gx)\urcorner$ is true iff $\ulcorner(x)(Fx \supset Gx)\urcorner$ is entailed by the generalizations in $\forall N$. But if the supposition $\ulcorner(\exists x)Fx\urcorner$ is inconsistent with $\forall N$ (so the generalization itself is counter-legal), it was argued that $\ulcorner(Fx \Rrightarrow Gx)\urcorner$ is true iff $\ulcorner(x)(Fx \supset Gx)\urcorner$ is entailed by every maximal-$\ulcorner(\exists x)Fx\urcorner$-consistent subset of $\forall N$. For present purposes, the best way to think of this is as follows. The different maximal-$\ulcorner(\exists x)Fx\urcorner$-consistent subsets of $\forall N$ represent the results of making minimal changes to $N$ in order to accommodate the truth of $P$, which in turn represent the possible choices regarding what *might* be the set of all true basic strong generalizations if $\ulcorner(\exists x)Fx\urcorner$ were true. In other words, these are the sets of basic strong generalizations in different $\ulcorner(\exists x)Fx\urcorner$-worlds. This immediately suggests a general way to deal with counter-legal subjunctive conditionals. If $P$ is inconsistent with $\forall N$, then the results of making different minimal $P$-changes to $N$ represent the sets of true basic strong generalizations in different $P$-worlds. Thus to accommodate counter-legal conditionals whose antecedents are inconsistent with the set of strong generalizations, we impose the following requirement on **M**:

(4.1) CONSERVATION OF STRONG GENERALIZATIONS: If $P$ is inconsistent with $\forall N$, and $\alpha \mathbf{M} P$, then $N_\alpha$, the set of basic strong generalizations true in $\alpha$, must be the result of making a minimal $\{P\}$-change to $N$.

The idea is that $P$ may force us to reject some of our strong generalizations, but we are constrained to reject as few as possible. As there may be different choices regarding which to reject in order to render the resulting set $P$-consistent, we must look at all of those choices in order to determine what would be true if $P$ were true.

In Chapter II, we identified the results of minimal $P$-changes to $N$ with the sets of subjunctive generalizations corresponding to the material generalizations in maximal-$P$-consistent subsets of $\forall N$. This, in effect, is to identify minimal $P$-changes to $N$ with strictly minimal $P$-changes to $N$, and in light of the discussion of section two, such an

identification is at least suspicious. It is plausible to suppose that we should instead proceed in a manner analogous to our analysis of minimal $P$-changes to $S_\alpha(t)$. I have been unable to find persuasive examples indicating that this level of sophistication is necessary at this point, but in the interest of safety, it seems best to follow this course anyway. The $P$-changes to $N$ that arise from indicative propositions are simply deletions, so it seems that we can analyze this notion as follows:

(4.2) If $X$ and $Y$ are sets of subjunctive generalizations $\Gamma$ is a set of propositions, and $k$ is a natural number, then $Y$ is the result of making a minimal$_k$ $\Gamma$-change to $X$ iff either:
(a) $\forall Y$ is a maximal-$\Gamma$-consistent subset of $\forall X$; or
(b) there is a sequence $\Lambda$ of sets of propositions such that $\Gamma \leftrightarrow \Pi\Lambda$, and for each $i \in \omega$, $\Lambda \rightarrow \Lambda_{i+1}$, and there is a sequence $\chi$ such that for each $i \in \omega$, $\chi_{i+1} \subseteq \chi_i$ and for some $j < k$, $\chi_i$ is the result of making a minimal$_j$ $\Lambda_i$-change to $X$, and $Y = \cup\chi$; or
(c) there is a sequence $\Lambda$ of sets of propositions such that $\Gamma \leftrightarrow \Sigma\Lambda$, and for each $i \bullet \omega$, $\Lambda_i \rightarrow \Lambda_{i+1}$, $Y$ is the result of making a $\Gamma$-change to $X$, and for some $j < k$, $X$ is the result of making a minimal$_j$ Neg $\Gamma$-change to $Y$.

(4.3) $Y$ is the result of making a minimal $\Gamma$-change to the set of $X$ of subjunctive generalizations iff for some natural number $k$, $Y$ is the result of making a minimal$_k$ $\Gamma$-change to $X$.

Next consider what happens when $P$ conflicts with the class of weak generalizations. Weak generalizations have a structure analogous to that of strong generalizations. There is a class $W$ of basic weak generalizations, and then others are inferred from the basic ones together with the strong generalizations. When $P$ conflicts with the weak generalizations, or more generally with $N \cup W$, we first modify $N$ as above to render it $P$-consistent, and *then* we delete as few things as possible from $W$ so as to render $W_\alpha$ consistent with $\forall N_\alpha \cup \{P\}$. The basic idea is that strong generalizations take precedence over weak generalizations, so we first conserve as many of the strong generaliza-

tions as possible, and then having done that, we proceed to conserve as many of the weak generalizations as possible:

(4.4) CONSERVATION OF WEAK GENERALIZATIONS: If $P$ is inconsistent with $\forall W$, then $W_\alpha$, the set of weak generalizations true in $\alpha$, must be the result of making a minimal $\forall N_\alpha \cup \{P\}$-change to $W$.

If we put all of these restrictions together, we obtain what I believe is a complete analysis of **M** for the case of indicative antecedents. Let $\alpha_0$ be the actual world. Then the analysis of **M** can be stated as follows:

(4.5) ANALYSIS VII: $\alpha \mathbf{M} P$ iff

(i) $P \in T_\alpha$;

(ii) $N_\alpha$ is the result of making a minimal $P$-change to $N_{\alpha_0}$;

(iii) $W_\alpha$ is the result of making a minimal $\forall N_\alpha \cup \{P\}$-change to $N_{\alpha_0}$;

(iv) for every time $t$, $(S_{\alpha_0}(t) \triangle S_\alpha(t))$ is a minimal $(\forall N_\alpha \cup \forall W_\alpha \cup \{P\})$-change to $T_{\alpha_0}$ at time $t$.

This analysis, together with the definition:

(4.6) DEFINITION: $\ulcorner P > Q \urcorner$ is true iff, for every possible world $\alpha$, if $\alpha \mathbf{M} P$ then $Q \bullet T_\alpha$.

constitutes my analysis of simple subjunctives for the case in which $P$ and $Q$ are both indicative statements.

Although our analysis is restricted to the case in which the antecedent and consequent of a conditional are indicative, it is still of some interest to see what the analysis entails about the logical properties of these conditionals. First, if the analysis is correct, logical necessity, and hence logical entailment, can be defined in terms of the subjunctive conditional. I have argued that a correct analysis must have the consequence that every $P$-change contains a minimal $P$-change, and this has the result that $\ulcorner P > \sim P \urcorner$ is true iff there are no $P$-changes. Thus, on the as yet unverified assumption that my analysis does satisfy my stated criterion of adequacy, we have:

(4.7) $\ulcorner \Box P \urcorner$ is true iff $\ulcorner (\sim P > P) \urcorner$ is true.

Second, the principles 6.1–6.8 of Chapter II are all valid. This means that the axiomatic theory SS formulated in Chapter II is sound. Third,

the principle 7.1 of Chapter II, which distinguished *SS* from David Lewis' theory *C*1, is not valid. Thus Analysis VII is in accord with the conclusions defended intuitively in Chapter II.

Now what remains is to generalize Analysis VII to include the case in which *P* and *Q* are not indicative sentences. This is basically a matter of turning the analysis into a recursive definition, and will be the topic of Chapter VI. We will also want to combine quantifiers with subjunctive conditionals. The groundwork for this will be laid in Chapter V, and then the full theory will be developed in Chapter VI.

## 5. Subject preference

Before proceeding with the further analysis of the simple subjunctive, let us turn to a different subjunctive conditional which is really a variant of the simple subjunctive. In Chapter I, in discussing whether subjunctive conditionals are subject to pragmatic ambiguity, we had occasion to discuss pairs of conditionals like:

(A) If that *were* gold, it would be malleable.

(B) If *that* were gold, some gold things would not be malleable.

At that point it was suggested that the apparent conflict between these two conditionals could be resolved by resolving an ambiguity in the antecedent. It was claimed that despite appearances, the antecedents of these conditionals do not have the same meaning. The difference in emphasis changes the meaning so that (B) actually means something like:

(B*) If some gold were like that (i.e., had the properties that has), then some gold things would not be malleable.

I believe that as a resolution of the conflict between the members of this particular pair of conditionals, this is acceptable. However, this pair of conditionals constitutes an example of a more general phenomenon which cannot always be dealt with so simply. Conditional (A) is a perfectly straightforward simple subjunctive conditional and is to be analyzed in accordance with Analysis VII. But conditional (B) is *not* a simple subjunctive conditional, although it is a close cousin to the

simple subjunctive. Conditional (B) is an example of what I will call a 'preferred subject conditional'. Before considering how preferred subject conditionals are to be analyzed, let us consider some additional examples. An interesting pair of conditionals is:

(C) If I were a member of the *Lakers*, I would be a good basketball player.

(D) If *I* were a member of the Lakers, they would not have a very good team.

In defense of (C) we simply observe that there is a weak subjunctive generalization to the effect that no one could be a member of the Lakers unless he were a good basketball player. Thus the truth of (C) is entirely in accordance with our analysis of the simple subjunctive. But in defense of (D), we observe that I am a lousy basketball player, and hence were I a member of the Lakers they would not have a very good team. The effect of emphasizing 'I' in the antecedent of (D) seems to be to indicate that in evaluating the truth of this conditional we seek to preserve my attributes in preference to those of the Lakers. We might try to paraphrase (D) as we did (B) to get something like:

(D*) If some members of the Lakers were like I am (i.e., had the properties I have), then the Lakers would not have a very good team.

But we can see that neither (D*) nor (B*) is an entirely adequate paraphrase. The difficulty is that, presumably, one of my properties is that of not being a member of the Lakers, but we do not want (D*) to say anything like 'If there were some members of the Lakers who were not members of the Lakers, then . . .'. We are naturally led to amend (D*) to talk about 'non-relational properties', but this is not a very clear notion and it will turn out below that such a move is not adequate to deal with other examples of subject preference anyway.

As a final example,[8] consider the Shah who has a famous collection of blue-white diamonds, and consider a grubby little industrial diamond. Then we have:

(E) If that were one of the *Shah's* diamonds, it would be blue-white.

(F) If *that* were one of the Shah's diamonds, not all of the Shah's diamonds would be blue-white.

The antecedent of (E) or (F) is ambiguous between 'If that diamond belonged to the Shah, . . .' and 'If that were one of $a_1, \ldots, a_n$ (where $a_1, \ldots, a_n$ are the diamonds belonging to the Shah) . . .'. On the former reading, this becomes analogous to (C) and (D) above. But let us take it instead in the latter, counter-identical, sense. Then we would agree that both (E) and (F) are true. However, neither (E) nor (F) would be true according to Analysis VII. In order to make the antecedent true, we must modify some simple truths regarding either the industrial diamond or the Shah's diamonds, but we have a choice regarding which to modify. Thus all that the results from Analysis VII is the uninteresting conclusion that if the industrial diamond were one of the Shah's diamonds, it might be blue-white, or it might instead be the case that not all of the Shah's diamonds would be blue-white.

It seems that the effect of emphasizing 'that' in (F) is to give preference to the properties of the industrial diamond over the properties of the Shah's diamonds in modifying the set of simple truths so as to make it consistent with the supposition that the industrial diamond is one of the Shah's diamonds. The effect of emphasizing 'Shah's diamonds' in (E) is just the opposite – to give preference to the properties of the Shah's diamonds. This seems to be what happens in general in cases of subject preference. In (B) the effect of emphasizing 'that' is to give preference to the properties of the piece of cast iron, and in (D) the effect of emphasizing 'I' is to give preference to my properties. This seems to be the effect in general of preferred subjects. In a conditional ⌜If it were true that $P$ then it would be true that $Q$⌝, if there is no preferred subject, then in constructing $P$-worlds we treat everything on a par with everything else, throwing all simple truths into the pot together. But if there is a preferred subject, then we give preference to the simple truths regarding that subject. That is, we only modify them insofar as they are themselves incompatible with $P$, and then we modify the other simple truths subject to the constraint that the result must be compatible with the surviving simple truths about the preferred subject.

Let us symbolize a preferred subject conditional by subscripting the connective with the individual term for the preferred subject: $\ulcorner P \gtrsim_a Q \urcorner$.

What is required for this to be true is, not for $Q$ to be true in every $P$-world, but for $Q$ to be true in every $\langle P, a\rangle$-world, where these are the worlds constructed by giving preference to the simple truths about the preferred subject. How, exactly, do we construct $\langle P, a\rangle$-worlds? First, let us ignore any effect on the subjunctive generalizations and just look at what happens to the simple truths. Letting $S_a$ be the set of simple propositions about the preferred subject $a$, we first set about making $(S_a \cap T_{\alpha_0})$ $P$-consistent and then afterwards we render $(S \cap T_{\alpha_0})$ consistent with both $P$ and our decisions regarding $(S_a \cap T_{\alpha_0})$. In deciding how to modify $(S_a \cap T_{\alpha_0})$, it may naturally seem that we must still take undercutting into account. However, let us reserve our opinion on this for the moment, and simply assume that $(S_a \cap T_\alpha)$ must be $P$-consistent.

Ignoring any changes in the subjunctive generalizations, it seems that we should modify Analysis VII by replacing clause (iv) by the two clauses:

(iv) $((S_a \cap T_{\alpha_0}) \triangle (S_a \cap T_\alpha))$ is a minimal $(\forall N_\alpha \cup \forall W_\alpha \cup \{P\})$-change to $(S_a \cap T_{\alpha_0})$ (in the sense of definition 2.20);

(v) for every time $t$, $(S_{\alpha_0}(t) \triangle S_\alpha(t))$ is a minimal $(\forall N_\alpha \cup \forall W_\alpha \cup (S_a \cap T_\alpha) \cap \{P\})$-change to $T_{\alpha_0}$ at time $t$.

So far we are ignoring any effect that the counterfactual hypothesis may have on the subjunctive generalizations, but before looking at that we must observe that a slight generalization of the above is required in order to deal with the example of the Shah's diamonds. We can handle (F) as above, but in the case of (E) our preferred subject is not an individual but rather a set of individuals (the set of all the Shah's diamonds). In evaluating the truth of (E), we consider worlds constructed by giving preference to the simple truths about each of the Shah's diamonds. Thus we must define a more general kind of preferred subject conditional: $\ulcorner P \underset{X}{>} Q \urcorner$ where $X$ is a set of individual terms. We can then define $\ulcorner P \underset{a}{>} Q \urcorner = \ulcorner P \underset{\{a\}}{>} Q \urcorner$. Defining:

(5.1) $$S_X = \bigcup_{a \in X} S_a$$

we can modify clauses (iv) and (v) above by replacing '$S_a$' throughout by '$S_X$'.

Next, let us ask what effect subject preference has on the subjunctive generalizations. We might naturally suppose that it has no effect, i.e., that clauses (ii) and (iii) of Analysis VII are left unchanged. However, examples (B) and (D) indicate that this is incorrect. In (D) we are using the simple truths about my athletic prowess to override the weak subjunctive generalization that no one could be a member of the Lakers unless he were a good basketball player. And in (B) we are using the simple truths about the piece of cast iron to override the strong subjunctive generalization that all gold is malleable. Thus in a preferred subject conditional, the simple truths about the preferred subject take absolute precedence over everything else. We first set them consistent with the counterfactual hypothesis, and then we proceed to make our other modifications to the world in accordance with the clauses of Analysis VII. As we deal with the simple truths about the preferred subject before we deal with subjunctive generalizations, we cannot deal with them in terms of their historical antecedents, and hence undercutting cannot be involved. So I propose the following:

(5.2) $\alpha$ is a $\langle P, X\rangle$-world iff

(i) $P \in T_\alpha$;

(ii) $((S_X \cap T_{\alpha_0}) \triangle (S_X \cap T_\alpha))$ is a minimal $P$-change to $(S_X \cap T_{\alpha_0})$;

(iii) $N_\alpha$ is the result of making a minimal $((S_X \cap T_\alpha) \cap \{P\})$-change to $N_{\alpha_0}$;

(iv) $W_\alpha$ is such that $(W_\alpha \cup N_\alpha)$ is the result of making a minimal $((S_X \cap T_\alpha) \cup \{P\})$-change to $(W_{\alpha_0} \cup N_{\alpha_0})$;

(v) for every time $t$, $(S_{\alpha_0}(t) \triangle S_\alpha(t))$ is a minimal $(\forall N_\alpha \cup \forall W_\alpha \cup (S_X \cap T_\alpha) \cup \{P\})$-change to $T_{\alpha_0}$ at time $t$.

(5.3) $\ulcorner P \underset{X}{\geqslant} Q \urcorner$ is true iff $Q$ is true in every $\langle P, X\rangle$-world.

One glaring shortcoming of this analysis is that it proceeds in terms of the notion of a statement being 'about' a particular object. This is a very problematic notion in general, and I am not prepared to give an account of it. However, some solace can be gained from the observation that the application of this notion seems relatively more obvious and transparent in the case of simple propositions than it is for

propositions in general. In particular, for the case of simple propositions 'about' appears to be extensional:

(5.4) $a = b \supset S_a = S_b$

This means that we can think of the subscripts on preferred subject conditionals as simply picking out individuals or sets of individuals rather than thinking of them linguistically as picking out individual terms or sets of terms.

It should be emphasized here that preferred subject conditionals constitute a genuinely different kind of subjunctive conditional. They are close cousins of simple subjunctive conditionals, but they are not the same. We have given a different analysis for preferred subject conditionals. Actually, turning things around, we can regard the simple subjunctive as a kind of limiting case of preferred subject conditionals, viz., the case in which the set of preferred subjects is empty:

(5.5) $(P > Q) \equiv (P \underset{\varnothing}{>} Q)$.

In Chapter I, in arguing against the purported pragmatic ambiguity of the simple subjunctive, I maintained that many putative counter-identicals are not really counter-identicals. We can now see how that claim should be understood. What is true is that many putative counter-identicals are not counter-identical simple subjunctive conditionals, but many of them do turn out to be counter-identical preferred subject conditionals. For example, consider:

(G) If Richard Nixon were Golda Meir, he would be a woman.
(H) If Golda Meir were Richard Nixon, she would be a man.

These can be analyzed as:

(G*) $RN = GM \underset{GM}{>} RN$ is a woman.

(H*) $GM = RN \underset{RN}{>} GM$ is a man.

In counter-identicals, the preferred subject is often indicated by the word order. However, this can often be overridden through the use of emphasis.

We have seen that some putative counter-identicals really are

counter-identical preferred subject conditionals. However, there still remain a number of putative counter-identicals which are not really counter-identicals on any reading, viz.,

(I) If I were Gerald Ford, I would sign the education bill.

This can only be analyzed in terms of roles as meaning:

(I*) If I were in the role of Gerald Ford, I would sign the education bill.

It seems that preferred subject conditionals are of some importance in understanding a number of our subjunctive locutions. I will not pursue their analysis further as we go on to consider quantification and the case of non-indicative antecedents and consequents, but it should be quite obvious how to extend the analysis to those cases.

## NOTES

[1] My terminology differs from that of David Lewis, who calls any world in which $P$ is true a $P$-world.

[2] For example, $Q$ is equivalent to $\ulcorner(Q \vee R) \,\&\, (Q \vee \sim R)\urcorner$.

[3] This is defended in Pollock (1974).

[4] $(X \triangle Y)$ is defined to be $(X \cup Y) - (X \cap Y)$.

[5] Notice that we do need the qualification 'non-inductive'. It is quite possible to have an inductive reason for believing a disjunction which is not composed of reasons for believing each disjunct separately.

[6] More precisely, in the jargon of Pollock (1974) the justification conditions for the statement 'The car is white' include non-conclusive non-inductive logical reasons, whereas the justification conditions for conjunctions and disjunctions include only conclusive non-inductive logical reasons.

[7] This has the effect of excluding subjunctive generalizations more or less arbitrarily. It makes no difference to the analysis of subjunctive conditionals whether we count subjunctive generalizations as simple, because they are preserved anyway in constructing $P$-worlds.

[8] I am indebted to Keith Lehrer for this example, and for getting me to think about subject preference in the first place.

CHAPTER V

# QUANTIFICATION, MODALITIES, AND CONDITIONALS

## 1. Referential opacity

In this chapter I want to begin the investigation of what happens when we combine quantifiers with subjunctive conditionals. This is a more difficult problem than might at first be supposed. It is not entirely obvious that quantification into subjunctive conditionals makes sense. Part of the problem can be seen as follows. We have seen that if our analysis is correct, then logical necessity can be defined in terms of subjunctive conditionals:

$$\Box\varphi \equiv (\sim\varphi > \varphi).$$

Thus quantification into subjunctive conditionals contains as a special case that of quantification into modal contexts, and many philosophers believe the latter to be illegitimate.

The difficulty can be made manifest as follows. Ordinarily, $\ulcorner(x)\psi\urcorner$ is taken to mean that every object in the universe *satisfies* the open formula $\psi$. Thus if quantification into $\psi$ is to be legitimate, it must make sense to talk about an object satisfying $\psi$. If $\psi$ is referentially opaque, it seems not to make sense to talk about an object satisfying $\psi$. For a referring term $t$, let us write $\ulcorner\psi t\urcorner$ for the result of replacing the free occurrences of $x$ in $\psi$ by $t$. Then it seems that a necessary condition for an object to satisfy $\psi$ is that if $t$ is any term denoting the object, then $\ulcorner\psi t\urcorner$ must be true. In other words, if the object itself satisfies $\psi$, then it must make no difference how we describe the object. But this requires that $\psi$ be referentially transparent. If $\psi$ is referentially opaque, then it seems that the most we can say is that the *object cum description* satisfies $\psi$, which is tantamount to saying that it is not objects but rather individual concepts that satisfy $\psi$.

The received view on modal operators is that they generate referentially opaque contexts, so in light of the above equivalence it would follow that subjunctive conditionals are also referentially opaque. We can

apparently verify this by turning to particular examples. For example, Neil Armstrong was the first man to set foot on the moon. It is true that if the first man to set foot on the moon had been only three feet tall, the rungs on the ladder of the lunar lander would have been close together. But it is not true that if Neil Armstrong had been only three feet tall then the rungs on the ladder of the lunar lander would have been close together – on the contrary, had Neil Armstrong been only three feet tall, he would not have been admitted to the astronaut training program and someone else of normal height would have been the first man to set foot on the moon.

Are we to conclude then that subjunctive conditionals are referentially opaque and hence that it is illegitimate to quantify into them? The situation is not as clear cut as all that. The difficulty is with the notion of referential opacity itself. The customary way of defining referential transparency is that a formula $\varphi x$ is referentially transparent iff the substitutivity of identity holds for it:

$$a = b \supset (\varphi a \equiv \varphi b).$$

However, this is an unreasonably strong requirement. A term may denote a certain object when taken in isolation, but denote a different object in the context of a certain statement. For example, in isolation the definite description 'the first man to set foot on the moon' denotes Neil Armstrong, but it does not denote Neil Armstrong in the context of the subjunctive conditional 'If the first man to set foot on the moon had been only three feet tall, then the rungs on the ladder of the lunar lander would have been close together'. Let us say that a singular term is used *rigidly* in a statement if, in the context of that statement, it denotes the same thing it denotes in isolation.

Many referentially opaque contexts arise from the fact that terms are not used rigidly in those contexts. But this should not be taken as showing that quantification into those contexts is illegitimate. In order for an object to satisfy a formula $\ulcorner \varphi x \urcorner$, it is not required that if $t$ is any term denoting the object *in isolation* then $\ulcorner \varphi t \urcorner$ is true. That is much too strong a requirement. All that should be required is that if $t$ denotes the object *in the context* $\ulcorner \varphi t \urcorner$ then $\ulcorner \varphi t \urcorner$ is true. Thus referential transparency, as traditionally defined, is not required for the legitimacy of quantification.

What the above considerations really indicate is that the traditional definition of referential transparency does not capture the notion it is intended to capture. A reasonable course would be to retain the requirement that quantification presupposes referential transparency but redefine referential transparency in terms of individual terms used rigidly:

(1.1) $\ulcorner \varphi x \urcorner$ is referentially transparent iff it is possible to use a term $a$ rigidly in $\ulcorner \varphi a \urcorner$, and for any terms $a$ and $b$ used rigidly in $\ulcorner \varphi a \urcorner$ and $\ulcorner \varphi b \urcorner$, the following is necessarily true: $\ulcorner a = b \supset (\varphi a \equiv \varphi b) \urcorner$.

Notice that if $\ulcorner \varphi x \urcorner$ is referentially transparent in this weaker sense, then the following principle must hold:

(1.2) $(x)(y)[x = y \supset (\varphi x \equiv \varphi y)]$.

Are modal and subjunctive contexts referentially transparent in this weaker sense? I believe they are. Referential opacity requires the failure of substitutivity of identity, but it requires more than just that – it requires such failure even in the case of terms used rigidly. It is easy to find cases of the failure of substitutivity in modal and subjunctive cases, but it is not so easy to find such cases where the terms are used rigidly. The clearest examples of failure of substitutivity involve definite descriptions, and in those cases the terms are rather obviously not used rigidly. This is illustrated by the example of Neil Armstrong and the first man to set foot on the moon. In contrast to that example, it is quite possible to use a definite description rigidly in a subjunctive context. For example, 'If the man at the end of the bar were outside now, it would be much quieter in here' clearly does not mean 'If it were the case that there is a unique man who is at the end of the bar and he is outside, then . . .'. Rather, it means, 'There is a unique man who is at the end of the bar, and if *he* were outside now, then . . .'. In this case, the substitutivity of identity goes through without difficulty. For example, if the man at the end of the bar is Edward, we can conclude, 'If Edward were outside now, it would be much quieter in here'.

The difference between the rigid and non-rigid use of a definite description in a subjunctive conditional seems to coincide with a

difference in the scope of the definite description. $\ulcorner[\varphi(\iota x\theta x) > \psi]\urcorner$ is ambiguous between the following two formulas, which represent wide scope and narrow scope respectively:

$$(\exists!x)\theta x \;\&\; (\exists x)[(\theta x \;\&\; (\varphi x > \psi)]$$
$$[(\exists!x)\theta x \;\&\; (\exists x)(\theta x \;\&\; \varphi x).> \psi]$$

If the conditional is understood in the former way (wide scope), then the description is being used rigidly and we can substitute coreferential terms for it with impunity.

Appealing to the scope ambiguity to account for the failure of substitutivity does not automatically resolve the problem in favor of quantification into subjunctive conditionals, because it begs the question by assuming that such quantification makes sense in the first place. But such an appeal does undercut the standard argument against quantification by showing, in effect, that it begs the question against quantification by assuming that we cannot make the distinction between wide scope and narrow scope. To resolve the issue of the legitimacy of such quantification we must appeal to other considerations.

Considerable support for the legitimacy of quantification into subjunctive conditionals can be mustered by appealing to examples. First, we have examples of subjunctive conditionals containing definite descriptions which seem clearly to have wide scope. We already discussed the example, 'If the man at the end of the bar were outside now, it would be much quieter in here'. It seems that the definite description in this conditional can only be construed as having wide scope, and if wide scopes are possible in such contexts it follows that quantification into subjunctive conditionals must make sense. We can support this same conclusion more directly by appealing to examples which appear to be quite straightforward cases of quantification into subjunctive conditionals. Two such examples are:

> Somewhere there is a man who would solve all our problems if he were elected president.
>
> There is a lion out there who would eat you if he caught you.

These seem very obviously to be examples of quantification into subjunctive conditionals, and they seem to make perfectly good sense.

I believe that these considerations satisfactorily meet the difficulties involving the putative referential opacity of subjunctive conditionals. If it were not for the difficulties discussed in the next section, we could leave it at this and assume that quantification into subjunctive conditionals, and hence into modal contexts, is unobjectionable. But as we will see, there are further difficulties for this view.

## 2. Transworld identity

The semantics for subjunctive conditionals and modal operators require us to look at more than one possible world. $\ulcorner\Box\varphi\urcorner$ is true iff $\varphi$ is true in every possible world, and $\ulcorner\varphi > \psi\urcorner$ is true iff $\psi$ is true in every $\varphi$-world. By analogy, it would seem that $\ulcorner(\exists x)\Box\varphi\urcorner$ is true iff there is some object (in the actual world) which satisfies the open formula $\varphi$ in every possible world. But it seems that in order to know whether an object in one world satisfies a formula in another world, we must somehow locate that object in the second world. Thus the legitimacy of quantification into modal contexts seems to presuppose that it makes sense to talk about an object in one world being the same as an object in another world, i.e., it presupposes the meaningfulness of transworld identity.

But what could it possibly mean to say that an object in one world is the same object as one in another world? The attempt to answer this question has sometimes led philosophers to talk about essences and to suppose that objects can be reidentified across possible worlds in terms of their essences.[1] Although this seems at first to be the only way to make sense of transworld identity, it also seems totally implausible to suppose that objects really have such essences. This is not to deny that objects might have some essential attributes. For example, it is not implausible to suppose that certain basic sortal properties are essential (Brody 1973, Pollock, 1974, pp. 157–174). But it is not to be supposed that objects have sufficiently many essential attributes to enable us to reidentify them across possible worlds in terms of those attributes.

For many years the above considerations led me to despair of making sense of transworld identity and hence to disdain quantification

(over individuals) into modal and subjunctive contexts. However, just because we reject quantification over individuals, we are not precluded from endorsing some other sort of quantification which is related to individual quantification and in terms of which we can try to make sense of putative examples of quantification into modal and subjunctive contexts. There are two prime candidates for such alternative modes of quantification.

A simple and safe way of making sense of quantification into modal contexts is to let the quantifiers range over *individual concepts.* Formally, an individual concept can be taken to be a function on possible worlds which picks one object out of each world to be the designatum of the concept in that world[3] (Kanger, 1957, Kaplan, 1964). I can find nothing objectionable about quantification over individual concepts. The only difficulty is that such a device does not enable us to make sense of some statements that seem to be perfectly sensible. Although it may be reasonable to interpret quantifiers in terms of individual concepts, it is not reasonable to interpret individual terms as expressing individual concepts. For example, I would be saying something true if I were to point to a particular table and say, 'Although that table could be red, it could not be the number two.' There are infinitely many different individual concepts which designate that table in this world. Some of those individual concepts do not designate a red object in any world, and others designate the number two in some worlds. Thus we cannot interpret the above statement as meaning that every individual concept which designates that table in this world is such that it designates something red in some world and designates the number two in no worlds. Apparently we are not talking about every possible individual concept which designates the table in this world. Are we then talking about some unique 'privileged' individual concept chosen from among those that designate the table in this world? I can see no way to pick out such a concept. Such a move would only make sense on the kind of essentialism we have already rejected. Thus I think we must conclude that the appeal to individual concepts does not allow us to make sense of many perfectly sensible modal statements. Something more is required.

The difficulty is that we really do seem to make modal and subjunctive assertions which are literally *about* some particular thing, such as

the table. The natural way to make sense of such assertions is in terms of transworld identity, but that involves us in immense difficulties. David Lewis [1968] has suggested an ingenious alternative. This is that although other worlds do not literally contain the same individuals as this world, they may contain 'counterparts' of individuals in this world. Lewis proposes that if $a$ is an object in one world and $b$ is an object in a second world, $b$ is the counterpart of $a$ in the second world just in case $b$ bears at least a minimal similarity to $a$ and is more similar to $a$ than is any other object in that second world. Then the proposal is that to say that something is necessarily true of $a$ is to say that it is true of $a$ in this world and true of all counterparts of $a$ in other worlds. Thus, for example, the table could be red but could not be the number two because there are worlds containing counterparts of the table which are red, but there are no worlds containing counterparts of the table which are also counterparts of the number two.

The appeal to counterparts is attractive. It allows us to make sense of *de re* modal statements and quantification into modal contexts while absolving us of having to make sense of transworld identity. However, as Lewis formulated it, counterpart theory leads to difficulties when we consider subjunctive conditionals. Counterpart theory leads us to say that a conditional $\ulcorner Fa > Ga \urcorner$ is true iff $G$ is true of the counterpart of $a$ in every $\ulcorner Fa \urcorner$-world, where an $\ulcorner Fa \urcorner$-world is defined in terms of the counterparts of $a$. But look what happens when we consider a conditional like 'If I had been born in the place of Richard Nixon, with all of his genes, etc., and raised as he was raised, and he in turn had been born and raised in my place, then I would have been a president threatened with impeachment and he would have been an interested bystander'. If we symbolize the antecedent of this conditional as $\ulcorner Fab \urcorner$, where $a$ designates Richard Nixon and $b$ designates me, then to evaluate its truth we look at all $\ulcorner Fab \urcorner$-worlds. But according to counterpart theory, there are no $\ulcorner Fab \urcorner$-worlds. This is because in an $\ulcorner Fab \urcorner$-world, the counterpart of Richard Nixon would be less like Richard Nixon than would my counterpart be, and vice versa, in which case they could not be Richard Nixon's counterpart and my counterpart.

The above example indicates that it is unreasonable to require that the counterpart of an object be more like that object than is anything

else existing in the same world as the counterpart. It makes perfectly good sense to assert counterfactuals about what would happen if one object were quite different than it is and a second object was just like the first object is now. We must construct the counterpart relation more liberally. But this leads to new difficulties. If we cannot define the counterpart relation in the simple way proposed by David Lewis, then what we need is a criterion for counterparthood. But what could such a criterion be? The difficulties that arise here are precisely the same as those that arose in looking for a criterion for transworld identity. It seems that the only way to make sense of counterparts is in terms of some kind of essentialism, but essentialism is no more plausible now than it was before. It appears that the appeal to counterparts in place of transworld identity solves nothing.

## 3. Kripke's observation

Faced with the above difficulties, it would be very tempting to conclude that *de re* modalities and quantification into subjunctive conditionals do not make sense. But such a conclusion flies in the face of what appear to be clear examples to the contrary. I do not see how anyone can deny the meaningfulness of either 'That table could be red, but it could not be the number two' or 'There is a lion out there who would eat you if he caught you'. Where do we go from here?

The key to the solution to our problem has been provided by Saul Kripke (1971, 1972). Kripke's point is very simple. We are thinking of possible worlds in the wrong way. We are thinking of them as being given to us 'structurally', by giving a domain of objects and saying what properties those objects have, and then leaving it up to us to somehow identify the objects in the domain on the basis of those properties. Why shouldn't we instead have the identity of the objects be part of the specification of the world?

Recall once more what the basic intuition is behind our analysis of subjunctive conditionals. We begin with the set of all true statements, and then modify that set as little as possible to accommodate the truth of the counterfactual hypothesis. This intuition is in terms of *truths*, not *worlds*. It was asserted that we could capture this intuition by talking

about how possible worlds compare with one another, but in order to accomplish this the possible worlds must be closely related to sets of propositions. What is important here is that the truths which go into the makeup of the possible worlds are truths *about* particular objects in this world. Thus if we identify a possible world with the set of propositions true in it, the identity of the objects in that world will not be open to question. Their identity will be given by some of those truths. The truths are truths *about* particular objects. Here is a place in which the platonistic view of possible worlds is apt to mislead us (although it need not), whereas if we construe possible worlds to be sets of propositions we will not be led astray. As Kripke puts it, at least for present purposes we should think of possible worlds as *counterfactual situations.* As such, the identity of the objects which exist in any possible world will be given as part of the specification of the world. There is no need for a *criterion* of transworld identity. The search for such a criterion was doomed from the start because most any object can be most any way, and hence an object described in terms of its properties in some possible world could be identified with most any object in the actual world. There is no question of getting the identification right. We can identity objects however we choose, and that identification will be part of the specification of the world. This follows from the simple fact that we can have non-vacuous counterfactual hypotheses (and hence counterfactual situations, i.e., possible worlds) in which we suppose most anything we like about an object.

The above considerations seem to show that transworld identity makes sense after all, in a trivial sort of way. Instead of solving the problem of transworld identity, we have resolved it by showing that there isn't really any problem. But in spite of appearances, all is not yet smooth sailing.

Suppose I entertain the counterfactual hypothesis, 'If my house were painted blue . . .'. This is a hypothesis that is literally about my house. Does it follow then that possible worlds in which this hypothesis is true contain an object that is literally the same object as my house (in this world)? Building this hypothesis into the construction of a possible world in some sense fixes the identity of an object in that world, but does this really mean that the object in that world is the same object as my house? In other words, is the transworld relation between objects 'of the same identity' literally that of identity?

It may seem that the answer to this question is trivially 'Yes'. After all, the statement 'My house is painted blue' is literally *about* my house. But it is not clear what this establishes. One must be wary of putting too much weight on the word 'about'. After all, the negative existential 'My house does not exist' is also about my house, but it certainly does not follow that my house exists in any world in which the negative existential is true. It may be protested that it is unfair to appeal to negative existentials, because everyone knows that they are peculiar, but there is another kind of statement which makes it seem very suspicious to suppose that objects in different worlds are literally identical. This is that kind of counterfactual called a 'counteridentical'. We sometimes assert counterfactuals about what would be the case if two objects which are in fact distinct were one and the same object. For example, I might entertain the counterfactual hypothesis that Richard Nixon and Donald Nixon are one and the same person, this being brought off through the clever use of makeup, with the person in question sneaking away from San Clemente occasionally to play the role of Donald Nixon, and then sneaking back to play the role of Richard Nixon. A counterfactual with this antecedent is just as literally about both of the two Nixons as the statement 'My house is painted blue' is about my house. One might feel some temptation to reply that the conditional is really about one of the two Nixons and the other would not exist if the antecedent were true, but this is unsatisfactory because there is no way to say which of the Nixons it is about and which would not exist. I think it must be admitted that the relation between the two Nixons in this world and the single Nixon in the world in which the counterfactual hypothesis is true is precisely the same as the relation between my house in this world and my house in the world in which it is painted blue. But in the former case it is a relation between two distinct objects in one world and a single object in another world. How can this be identity?

Let us return to Lewis' terminology and call the relation between objects of the same identity in different worlds the *counterpart relation*, leaving open for the moment whether the counterpart relation is literally a relation of transworld identity. Let us write '$xCy$' for '$x$ is a counterpart of $y$'. Counteridenticals indicate that we can have $xCy$ and $xCz$ but $y \neq z$. If $C$ is the identity relation, this amounts to having $x = y$ and $x = z$, but $y \neq z$, which is to deny that identity is transitive. No

doubt many philosophers will regard the non-transitivity of identity as manifestly absurd, in which case it follows that the counterpart relation is not a relation of transworld identity. Personally, I am not so sure. I don't really know what to say about putative cases of the non-transitivity of identity. At this point I want merely to counsel caution. We must not be overly hasty in claiming that the counterpart relation is not an identity relation on the grounds that if it were then identity would not be transitive. I think the real problem here is that there is no clear meaning to the question whether the counterpart relation is or is not the relation of identity. Before we can answer this question, we must know what it would mean to have two objects in two different possible worlds literally identical or not identical. This is a notion I can get no clear grasp on. Furthermore, as far as I can see it makes not the slightest difference to anything whether the counterpart relation is an identity relation or not. It seems far better to simply call the relation 'the counterpart relation' and not worry about whether it is identity.

But now what becomes of Kripke's observation that there is no problem of transworld identity because the identity of an object is part of the specification of the world in which it is found? We can translate this into an observation about the counterpart relation. Lewis' counterpart theory floundered on the difficulty of finding a criterion for one object being a counterpart of another. We now see that what Kripke's observation amounts to is that the identity of an object is built into the world in which it is found in the sense that the counterpart relation is part of the specification of that world. Starting from the actual world, part of the specification of a possible world consists of saying what is a counterpart of what objects in the actual world. There is no problem of finding a criterion for counterparthood.

Counteridenticals show that the counterpart relation can converge – two distinct objects in the actual world can have a single counterpart in some possible world. Can the counterpart relation also diverge? That is, can a single object in the real world have two distinct counterparts in some possible worlds? To establish this, we would naturally turn to what we might call 'counter-nonidenticals'. It is easy enough to formulate counterfactual hypotheses which deny identities that hold in the actual world. For example, we can reason about what would be the case if the tallest man on the basketball team were not also the worst

shot on the team. But this shows nothing about the counterpart relation, because the individual terms involved are not being used rigidly. What we need is a counter-non-identical in which the terms flanking the identity sign are used rigidly to denote the same object. Unfortunately, I cannot find an intelligible way of even formulating such a counter-non-identical. If we try to do it by telling a story as in the Donald and Richard Nixon case, we do not get the desired result. For example, if we hypothesize that the elder Nixons had two twin brothers, and that they have been fooling people by alternately hiding and playing the role of Richard Nixon, this would not be a case in which we could *literally* say that Richard Nixon is two people. On the contrary, what would really be the case is that there is no such person as Richard Nixon. This seems to be what happens in general when we try to formulate counter-nonidenticals with individual terms used rigidly. To suppose that one thing is really two things is to suppose that that one thing does not really exist.

I must admit that these examples are not as compelling as we might wish. To further muddy the waters, it does seem to be possible to formulate counter-non-identicals using proper names, and it has sometimes been maintained that proper names can only be used rigidly. It does not seem possible to maintain the latter thesis with respect to negative existentials involving proper names, and I do not really think it can be maintained with respect to counter-non-identicals either. But the situation here is not clear. For example, what are we to say about a conditional which begins, 'If Hesperus and Phosphorus were distinct heavenly bodies, then . . .'? My own intuitions here alternate between on the one hand really taking seriously the denotation of the terms 'Hesperus' and 'Phosphorus' and finding this conditional unintelligible, and on the other hand taking 'Hesperus' and 'Phosphorus' as short for definite descriptions on the order of 'the first heavenly body regularly seen in the morning' and 'the last heavenly body regularly seen in the evening'. I think the weight of intuition is thus on the side of denying that the counterpart relation can diverge, but these intuitions are not clear and are not really compelling. What we need is an argument based upon other considerations which will settle the matter.

Such an argument can be provided. On purely intuitive grounds it seems that quantification into modal and subjunctive contexts makes

sense. The appeal to the counterpart relation explains how such quantification makes sense. But it was argued in section one that a necessary condition for quantification into a context $\ulcorner \varphi x \urcorner$ to make sense is that the context be referentially transparent in the sense of 1.1. This in turn implies that principle 1.2 holds:

$$(x)(y)[x = y \supset (\varphi x \equiv \varphi y)].$$

Applying this to modal contexts, because quantification makes sense in such contexts, the following must hold:

(3.1) $(x)(y)\{x = y \supset [\Box((\exists z)x = z \supset x = x) \equiv \Box((\exists z)x = z \supset x = y)]\}$.

$\ulcorner(\exists z)x = z\urcorner$ is a way of saying that $x$ exists. It is clearly true of anything in this world that in any other world in which it exists, it is self-identical:

(3.2) $(x)\Box((\exists z)x = z \supset x = x)$.

This immediately implies:

(3.3) $(x)(y)\{x = y \supset \Box[(\exists z)x = z \supset x = y]\}$.

Principle 3.3 is precisely the statement that the counterpart relation cannot diverge. It says that if $x$ and $y$ are the same object in this world, then they are the same object in any world in which they exist. Thus it seems that the very legitimacy of quantification into modal contexts requires that the counterpart relation cannot diverge. The counterpart relation must be a many-one relation, i.e., a function.[2]

This generates a picture of possible worlds which is rather different from the classical one. In the next section I will attempt to make it precise and investigate what sort of modal logic results from it. Then in the next chapter we will apply these results to an examination of quantified subjunctive conditionals.

## 4. Quantified modal logic

Let us begin by constructing a formal language for a first-order modal logic. The logical constants are '(', ')', '~' '&', '=', and '□'. We have a denumerable set *Cn* of individual constants, a denumerable set *Vr* of individual variables, and for each $n \in \omega$, a denumerable set $\mathfrak{R}_n$ of

$n$-place relation symbols. We define the set *Fm* of formulas and the set *Sn* of sentences (closed formulas) in the normal way. We define '$\vee$','$\supset$', '$\equiv$', '$\exists$', and '$\Diamond$' in the usual way.

Our semantics will proceed in terms of possible worlds. A possible world has two parts – a structure consisting of a domain of objects and a determination of what properties each object has, and an identification of some of the objects in the domain with objects in the real world. The latter is in terms of the counterpart relation. Before considering the details of this, notice that this makes our definition of 'possible world' always relative to the real world, because the counterpart relation must relate objects in the domain of the possible world to objects in the real world. This is rather like those semantics which contain an accessibility relation together with the ruling that not all worlds are possible (or accessible) relative to a given world.

Notice further that we can have many different possible worlds having the same structure. Such worlds will differ only in their counterpart relation. Different such worlds correspond to similar counterfactual hypotheses about different objects. It would be impossible to have different worlds with the same structure on the traditional views according to which counterparthood or transworld identity is a function of the structural properties of the objects in the world.

A possible world will be an ordered pair $\langle S, C\rangle$ where $S$ is its structure and $C$ the counterpart relation. In specifying a structure, we must first specify a domain $D$. Then we must specify the extensions of all relation symbols in this domain. This can be accomplished by a function $\mu$ which assigns to each $n$-place relation symbol some subset of $D^n$ (the set of ordered $n$-tuples of members of $D$). So let us define:

(4.1) A *structure* is an ordered pair $\langle D, \mu\rangle$ where $D$ is a set of objects and $\mu$ is a function such that for each $n \in \omega$ and $R \in \Re_n$, $\mu(R) \subseteq D^n$.

Now consider the counterpart relation. Let $D_0$ be the set of objects in the real world. The domain of $C$ may be a proper subset of $D_0$, because some actual objects may not have counterparts in the possible world. Thus $C$ may be any function from a subset of $D_0$ into $D$.

Before giving a precise definition of a possible world, we must decide how to represent the actual world. Is it also a possible world? In

fact, we can represent the actual world by its structure alone, because insofar as there is a counterpart relation in the actual world, it is the identity relation. In general, given any structure we can define the notion of a possible world relative to that structure, which is to say what the set of possible worlds would be if that structure represented the actual world:

(4.2) A possible world relative to a structure $\langle D, \mu \rangle$ is an ordered pair $\langle S, C \rangle$ where $S$ is a structure $\langle E, \mu^* \rangle$, and $C$ maps a subset of $D$ into $E$.

How should we define validity for our modal language? Informally, we want a sentence to be valid iff it comes out necessarily true no matter how we interpret the non-logical constants of our language.[3] From an informal point of view an interpretation of the language does two things. First, it assigns denotations in the real world to the individual constants of the language. We are going to want to talk about models in which the 'real world' is a world that is just a possible world relative to this world, and in such a world some individual constants may fail to denote, so we must make provision for this. The second function of an interpretation of the language is to assign meanings to the relation symbols. Such an assignment has the effect that certain structures are ruled out as being 'really inconsistent' given the meanings of the relation symbols. This second function can be accomplished formally by simply specifying what the actual set of permissible structures is. However, a fact which logicians have generally overlooked is that not just any set of structures can constitute the set of all permissible structures under an assignment of meanings to the relation symbols. For example, if we consider a set of structures in which no structure has a domain of cardinality 3, to take this as our set of permissible structures would be to suppose that the meanings of the relation symbols somehow make it logically impossible for there to exist just three things in the universe. But the meanings of the relation symbols have nothing at all to do with what size universes are logically possible. I am supposing that the residents of our possible worlds are ordinary contingent objects like tables, chairs, electrons, etc., and for such objects it seems that it must be logically possible for the universe to have any finite or denumerable size. I don't know whether it should

be possible for the universe to be non-denumerably infinite, but fortunately, this is a question we can leave undecided as long as we are only dealing with first-order logic, because countenancing larger universes makes no difference to the satisfiability of any formula. At any rate, it is clear that the set of permissible structures must contain structures having domains of all finite or denumerable sizes. Furthermore, in light of the Skolem-Löwenheim theorem, we will only consider countable structures, i.e., structures having countable domains. Should we impose any other restrictions on sets of permissible structures? There is no need to do so. The only contingent facts that can be expressed in an uninterpreted first-order language are those having to do with the size of the universe, so any set of structures satisfying our restriction concerning the variety of cardinalities will correspond to some assignment of meanings to relation symbols.

To say that a sentence is necessarily true under every interpretation of the language is to say that no matter what structure we start with as the actual world, the sentence comes out true under every interpretation of the language. So let us define:

(4.3) A *model* is an ordered triple $\langle\langle D, \mu\rangle, \eta, H\rangle$ where $H$ is a set of countable structures, $\langle D, \mu\rangle \in H$, for each $\alpha \leq \omega$ there is a structure $S$ in $H$ whose domain is of cardinality $\alpha$, and $\eta$ maps a subset of $Cn$ into $D$.

Then we define:

(4.4) A sentence $\varphi$ is *valid* iff $\varphi$ is true in every model.

What remains is to define truth in a model. This definition is relatively straightforward except for the modal case. We first define:

(4.5) A model $\langle S^*, \eta^*, H^*\rangle$ is an *a-variant* of a model $\langle S, \eta, H\rangle$ iff $a \in Cn$, $S^* = S$, $H^* = H$, and $\eta^*$ agrees with $\eta$ except possibly for the value of $\eta^*(a)$.

Next we must decide how to handle sentences containing non-denoting individual constants. I will adopt what can probably be called 'the standard procedure' of giving every sentence a truth value, ruling that atomic sentences containing non-denoting constants are automatically false. Given this we can define truth in a model by induction on

the length of a sentence. Letting $\mathcal{D}(\eta)$ be the domain of $\eta$:

(4.6) If $M$ is the model $\langle\langle D, \mu\rangle, \eta, H\rangle$, *then*:

(i) if $a, b \in Cn$ then $\ulcorner a = b \urcorner$ is true in $M$ iff $a, b \in \mathcal{D}(\eta)$ and $\eta(a) = \eta(b)$;

(ii) if $R \in \Re_n$ and $a_1, \ldots, a_n \in Cn$, then $\ulcorner Ra_1, \ldots a_n \urcorner$ is true in $M$ iff $a_1, \ldots, a_n \in \mathcal{D}(\eta)$ and $\langle \eta(a_1), \ldots, \eta(a_n)\rangle \in \mu(R)$;

(iii) $\ulcorner \sim \varphi \urcorner$ is true in $M$ iff $\varphi$ is not true in $M$;

(iv) $\ulcorner (\varphi \,\&\, \psi) \urcorner$ is true in $M$ iff $\varphi$ and $\psi$ are both true in $M$;

(v) $\ulcorner \Box \varphi \urcorner$ is true in $M$ iff $\varphi$ is true in every model $\langle S, \eta^*, H^*\rangle$ such that $H^* = H$ and for any $c$, $d \in \mathcal{D}(\eta)$, if $\eta(c) = \eta(d)$ and $c \in \mathcal{D}(\eta^*)$ then $\eta^*(c) = \eta^*(d)$;

(vi) $\ulcorner (x)\varphi \urcorner$ is true in $M$ iff, if $c$ is some constant not occurring in $\varphi$, and $\ulcorner \varphi c \urcorner$ is the result of substituting $c$ for every free occurrence of $x$ in $\varphi$, then $\ulcorner \varphi c \urcorner$ is true in every $c$-variant of $M$.

The only clause of definition 4.6 that needs explanation is clause (v). We want $\ulcorner \Box \varphi \urcorner$ to be true in $M$ just in case $\varphi$ is true in every world possible relative to $\langle D, \mu\rangle$. Thus the most straightforward definition of truth for modal sentences would be that $\ulcorner \Box \varphi \urcorner$ is true in $M$ iff $\varphi$ is true in every model $\langle S, \eta^*, H\rangle$ for which there is a function $C$ such that $\langle S, C\rangle$ is a possible world relative to $\langle D, \mu\rangle$, and $\eta^*$ 'agrees' with the counterpart relation, i.e., if $c \in \mathcal{D}(\eta) \cap \mathcal{D}(\eta^*)$, then $C(\eta(c)) = \eta^*(c)$. But this is equivalent to requiring that any identities which hold in $\langle\langle D, \mu\rangle, \eta, H\rangle$ continue to hold in $\langle S, \eta^*, H\rangle$ provided that the terms in the identity continue to denote. This gives us clause (v).

It will be useful to define an 'existence' predicate as follows:

(4.7) If $t$ is an individual term, $\ulcorner Et \urcorner$ is $\ulcorner (\exists x) x = t \urcorner$.

The modal logic resulting from this semantics has a number of peculiar features. Some of these arise out of the fact that our modal operator is basically a *de re* operator. We can quantify into modal contexts, and appending '$\Box$' to a sentence yields a sentence that is literally *about* the objects mentioned in the sentence. In effect, the modal operator of logical necessity is a *de re* operator, and although *de dicto* modal statements do occur, they are just limiting cases of *de re*

statements. Let us define:

(4.8) $\varphi$ is a *de dicto* sentence iff $\varphi$ is a sentence and no subformula of $\varphi$ having the form $\ulcorner\Box\psi\urcorner$ contains occurrences of any free variables or individual constants. $\varphi$ is *de re* iff $\varphi$ is not *de dicto*.

It is easily verified that the logic of *de dicto* sentences is completely normal, satisfying exactly the propositional modal laws of *S*5. Surprisingly enough *de re* sentences behave differently, satisfying only the laws of *S*4. It is trivial to verify that sentences in general satisfy the modal laws of *S*4. The 'characteristic law' of *S*5 which is not satisfied is:

(4.9) $\Diamond\varphi \supset \Box\Diamond\varphi$.

Given the valid principle $\ulcorner\varphi \supset \Diamond\varphi\urcorner$, 4.9 implies:

(4.10) $\varphi \supset \Box\Diamond\varphi$.

Due to the fact that our counterpart relation can converge but not diverge (that is, it is a function but not necessarily one-one), it is simple to construct counter-examples to 4.10. First, notice that because the counterpart relation cannot diverge, both of the following are valid:

(4.11) $a = b \supset \Box(Ea \ \& \ Eb. \supset a = b)$.

(4.12) $(x)(y)[x = y \supset \Box(Ex \ \& \ Ey. \supset x = y)]$.

That is, individual constants and variables act as 'rigid designators'. By virtue of the validity of 4.11, $\ulcorner\Diamond a = b. \supset \Diamond\Box(Ea \ \& \ Eb. \supset a = b)\urcorner$ is valid, and hence $\ulcorner\Diamond a = b. \supset \sim\Box\Diamond[Ea \ \& \ Eb \ \& \ a \neq b]\urcorner$ is valid. But this implies the invalidity of:

(4.13) $(Ea \ \& \ Eb \ \& \ a \neq b) \supset \Box\Diamond(Ea \ \& \ Eb \ \& \ a \neq b)$.

This is a counterexample to 4.10. 4.13 is invalid for the following reason. Suppose '$Ea \ \& \ Eb \ \& \ a \neq b$' is true in the real world. There are worlds possible relative to the real world in which $\ulcorner a = b\urcorner$ is true, i.e., $\ulcorner\Diamond a = b\urcorner$ is true in the real world. But then $\ulcorner\sim\Box\Diamond[Ea \ \& \ Eb \ \& \ a \neq b]\urcorner$ is true.

Lest it be thought that the only counterexamples to 4.9 and 4.10 are those involving identity, notice that the following is invalid for exactly

the same reason 4.13 is invalid:

(4.14) $(Ea \ \& \ Eb \ \& \ Fa \ \& \ \sim Fb) \supset \Box\Diamond(Ea \ \& \ Eb \ \& \ Fa \ \& \ \sim Fb)$.

Thus our semantics gives us a version of quantified *S*4. The treatment of identities is interesting. Because the counterpart relation is a function, we get the validity of 4.11 and 4.12. However, because it need not be one-one, neither of the following is valid:

(4.15) $a \neq b \supset \Box a \neq b$.

(4.16) $(x)(y)(x \neq y \supset \Box x \neq y)$.

We have a version of quantified *S*4, but as we have seen, it is not quite the same as any of the familiar versions of quantified *S*4. This results in part from the treatment of identity and the counterpart relation. Another source of differences results from our 'variety of cardinalities' requirement on models. For each $n \in \omega$, let $\ulcorner En \urcorner$ be a first-order sentence which says that the universe contains exactly $n$ objects. Then $\ulcorner \Diamond En \urcorner$ is valid on our semantics. No doubt some logicians will find this objectionable. They will reason that our logic should be adequate for dealing with any subject matter, and some subject matters will restrict the size of the universe. For example, we might want to talk about the natural numbers. Thus, it will be urged, we should eliminate the 'variety of cardinalities' requirement. I have no particular objection to building a modal logic on such a basis. However, unlike ordinary first-order logic, I do not find modal logic very useful for talking about anything except contingent objects. As long as we are going to take our domains to consist exclusively of contingent objects, then we should impose the 'variety of cardinalities' requirement on models. Of course, there is nothing to prevent one from having both kinds of modal logic, using one for some purposes and the other for other purposes.

Another unusual valid sentence is $\ulcorner \Box \varphi a \supset \Box(x)\varphi \urcorner$ (where no individual constants occur in $\varphi$). This theorem results from the fact that in constructing a model, we take the set of structures as basic and then look at all the possible worlds that can be built out of those structures using different counterpart relations. The more conventional procedure is to take the set of possible worlds itself as basic in constructing a

model. However, this conventional procedure seems to me quite definitely wrong. Insofar as different sets representing the set of all possible worlds are supposed to result simply from different assignments of meanings to the relation symbols, there should be no restrictions on the counterpart relations. If a certain structure is permissible in terms of a certain assignment of meanings, then it should be possible for any object in the real world to have the properties of *any* object in the domain of that structure. Thus given the possibility of one world having a certain structure, any other world having that same structure but a different counterpart relation should also be possible. Hence the validity of $\ulcorner\Box\varphi a \supset \Box(x)\varphi\urcorner$ when no individual constants occur in $\varphi$. This constitutes a major divergence between the above semantics and the conventional semantics for quantified modal logic.

It would be remiss not to admit that there is one possible exception to what I have just argued. The validity of $\ulcorner\Box\varphi a \supset \Box(x)\varphi\urcorner$ has the effect of ruling out essential properties. For the most part, I find this an exemplary result. However, as I suggested earlier, there may be certain kinds of properties which are essential. These are what might be called 'basic sortals'. Basic sortals, in some sense, would tell us 'what kind of thing' an object is. Basic sortals are going to be properties like 'physical object', 'person', 'work of art', etc. These basic sortals seem to impose restrictions on the counterpart relation, because it seems to be the case that an object satisfying one basic sortal could not shed that sortal and come to satisfy a different basic sortal. For example, a person could not become a table. I am not sure about any of this, and I am even less sure about how to accommodate these basic sortals in our formal semantics. For this reason I have ignored them, but it must be recognized that we may want to modify the semantics at some future time so that we can deal with them.

It is of interest to look at what happens to a few other problematic principles of quantified modal logic. The 'Barcan formula', $\ulcorner(x)\varphi \supset \Box(x)\varphi\urcorner$ fails to be valid for two reasons. First, it can fail to be valid when $\varphi$ contains individual constants. For example $\ulcorner(x)\Box(Ea \supset x = a) \supset \Box(x)(Ea \supset x = a)\urcorner$ is invalid. If we eliminate this source of invalidity by requiring that $\varphi$ contain no individual constants, the Barcan formula is still invalid but only because we can have structures with empty domains. The validity of the following weakened version of the

Barcon formula:

$(\exists x)(x = x) \supset [(x)\Box\varphi \supset \Box(x)\varphi]$ (where $\varphi$ contains no individual constants)

follows immediately from the validity of $\ulcorner\Box\varphi a \supset \Box(x)\varphi\urcorner$.

A related formula that is often considered problematic is $\ulcorner\Box(\exists x)\varphi \supset (\exists x)\Box\varphi\urcorner$. This formula is valid without restriction on our semantics, but this is because $\ulcorner\sim\Box(\exists x)\varphi\urcorner$ is valid. This results from the variety of cardinalities requirement.

## 5. Conditionals

It would now be a simple matter to add to our language subjunctive conditionals with indicative antecedents and consequents and extend the semantics to deal with them. This would involve modifying the notion of a structure so as to include $N$ and $W$, the sets of basic strong and weak subjunctive generalizations, as part of the structure, and it would require the introduction of some way to compare the dates of simple propositions. There are a number of little intricasies in this, but they are all rather straightforward. However, rather than delve into these details just for the special case of conditionals with indicative antecedents and consequents, let us go on to the full theory in which we allow conditionals to have modal and subjunctive antecedents and consequents. As we will see, some new problems arise when we undertake this.

## NOTES

[1] See Chisholm (1967) for a critical discussion of this.

[2] In trying to convince me of the contrary, philosophers have repeatedly appealed to the fact that both fission and fusion are possible in temporal contexts. That, of course, is quite true, but I do not see what it has to do with the issue at hand. I do not see any way out of the argument given in the above paragraph, so we should conclude that the temporal modalities do not work like the alethic modalities.

[3] See Pollock (1967) for a discussion of this notion of validity.

CHAPTER VI

# THE FULL THEORY

The objective of this chapter is to extend the analysis of subjunctive conditionals in such a way as to incorporate both quantifiers and conditionals with non-indicative antecedents and consequents. This is a matter of combining Analysis VII of Chapter IV with the treatment of quantification discussed in Chapter V, and then making the whole thing recursive in order to handle non-indicative antecedents and consequents. We will begin by constructing an artificial language within which to express our conditionals, and then our analysis will take the form of a formal semantics for this language.

## 1. SYNTAX

Our language must be a temporal language, because in order to express Analysis VII we must be able to talk about the dates of simple propositions. The simplest way to accomplish this is to make our language a two-sorted language with constants and variables for times, and a temporal relation '$\leq$' between times. Then the 'date' of an atomic formula will be indicated by a temporal subscript. In order for this to make sense, we must ensure that our atomic sentences always express propositions of the sort that have discrete dates. This is easily accomplished as follows. We must have some way of picking out the set Sim of simple sentences (those expressing simple propositions). Simple propositions are supposed to be those that are not logical compounds, so it seems reasonable to require that our atomic sentences be the simple ones. This will ensure the appropriateness of our temporal subscripts in atomic formulas.

The logical constants of our language will be '(', ')', '$\sim$', '&', '=', '$>$', '$\Rrightarrow$', '$\Rightarrow$', '$\Box_p$', '$\Box_a$', and '$\leq$'. We must now have a non-denumerable set *Cn* of individual constants, a denumerable set *Vr* of individual variables, a denumerable set $C_T$ of temporal constants, a denumerable set

$V_T$ of temporal variables, and for each $n \bullet \omega$ such that $n \neq 0$ we have a denumerable set $\Re_n$ of $n$-place relation symbols. We define the set of formulas in almost, but not quite, the usual manner. The difficulty is that the account of subjunctive generalizations in Chapter III only applies to subjunctive generalizations having indicative antecedents and consequents and it is not at all obvious how to extend it to the case of non-indicative antecedents and consequents. Hence we are led to count $\ulcorner(\varphi \Rrightarrow \psi)\urcorner$, $\ulcorner(\varphi \Rightarrow \psi)\urcorner$, $\ulcorner\underset{p}{\Box}\psi\urcorner$, and $\ulcorner\underset{a}{\Box}\psi\urcorner$ as well-formed only when $\varphi$ and $\psi$ are indicative. This leads, in effect, to a two-step syntax. We begin with a two-sorted first-order language with the logical constants '(', ')', '$\sim$'. '&', '=', and '$\leq$'. We define the set of atomic formulas as follows:

(1.1) An expression $\varphi$ is in $At$ iff either:

(i) there are $n \in \omega$, $R \in \Re_n$, $x_1, \ldots, x_n \in (Vr \cup Cn)$, $t \in (V_T \cup C_T)$, such that $\varphi = \ulcorner R_t x_1, \ldots, x_n \urcorner$; or

(ii) there are $x, y \in (Vr \cup Cn)$, $t \in (V_T \cup C_T)$, such that $\varphi = \ulcorner x \underset{t}{=} y \urcorner$; or

(iii) there are $t_1, t_2 \in (V_T \cup C_T)$ such that $\varphi = \ulcorner t_1 \leq t_2 \urcorner$.

We define the set $Fm_0$ of first-order formulas in the usual way. Then we define our full set of formulas recursively as follows:

(1.2) DEFINITION:

(i) if $\varphi \in Fm_0$ then $\varphi \in Fm$;
(ii) if $\varphi \in Fm_0$ then $\ulcorner\underset{p}{\Box}\varphi\urcorner \in Fm$;
(iii) if $\varphi \in Fm_0$ then $\ulcorner\underset{a}{\Box}\varphi\urcorner \bullet Fm$;
(iv) if $\varphi, \psi \in Fm_0$ then $\ulcorner(\varphi \Rrightarrow \psi)\urcorner \in Fm$;
(v) if $\varphi, \psi \bullet Fm_0$ then $\ulcorner(\varphi \Rightarrow \psi)\urcorner \in Fm$;
(vi) if $\varphi \in Fm$ then $\ulcorner\sim\varphi\urcorner \in Fm$;
(vii) if $\varphi \in Fm$ then $\ulcorner\Box\varphi\urcorner \in Fm$;
(viii) if $\varphi \in Fm$ and $\alpha \in Vr$, then $\ulcorner(\alpha)\varphi\urcorner \in Fm$;
(ix) if $\varphi, \psi \in Fm$ then $\ulcorner(\varphi \ \& \ \psi)\urcorner \bullet Fm$;
(x) if $\varphi, \psi \in Fm$ then $\ulcorner(\varphi > \psi)\urcorner \in Fm$.

In defining 'open formula' and 'sentence' ('closed formula'), we proceed in the usual manner except that '$\Rrightarrow$', '$\Rightarrow$', and correspondingly and '$\underset{p}{\Box}$' and '$\underset{a}{\Box}$' are treated as variable binding operators, binding all of the variables in the formula to which they attach.[1] Sn is the set of sentences. We define '$\exists$', '$\vee$', '$\supset$', '$\equiv$', '$\Diamond$', '$\underset{p}{\Diamond}$', '$\underset{a}{\Diamond}$', '$\gg$', '$M$', and '$E$' in the normal ways. We define $\ulcorner(\varphi \rightarrow \psi)\urcorner$ to be $\ulcorner\Box(\varphi \supset \psi)\urcorner$ and $\ulcorner(\varphi \leftrightarrow \psi)\urcorner$ to be $\ulcorner\Box(\varphi \equiv \psi)\urcorner$. As before, $\ulcorner\forall\varphi\urcorner$ is the universal closure of $\varphi$, and let $\ulcorner\exists\varphi\urcorner$ be the existential closure (i.e., $\ulcorner\sim\forall\sim\varphi\urcorner$). We also define:

(1.3) If $G$ is a set of subjunctive generalizations, then $\forall G = \{\ulcorner\forall(\varphi \supset \psi)\urcorner; \ulcorner(\varphi \Rrightarrow \psi)\urcorner \in G \text{ or } \ulcorner(\varphi \Rightarrow \psi)\urcorner \in G\}$.

$\forall G$ is the set of material generalizations corresponding to the subjunctive generalizations in $G$.

(1.4) $\varphi$ is *indicative* iff $\varphi \in Fm_0$. *Ind* is the set of indicative sentences.

We call '$\underset{p}{\Box}$', '$\underset{a}{\Box}$', '$\Rrightarrow$', '$\Rightarrow$', and '$>$' the *subjunctive operators.*

(1.5) $SD(\varphi)$, the *subjunctive depth* of $\varphi$, is the maximum number of nested subjunctive operators in $\varphi$.

Our definition of truth will be recursive on the subjunctive depth of a sentence.

In our definition of the relation **M** we will have to make use of internal negations, so now we must give a precise definition of that notion. It is simple to do this in the case of atomic sentences:

(1.6) If $\varphi$ is an atomic sentence and $a_1, \ldots, a_n$ are the individual constants occurring in $\varphi$, then

$$\ulcorner-\varphi\urcorner = \ulcorner(\exists x)x = a_1 \;\&\; \ldots \;\&\; (\exists x)x = a_n \;\&\; \sim\varphi\urcorner.$$

It is more difficult to extend this definition to the case of sentences in general. The definition must be recursive on the length of the sentence. At this point it is not worth the effort to carry this out, because it will turn out that we only need internal negation for atomic sentences.

## 2. Semantics

Our semantics will be based upon the semantics for quantified modal logic developed in Chapter V. As usual, validity will be defined as truth in all models, and a model will consist of an interpretation of the language and a specification of what possible world is the actual world. As in the case of modal logic, an interpretation of the language will determine entailment relations between sentences by specifying the set of allowable structures. Structures are more complicated than they were in Chapter V because our language is now a temporal language. We represent times by real numbers. Letting $Rl$ be the set of real numbers:

(2.1) A *structure* is an ordered pair $\langle D, \mu \rangle$ such that $D$ is a countable set of objects and $\mu$ is a function and:
(i) for each $t \in Rl$, $\mu(t)$ is a set $D_t$;
(ii) $D = \bigcup_{t \in Rl} D_t$;
(iii) for each $t \in Rl$, $n \in \omega$, and $R \in \Re_n$, $\mu(R, t) \subseteq D_t^n$.
(iv) for each $t \in C_T$, $\mu(t) \in Rl$.

In a structure $\langle D, \mu \rangle$, $D_t$ is that subset of the domain consisting of objects existing at time $t$. If $t \in C_T$, then $\mu(t)$ is the time denoted by $t$. Then we define:

(2.2) An *interpretation* is a set $H$ of structures such that for each $n \in \omega$ and $t \in Rl$, there is a structure $\langle D, \mu \rangle$ in $H$ for which $D_t$ has cardinality $n$.

In this definition of an interpretation we have extended our variety of cardinalities requirement to the domains at each instant of time. This is because in our uninterpreted temporal language we can express contingent propositions that we could not express in the language of Chapter V. These are propositions describing the size of the domain at a particular time. An interpretation of the uninterpreted symbols of our language cannot turn such propositions into necessary truths, so if the notion of an interpretation is to capture this intuitive notion our strengthened variety of cardinalities requirement is required.

Given an interpretation of the language, we can define the notion of

a possible world relative to that interpretation. If we were to follow strictly the treatment of possible worlds contained in Chapter V, we would require possible worlds to contain four pieces of information: (1) a structure $S$; (2) a counterpart relation relating $S$ to the real world; (3) a specification of what basic strong generalizations are true; (4) a specification of what basic weak generalizations are true. However, an inspection of the definition of truth for quantified modal logic reveals that the counterpart relation plays only a heuristic role and does not enter into any of the formal definitions. It is replaced there by the function $\eta$ which assigns denotations to individual constants. The same thing will be true in our present language. In Chapter V $\eta$ was taken to be part of the interpretation of the language, but I now propose to include it in the possible world, using it to replace the counterpart relation. Thus a possible world will become an ordered quadruple $\langle S, \eta, N, W \rangle$.

Normal procedure would be to require that $\eta$ maps a subset of $Cn$ into the domain $D$ of the structure $S$. This was the procedure followed in Chapter V. However, in the present context we must require instead that $\eta$ map a countable subset of $Cn$ *onto* $D$. We must require every object to have a name. The reason for this is that in our definition of the relation **M** we want to talk about the set of *all* simple truths in a world, and if we are to do this by talking metalinguistically about what simple sentences are true in the world, we must have every simple truth expressed by some simple sentence. This can only be done if every object has a name. This is also the reason we require our language to have non-denumerably many individual constants. Otherwise we could have a world $\alpha$ which uses them all up, in which case there could be no worlds $\beta$ possible relative to $\alpha$ (i.e., 'accessible' in the sense of 2.9) containing all of the objects of $\alpha$ and some new objects in addition.

We cannot define a possible world to be just any quadruple $\langle S, \eta, N, W \rangle$ of the appropriate sort. A possible world must be 'coherent' in the sense that if a subjunctive generalization is included in $N$ or $W$, then the corresponding material generalization is true in the world. This seems to put us in the position of having to define truth in a model before we can define the notion of a model, which would quickly become circular. Fortunately, there is no real difficulty here.

The ordered pair $\langle S, \eta \rangle$ constitutes a first-order model for the indicative fragment of our language, and we can define the set $T$ of true indicative sentences in the normal way in terms of $\langle S, \eta \rangle$.

It was argued in Chapter III that a logically contingent generalization of the form $\ulcorner(x)(\varphi \supset .x = a_1 \vee \ldots \vee x \underset{t}{=} a_n)\urcorner$ cannot be actually necessary. This constitutes a constraint on the sets $N$ and $W$. Combining this with the previous observations leads to the following definitions:

(2.3) A *possible world* relative to an interpretation $H$ is an ordered quadruple $\langle S, \eta, N, W \rangle$ such that:

(i) $S$ is a structure $\langle D, \mu \rangle$;
(ii) $S \in H$;
(iii) $\eta$ maps a countable subset of $Cn$ onto $D$;
(iv) $N$ is a set of strong generalizations;
(v) $W$ is a set of weak generalizations;
(vi) If $T$ is the set of indicative truths, then $\forall N \subseteq \forall W \subseteq T$, and there is no indicative formula $\varphi$, individual constants $a_1, \ldots, a_n$, and $t \in C_T$ such that $\forall W$ entails $\ulcorner(x)(\varphi \supset .x \underset{t}{=} a_1 \vee \ldots \vee x \underset{t}{=} a_n)\urcorner$ but $\ulcorner\Box(x)(\varphi \supset .x \underset{t}{=} a_1 \vee \ldots \vee x \underset{t}{=} a_n)\urcorner$ is false in $\langle S, \eta, N, W \rangle$.

Clause (vi) looks as though it might be circular because it refers to the truth of a modal sentence in the possible world. However, there is no circularity because the truth value of such a sentence will be independent of the contents of $N$ and $W$.

(2.4) A *model* is an ordered pair $\langle \alpha, I \rangle$ where $I$ is an interpretation and $\alpha$ is a possible world relative to $I$.

We will define the notion of truth in a model, and then:

(2.5) $\varphi$ is *valid* iff $\varphi$ is true in every model.

Truth will be truth in a model, but for convenience we will also allow ourselves to talk about a sentence being true in a possible world (relative to an interpretation). For convenience we define the following notation:

(2.6) If $\alpha = \langle S, \eta, N, W\rangle$, then $S_\alpha = S$, $\eta_\alpha = \eta$, $N_\alpha = N$, and $W_\alpha = W$.

(2.7) If $\alpha$ is a possible world, $T_\alpha$ is the set of indicative truths in $\alpha$.

(2.8) If $I$ is an interpretation, $[[I]] = \{\alpha;\ \alpha$ is a possible world relative to $I\}$.

In defining the notion of a possible world, we have ignored the relation of the possible world to the actual world. The result is that we will have more possible worlds than we really want. A world is not 'really possible' unless there is a counterpart function from the real world into the domain of the possible world such that the denotation function $\eta$ 'agrees' with the counterpart function. More precisely, let us define:

(2.9) If $I$ is an interpretation and $\alpha, \beta \in [[I]]$, $\beta$ is accessible from $\alpha$ iff for every $b$, $c \in \mathscr{D}(\eta_\alpha)$, if $\eta_\alpha(b) = \eta_\alpha(c)$ and $b \bullet \mathscr{D}(\eta_\beta)$, then $\eta_\beta(b) = \eta_\beta(c)$.

(2.10) If $\alpha \in [[I]]$, $[[I]]_\alpha = \{\beta;\ \beta \in [[I]]$ and $\beta$ is accessible from $\alpha\}$.

If $\alpha$ is the real world, then only worlds accessible from $\alpha$ are 'really possible'.

What remains is to define truth in a model. This can be accomplished by generalizing Analysis VII. We will define truth and the relation $\mathbf{M}$ by simultaneous recursion on the subjunctive depth of a sentence. Up to this point we have taken $\mathbf{M}$ to be a binary relation between a possible world and a sentence. But for a general semantics, we must construe it to be a ternary relation. If $\alpha$ and $\beta$ are possible worlds relative to some fixed interpretation, '$\beta \mathbf{M}_\alpha \varphi$' means 'If $\alpha$ were the real world, then $\beta$ would be a world that might be actual if $\varphi$ were true'. We need this ternary relation because we want to define the truth of a subjunctive conditional in any world and not just in the actual world.

Consider now how to define the relation '$\beta \mathbf{M}_\alpha \varphi$'. The first requirement is obvious, and is the same as in Analysis VII:

(i) $\varphi$ is true in $\beta$.

Next we must consider what is required of $N_\beta$. In Analysis VII we required that $N_\beta$ be the result of making a minimal $P$-change to $N_\alpha$.

This is still correct, however our earlier definition of the notion of a minimal $P$-change is no longer sufficient for two reasons. First, what we really wanted in Analysis VII was to ensure that $N_\beta$ was the result of making minimal changes to $N_\alpha$ in order to render it consistent with $P$. We could ensure that as we did by talking about $\forall N_\beta$ because we were only considering the case in which $\varphi$ was indicative. In that case, $\varphi$ is consistent with $N_\beta$ iff it is consistent with $\forall N_\beta$. But in the general case, where we can no longer assume that $\varphi$ is indicative, we cannot get by talking about $\forall N_\beta$. For example, $\varphi$ might be the negation of some member of $N_\alpha$. In that case, $\varphi$ would still be consistent with $\forall N_\alpha$. We must instead talk directly about the consistency of $N_\alpha$ with $P$. There is a second difficulty. As long as we were only considering an indicative $\varphi$, the only way the supposition that $\varphi$ is true could require us to alter $N_\alpha$ was by being inconsistent with $N_\alpha$, and hence the only kind of change that could be required was a deletion. But if $\varphi$ can be non-indicative, then the supposition that $\varphi$ is true may require us to enlarge $N_\alpha$. For example, $\varphi$ might be of the form $\ulcorner \psi \Rrightarrow \theta \urcorner$, in which case we would have to add generalizations to $N_\alpha$ which would be sufficient to allow us to derive $\varphi$ from the enlarged set. Thus in the general case, there are two kinds of changes that might be required in constructing $N_\beta$. We might have to eliminate some members of $N_\alpha$, and we might also have to add some new generalizations not already in $N_\alpha$. In either case, we want to require that the changes be as small as possible. Just as in the case of changes to simple sentences, we can represent the change to $N_\alpha$ by the indexed difference $(N_\alpha \triangle N_\beta)$. We can then take definitions 2.20 and 2.21 of Chapter IV to define the notion of a minimal change (replacing sets of propositions now by sets of sentences) if we supply a new definition of the notion of a strictly minimal change which will make the latter notion applicable to $N_\alpha$. This is readily accomplished:

(2.11) $X$ is a strictly minimal $\varphi$-change to $N_\alpha$ (relative to $I$ and $\alpha$) iff there is a world $\beta \in [[I]]_\alpha$ in which $\varphi$ is true such that $N_\beta = N_\alpha + X$, and there is no $Y$ such that $Y \subset X$ and there is a world $\gamma \in [[I]]_\alpha$ in which $\varphi$ is true such that $N_\gamma = N_\alpha + Y$.

We then require:

(ii) $(N_\alpha \triangle N_\beta)$ is a minimal $\varphi$-change to $N_\alpha$.

Analogous considerations apply to $W_\beta$. We now need the notion of a minimal $\Gamma$-change (where $\Gamma$ is a set of sentences), and we must relativize our notion of a minimal change to our choice of $N_\beta$. This can again be obtained from definitions 2.20 and 2.21 of Chapter IV if we extent the notion of a strictly minimal change as follows:

(2.12) $X$ is a strictly minimal $\Gamma$-change to $W_\alpha$ (relative to $N_\beta$) iff there is a world $\gamma \in [[I]]_\alpha$ in which all members of $\Gamma$ are true and such that $N_\gamma = N_\beta$ and $W_\gamma = W_\alpha + X$; and there is no $Y$ such that $Y \subset X$ and there is a world $\delta \in [[I]]_\alpha$ in which all members of $\Gamma$ are true and such that $N_\delta = N_\beta$ and $W_\delta = W_\alpha + Y$.

We then require:

(iii) $(W_\alpha \triangle W_\beta)$ is a minimal $(\forall N_\beta \cup \{\varphi\})$-change to $W_\alpha$ (relative to $N_\beta$).

We seek to preserve the truth values of simple sentences and their internal negations, as in Chapter IV. Simple sentences are now identified with the atomic sentences, so let us define:

(2.13) $Sim = \{\varphi; \varphi \bullet At \cap Sn$ or $(\exists \psi)(\psi \in At \cap Sn$ and $\varphi = \ulcorner - \psi \urcorner)\urcorner$.

Given an atomic formula $\varphi$ whose temporal subscript is $\tau$, we will define $\sigma(\ulcorner - \varphi \urcorner) = \sigma(\varphi) = \mu(\tau)$. Then we define, for $t \in Rl$:

(2.14) $S_\alpha(t) = \{\varphi; \varphi \in (Sim \cap T_\alpha)$ and $\sigma(\varphi) \leq t\}$.

We want to treat simple sentences just as we did in Analysis VII, giving as our final analysis:

(2.15) DEFINITION: $\beta \mathbf{M}_\alpha \varphi$ (relative to an interpretation $I$) iff $\beta \in [[I]]_\alpha$ and:

(i) $\varphi$ is true in $\beta$;

(ii) $(N_\alpha \triangle N_\beta)$ is a minimal $\varphi$-change to $N_\alpha$;

(iii) $(W_\alpha \triangle W_\beta)$ is a minimal $(\forall N_\beta \cup \{\varphi\})$-change to $W_\alpha$ (relative to $N_\beta$);

(iv) for every $t \in Rl$, $(S_\alpha(t) \triangle S_\beta(t))$ is a minimal $(\forall N_\beta \cup \forall W_\beta \cup \{\varphi\})$-change to $T_\alpha$ at time $t$.

With this definition, we have completed out account of the relation **M**.

Definition 2.15 makes use of the notion of $\varphi$ being true in a world $\gamma$. Now we want to define that notion. For subjunctive $\varphi$, the definition will use the relation $\mathbf{M}$. We want these two definitions to constitute a definition by simultaneous recursion on the subjunctive depth of $\varphi$. In order for this to work, we must ensure that in defining truth for $\varphi$, we only use the relation $\ulcorner\beta\mathbf{M}_\alpha\psi\urcorner$ for formulas $\psi$ of subjunctive depth less than that of $\varphi$.

It is obvious how to define truth for indicative sentences, and for conjunctions, negations, universal generalizations, modal sentences, and subjunctive conditionals. The first case that is not entirely obvious is that of a formula of the form $\ulcorner\underset{p}{\Box}\varphi\urcorner$. Intuitively, if $\varphi$ is a sentence then $\ulcorner\underset{p}{\Box}\varphi\urcorner$ is true in a world $\alpha$ (relative to an interpretation $I$) iff $\varphi$ is entailed by $N_\alpha$. This can be captured by requiring that $\varphi$ be true in every world $\beta$ for which $N_\beta$ includes at least $N_\alpha$. Recalling that if $\varphi$ is an open formula then $\ulcorner\underset{p}{\Box}\varphi\urcorner$ is supposed to mean the same thing as $\ulcorner\underset{p}{\Box}\forall\varphi\urcorner$, we have the general condition:

> $\ulcorner\underset{p}{\Box}\varphi\urcorner$ is true in $\alpha$ (relative to $I$) if $(\forall\beta\in[[I]]_\alpha)$[if $N_\alpha\subseteq N_\beta$ then $\ulcorner\forall\varphi\urcorner\in T_\beta$].

Truth for $\ulcorner\underset{p}{\Box}\varphi\urcorner$ is defined analogously.

A generalization $\ulcorner\varphi\Rrightarrow\psi\urcorner$ is supposed to be true in a world $\alpha$ iff for each world $\beta$ such that $\beta\mathbf{M}_\alpha\exists\varphi$, the set $\forall N_\beta$ entails $\ulcorner\forall(\varphi\supset\psi)\urcorner$. This can be captured by the following definition:

> $\ulcorner\varphi\Rrightarrow\psi\urcorner$ is true in $\alpha$ (relative to $I$) iff $(\forall\beta,\gamma\in[[I]]_\alpha)$[if $\beta\mathbf{M}_\alpha\exists\varphi$ and $\forall N_\beta\subseteq T_\gamma$ then $\ulcorner\forall(\varphi\supset\psi)\urcorner\in T_\gamma$].

Weak subjunctive generalizations are handled analogously. Putting these observations together into a definition, we obtain:

(2.16) If $t\in C_T$ and $\langle\alpha, H\rangle$, $\langle\beta, I\rangle$ are models, then $\langle\beta, I\rangle$ is a *t-variant* of $\langle\alpha, H\rangle$ iff $\langle\alpha, H\rangle$ and $\langle\beta, I\rangle$ are identical except possibly for the value of $\eta(t)$.

(2.17) DEFINITION: Truth in $\alpha$ relative to $I$:

(i) If $b,c\in Cn$ and $t\in C_T$ then $\ulcorner b\underset{t}{=}c\urcorner$ is true iff $b, c\in\mathscr{D}(\eta_\alpha)$ and $\eta_\alpha(b)=\eta_\alpha(c)$ and $\eta_\alpha(b)\in D_{\alpha,\eta(t)}$;

(ii) if $t_1, t_2 \in C_T$ then $\ulcorner t_1 \leq t_2 \urcorner$ is true iff $\eta(t_1) \leq \eta(t_2)$;

(iii) if $R \in \mathfrak{R}_n$, $t \in C_T$, and $a_1, \ldots, a_n \in C_n$, then $\ulcorner R_t a_1 \ldots, a_n \urcorner$ is true iff $a_1, \ldots, a_n \in \mathscr{D}(\eta_a)$ and $\langle \eta_\alpha(a_1), \ldots, \eta_\alpha(a_n) \rangle \in \mu(R, t)$;

(iv) $\ulcorner \sim \varphi \urcorner$ is true iff $\varphi$ is not true;

(v) $\ulcorner (\varphi \ \& \ \psi) \urcorner$ is true iff $\varphi$ and $\psi$ are both true;

(vi) if $x \in Vr$, and $\ulcorner \varphi(c/x) \urcorner$ is the result of substituting a constant $c$ for every free occurrence of $x$ in $\varphi$, then $\ulcorner (x)\varphi \urcorner$ is true in $\alpha$ iff for every $c \in \mathscr{D}(\eta)$, $\ulcorner \varphi(c/x) \urcorner$ is true;

(vii) if $t \in V_T$, $\tau \in C_T$ and $\tau$ does not occur in $\varphi$ and $\ulcorner \varphi(\tau/t) \urcorner$ is the result of substituting $\tau$ for every free occurrence of $t$ in $\varphi$, then $\ulcorner (t)\varphi \urcorner$ is true in $\alpha$ iff $\ulcorner \varphi(\tau/t) \urcorner$ is true (relative to $H$) in every $\beta$ such that $\langle \beta, H \rangle$ is a $\tau$-variant of $\langle \alpha, I \rangle$;

(viii) $\ulcorner \Box \varphi \urcorner$ is true in $\alpha$ iff $\varphi$ is true in every member of $[[I]]_\alpha$;

(ix) $\ulcorner \underset{p}{\Box} \varphi \urcorner$ is true in $\alpha$ iff $(\forall \beta \in [[I]])_\alpha$[if $N_\alpha \subseteq N_\beta$ then $\ulcorner \forall \varphi \urcorner \in T_\beta$];

(x) $\ulcorner \underset{a}{\Box} \varphi \urcorner$ is true in $\alpha$ iff $(\forall \beta \in [[I]])_\alpha$ [if $W_\alpha \subseteq W_\beta$ then $\ulcorner \forall \varphi \urcorner \in T_\beta$];

(xi) $\ulcorner (\varphi \Rrightarrow \psi) \urcorner$ is true in $\alpha$ iff $(\forall \ \beta, \gamma \in [[I]]_\alpha)$[if $\beta \mathbf{M}_a \exists \varphi$ and $\forall N_\beta \subseteq T_\gamma$ then $\ulcorner \forall (\varphi \supset \psi) \urcorner \in T_\gamma$]

(xii) $\ulcorner (\varphi \Rrightarrow \omega) \urcorner$ is true in $\alpha$ iff $(\forall \beta, \ \gamma \in [[I]]_\alpha)$ [if $\beta \mathbf{M}_\alpha \exists \varphi$ and $\forall W_\beta \subseteq T_\gamma$ then $\ulcorner \forall (\varphi \supset \psi) \urcorner \in T_\gamma$];

(xiii) $\ulcorner (\varphi > \psi) \urcorner$ is true in $\alpha$ iff $(\forall \beta \in [[I]]_\alpha)$ [if $\beta \mathbf{M}_\alpha \varphi$ then $\psi$ is true in $\beta$].

Definitions 2.15 and 2.17 constitute definitions by simultaneous recursion on the subjunctive depth of a formula.

## 3. Infinitary operators

It is customary to define logical entailment both as an object language relation between sentences and as a metalinguistic relation between

sets of sentences and sentences. We define $\ulcorner \varphi \rightarrow \psi \urcorner$ to mean that $\ulcorner \varphi \supset \psi \urcorner$ is necessarily true. However, where $\Gamma$ is an infinite set of sentences, we cannot similarly define '$\Gamma \rightarrow \psi$', because there is no way to talk about the set $\Gamma$ in the object language. Thus we retreat into the metalanguage and define it to mean that it is necessarily true that if all the sentences in $\Gamma$ are true then $\psi$ is true.

In the next chapter, where we construct an analysis of causal statements, we will find ourselves in a similar position with respect to subjunctive conditionals. We will find it necessary to employ subjunctive conditionals whose antecedents and consequents are infinite conjunctions or disjunctions. When $\Gamma$ is a finite set, then the disjunction '$\Sigma\Gamma$' and the conjunction '$\Pi\Gamma$.' of the members of $\Gamma$ make perfectly good sense, and we can express statements like $\ulcorner \Sigma\Gamma > \psi \urcorner$ or $\ulcorner \Pi\Gamma > \psi \urcorner$ in the object language. However, if $\Gamma$ is infinite, this can no longer be done. Our language does not contain infinite conjunctions or disjunctions, so we must retreat into the metalanguage once more. We can straightforwardly define the metalinguistic relation '$\underset{\alpha}{>}$' between single sentences as follows:

(3.1) $\varphi \underset{\alpha}{>} \psi$ iff $\ulcorner \varphi > \psi \urcorner$ is true in the world $\alpha$.

Then if $\Gamma$ is finite, we can write things like '$\Pi\Gamma \underset{\alpha}{>} \psi$' and '$\Sigma\Gamma \underset{\alpha}{>} \psi$'. But we want to extend our definition of this metalinguistic relation to the case in which $\Gamma$ is infinite. We want, ultimately, to be able to write all of the following:

(a) $\Pi\Gamma \underset{\alpha}{>} \psi$

(b) $\Sigma\Gamma \underset{\alpha}{>} \psi$

(c) $\varphi \underset{\alpha}{>} \Sigma\Gamma$

(d) $(\varphi \ \& \ \Sigma\Gamma) \underset{\alpha}{>} \psi$

(e) $(\varphi \ \& \ \sim\Sigma\Gamma) \underset{\alpha}{>} \psi$

(f) $(\varphi \ \& \ \Pi\Gamma) \underset{\alpha}{>} \psi$

(g) $\left(\varphi \ \& \ \prod_{\beta<\gamma} \sim\Sigma\Gamma_\beta\right) \underset{\alpha}{>} \psi$

(h) $\left(\varphi \;\&\; \prod_{\beta<\gamma} \Sigma\Gamma_\beta\right) > \Sigma\Lambda$

(i) $\left(\varphi \;\&\; \prod_{\beta<\gamma} \Sigma\Pi_\beta\right) > \Sigma\Lambda$

and other similar metalinguistic statements. Unfortunately, it does not appear to be possible to give a definition of a single relation '$>$' which will then give us all of the above as special cases. The difficulty is that, e.g., (b) does not express a relation between two objects $\Sigma\Gamma$ and $\psi$. '$\Sigma\Gamma$' is not an object. $\Gamma$ is a set, but as we don't really have infinite disjunctions, '$\Sigma\Gamma$' isn't anything. We must instead treat the whole expression '$\Sigma \ldots > \ldots$' as expressing a metalinguistic relation between $\Gamma$ and $\psi$. Each of (a)–(i) expresses a different metalinguistic relation and must be defined separately. This could be overcome by formalizing our metalanguage and then giving a recursive definition in the meta-metalanguage, but that is not worth the trouble. Once we have seen how to define one of the relations (a)–(i), it will be quite obvious how to define any of the others. Thus once we have gone through the definition of one of these, I will feel free to use other similar relations without actually defining them.

The key to constructing definitions for relations like (a)–(i) is to extend the definition of $\mathbf{M}$ so that we can write things like '$\beta\mathbf{M}_\alpha\Sigma\Gamma$' and '$\beta\mathbf{M}_\alpha\Pi\Gamma$'. It is trivial to do this. We simply go back to definition 2.15 and wherever that definition requires that the sentence $\varphi$ be true in some possible world, we instead require that there is a member of $\Gamma$ true in that world (in the case of $\Sigma\Gamma$) or that every member of $\Gamma$ is true in that world (in the case of $\Pi\Gamma$). In the case of each of the antecedents of (a)–(i) there is a similarly obvious condition to employ in the definition of $\mathbf{M}$. Then we can define our relation (a)–(i) quite straightforwardly. For example, we define (a) as:

> for every world $\beta$, if $\beta\mathbf{M}_\alpha\Pi\Gamma$ then $\psi$ is true in $\beta$.

Analogously, we define (h) as:

> for every world $\beta$, if $\beta\mathbf{M}_\alpha(\varphi \;\&\; \prod_{\zeta<\gamma} \Sigma\Gamma_\zeta)$, then some member of $\Lambda$ is true in $\beta$.

Using these metalinguistic relations we can write anything we could write if our language actually contained infinite conjunctions and disjunctions, the only difference being that what we write are expressions of the metalanguage rather than the object language. In addition there is the advantage that whereas one could raise philosophical difficulties about infinite conjunctions and disjunctions in the object language, similar difficulties do not arise for our metalinguistic procedures.

We will also want to use infinitary operators in connection with necessitation conditionals, writing things like '$\Pi\Gamma \underset{\alpha}{\gg} \psi$'. We can define this quite straightforwardly in terms of our use of infinitary operators with the simple subjunctive. The object language necessitation conditional $\ulcorner \varphi \gg \psi \urcorner$ is defined to be:

$$(\varphi > \psi) \ \& \ [(\sim\varphi \ \& \ \sim\psi) > (\varphi > \psi)].$$

Thus we would like to define '$\Pi\Gamma \underset{\alpha}{\gg} \psi$' to mean something like:

$$(\Pi\Gamma \underset{\alpha}{\gg} \psi) \ \& \ [(\sim\Pi\Gamma \ \& \ \sim\psi) \underset{\alpha}{>} (\Pi\Gamma > \psi)].$$

We cannot write quite this, because the final occurrence of '$>$' is not subscripted. But it is obvious what we want to say here. We want to say that '$\Pi\Gamma \underset{\beta}{\geq} \psi$' holds for every $(\sim\Pi\Gamma \ \& \ \sim\psi)$-world $\beta$. Thus our definition should be:

$$(\Pi\Gamma \underset{\alpha}{\geq} \psi) \ \& \ (\forall\beta)[\text{if } \beta \mathbf{M}_\alpha(\sim\Pi\Gamma \ \& \ \sim\psi) \text{ then } \Pi\Gamma \underset{\beta}{\geq} \psi].$$

Given our infinitary simple subjunctives, we can define our infinitary necessitation conditionals completely generally on the above model:

(3.1) $(\ ) \underset{\alpha}{\gg} [\ ]$ iff $(\ ) \underset{\alpha}{\geq} [\ ]$ and $(\forall\beta)\{$if $\beta\mathbf{M}_\alpha(\sim(\ ) \ \& \ \sim[\ ])$ then $(\ ) \underset{\beta}{\geq} [\ ]\}$.

## 4. The introduction of sets

One of the most important tasks of this chapter will be to prove theorems 4.20 and 4.21 of Chapter III. However, these principles cannot yet be formulated in our language. To do this we must augment

our language with some set-theoretic notation and extend our semantics to handle this additional richness. This is not difficult to do. We add a new non-denumerable class $V_S$ of set variables, and a non-denumerable class $C_S$ of set constants, together with the new logical constant '$\in$'. We then turn our language into a three-sorted language. We add the new atomic formulas $\ulcorner x \underset{t}{\in} X \urcorner$ and $\ulcorner X \underset{t}{=} Y \urcorner$ where $x \in (Cn \cup VR)$, $t \in (C_T \cup V_T)$, and $X, Y \in (C_S \cup V_S)$. Our other syntactical definitions remain as before, with the exception that we do not allow our set notation to be used in the formulation of subjunctive generalizations. Nor do we allow our set notation to be used in constructing simple sentences. The set *Sim* remains as before.

In formulating the semantics for our extended language, we define 'structure' and 'interpretation' just as before. We must modify the definition of 'possible world' by requiring that $\eta$ interpret the set constants in addition to the individual constants. Thus we add the condition:

(iii*) $\eta$ maps a countable subset of $C_S$ into the power set of $D$.

We require that $\eta$ only interpret countably many set constants in order to ensure that there are always plenty of set constants left over to denote new sets occurring in worlds possible relative to $\alpha$. This requirement has the effect that if $D$ is denumerable, then most sets will not receive names. Thus in defining truth for sentences involving quantification over sets, we must proceed conventionally in terms of $X$-variants. We define truth for our new atomic sentences as follows:

(4.1) (i) If $a \in Cn$, $t \in C_T$, and $A \in C_S$, then $\ulcorner a \in_t A \urcorner$ is true in $\alpha$ iff $a \in \mathscr{D}(\eta_\alpha)$, $A \in \mathscr{D}(\eta_\alpha)$, $\eta_\alpha(A) \subseteq D_{\alpha,\eta(t)}$, and $\eta_\alpha(a) \in \eta_\alpha(A)$;

(ii) if $A, B \in C_S$ and $t \in C_T$ then $\ulcorner A =_t B \urcorner$ is true in $\alpha$ iff $A, B \in \mathscr{D}(\eta_\alpha)$, $\eta_\alpha(A) = \eta_\alpha(B)$, and $\eta_\alpha(A) \subseteq D_{\alpha,\eta(t)}$.

Our definition of accessibility must be made more complicated to accommodate our set notion. We want the transworld identity of a set to be determined by the transworld identity of its members. Thus if $\beta$ is a world accessible from $\alpha$, and $A \in C_S$, then the set denoted by $A$ in $\beta$ must consist of the counterparts of the members of the set denoted

by $A$ in $\alpha$. Thus our definition becomes:

(2.9*) If $I$ is an interpretation and $\alpha, \beta \in [[I]]$, $\beta$ is accessible from $\alpha$ iff:

(i) for every $b, c \in Cn$, if $b, c \in \mathscr{D}(\eta_\alpha)$, $\eta_\alpha(b) = \eta_\alpha(c)$ and $b \in \mathscr{D}(\eta_\beta)$ then $\eta_\beta(b) = \eta_\beta(c)$; and

(ii) for every $A \in C_S$, if $A \in \mathscr{D}(\eta_\alpha)$ then $A \in \mathscr{D}(\eta_\beta)$ iff $\{c; c \in Cn \ \& \ \eta_\alpha(c) \in \eta_\alpha(A)\} \subseteq \mathscr{D}(\eta_\beta)$, and if $A \in \mathscr{D}(\eta_\beta)$ then $\eta_\beta(A) = \{\eta_\beta(c); \ c \in (Cn \cap \mathscr{D}(\eta_\alpha))$ and $\eta_\alpha(c) \in \eta_\alpha(A)\}$.

The rest of our semantics can be left unchanged. One noteworthy theorem which results immediately from our interpretation of set terms is the following:

(4.2) $A =_{t} B \supset \Box((\exists X) X =_{t^*} A \supset A =_{t^*} B)$.

That is, set terms are rigid designators.

## 5. Some consequences of the analysis

Definitions 2.15 and 2.17 constitute both an analysis and a formal semantics for subjunctive conditionals. A number of interesting theorems result from this analysis. The turnstile is defined as usual:

(5.1) $\vdash \varphi$ iff $\varphi$ is valid.

Our logic of subjunctive conditionals contains the quantified modal logic of Chapter V. Once again, the propositional fragment of our modal logic is $S4$. As noted before, a necessary condition for the correctness of the above analysis is that it have the consequence that every $P$-change contains a minimal $P$-change. In the present context, this should become a provable metatheorem. However, the extreme complexity of our language makes it very difficult to prove anything about it, and I can do no more than conjecture that this metatheorem holds. Assuming that it does, the necessity operator can be defined in terms of '$>$':

(5.2) $\vdash \Box\varphi \equiv (\sim\varphi > \varphi)$.

The axioms of *SS* are all valid:

(5.3) $\vdash(\varphi \ \& \ \psi) \supset (\varphi > \psi)$.

(5.4) $\vdash(\varphi \rightarrow \psi) \supset (\varphi > \psi)$.

(5.5) $\vdash[(\varphi > \psi) \ \& \ (\varphi > \theta)] \supset [\varphi > (\psi \ \& \ \theta)]$.

(5.6) $\vdash[(\varphi > \theta) \ \& \ (\psi > \theta)] \supset [(\varphi \vee \psi) > \theta]$.

(5.7) $\vdash[(\varphi > \psi) \ \& \ (\psi \rightarrow \theta)] \supset (\varphi > \theta)$.

(5.8) $\vdash(\varphi > \psi) \supset (\varphi \supset \psi)$.

(5.9) $\vdash[(\varphi \leftrightarrow \psi) \ \& \ (\psi > \theta)] \supset (\varphi > \theta)$.

(5.10) $\vdash[(\varphi > \psi) \ \& \ (\varphi > \theta)] \supset [(\varphi \ \& \ \psi) > \theta)]$.

At the end of Chapter I, it was suggested that there is a closer affinity between the linguistic theory of conditionals and the possible worlds theory than might be supposed. This can now be demonstrated. According to the linguistic theory, $\ulcorner(\varphi > \psi)\urcorner$ is true just in case $\psi$ is entailed by a combination of $\varphi$, some laws, and some truths satisfying an as yet unanalyzed condition of 'contenability with $\varphi$'. I suggested in Chapter I that the proper analysis of $\ulcorner\theta$ is contenable with $\varphi\urcorner$ is $\ulcorner\theta E\varphi\urcorner$. Of course, this does not give an analysis of subjunctive conditionals until we have an analysis of 'even if', but that it is right as far as it goes is indicated by the following theorem:

(5.11) $\ulcorner(\varphi > \psi)\urcorner$ is true in a model iff for some sentence $\theta$, $\ulcorner\{[(\varphi \ \& \ \theta) \Rrightarrow \psi] \ \& \ (\theta E\psi)\}\urcorner$ is true in the model.

Furthermore, Analysis VII proceeds by telling us what must be preserved in a $\varphi$-world, and so is in effect an analysis of 'even if'. Thus in the end, the linguistic and possible worlds approaches to the analysis of subjunctive conditionals converge.

Our logic is also a logic of law statements and physical necessity. All of the principles 3.15–3.27 and 4.5–4.16 of Chapter III are valid. A few other simple principles are:

(5.12) $\vdash[\underset{p}{\Box}\varphi \ \& \ \underset{p}{\Diamond}\psi] \supset [(\underset{p}{\Box}\varphi)E\psi]$.

(5.13) $\vdash[\underset{a}{\Box}\varphi \ \& \ \underset{a}{\Diamond}\psi] \supset [(\underset{a}{\Box}\varphi)E\psi]$.

(5.14) $\vdash(\varphi > \psi) \;\&\; \underset{p}{\Diamond}\varphi. \supset \underset{p}{\Diamond}\psi.$

(5.15) $\vdash(\varphi > \psi) \;\&\; \underset{a}{\Diamond}\varphi. \supset \underset{a}{\Diamond}\psi.$

(5.16) $\vdash(\varphi \Rrightarrow \psi) \supset \forall(\varphi \supset \psi).$

(5.17) $\vdash\Diamond\exists(\varphi \;\&\; x_1, \ldots, x_n \notin A) \supset [\underset{a}{\Box}\forall(\varphi \supset x_1, \ldots, x_n \in A) \equiv \underset{a}{\Box}\forall\sim\varphi].$

(5.18) $\vdash\Diamond\exists(\varphi \;\&\; x_1, \ldots, x_n \notin A) \supset \Box[\underset{a}{\Diamond}\exists(\varphi \;\&\; x_1, \ldots, x_n \in A) \equiv \underset{a}{\Diamond}\exists\varphi].$

(5.19) $\vdash(\varphi > \psi) \equiv [\underset{a}{\Diamond}\varphi > (\varphi > \psi)].$

(5.20) $\vdash[(\varphi > \psi) \rightarrow \theta] \supset [(\varphi > \psi) \rightarrow (\underset{a}{\Diamond}\varphi > \theta)].$

(5.21) $\vdash(\varphi \Rrightarrow \psi) \supset \{\underset{p}{\Diamond}(\varphi \;\&\; \theta) > [(\varphi \;\&\; \theta) \Rrightarrow \psi]\}.$

In Chapter III we discussed the possibility of analyzing $\ulcorner(\varphi \Rrightarrow \psi)\urcorner$ as a universally quantified conditional of some sort, but were unable to do so because of the difficulty concerning what the quantifier would range over in the case of a counter-legal. It was suggested instead that the strong subjunctive generalization was analyzable as $\ulcorner\underset{p}{\Diamond}\exists\varphi > \underset{p}{\Box}\forall(\varphi > \psi)\urcorner$. We can now prove that this proposed characterization is correct. According to our semantics, in deciding whether $\ulcorner(\varphi \Rrightarrow \psi)\urcorner$ is true we make minimal changes to $N$ in order to render it $\ulcorner\exists\varphi\urcorner$-consistent, and then we see whether the resulting set entails $\ulcorner\forall(\varphi \supset \psi)\urcorner$. But this is the same thing as looking either at each $\ulcorner\exists\varphi\urcorner$-world or at each $\ulcorner\underset{p}{\Diamond}\exists\varphi\urcorner$-world $\alpha$ and seeing whether $N_\alpha$ entails $\ulcorner\forall(\varphi \supset \psi)\urcorner$. Thus we have both of the following equivalences:

(5.22) $\vdash(\varphi \Rrightarrow \psi) \equiv [\exists\varphi > \underset{p}{\Box}\forall(\varphi \supset \psi)].$

(5.23) $\vdash(\varphi \Rrightarrow \psi) \equiv [\underset{p}{\Diamond}\exists\varphi > \underset{p}{\Box}\forall(\varphi \supset \psi)].$

Analogously:

(5.24) $\vdash(\varphi \Rrightarrow \psi) \equiv [\exists\varphi > \underset{a}{\Box}\forall(\varphi \supset \psi)].$

(5.25) $\vdash(\varphi \Rrightarrow \psi) \equiv [\underset{a}{\Diamond}\exists\varphi > \underset{a}{\Box}\forall(\varphi \supset \psi)].$

An immediate consequence of 5.24 is:

(5.26) $\vdash(\varphi \Rightarrow \psi) \equiv [\exists\varphi > (\varphi \Rightarrow \psi)]$.

Now we are in a position to prove theorems 4.20 and 4.21 of Chapter III:

(5.27) If $\varphi$ contains no set variables or set constants, and the free individual variables of $\varphi$ are $x_1, \ldots, x_n$, then $\vdash(\varphi \Rightarrow \psi) \equiv (\exists A)\{(x)(x \in A \equiv x = x) \ \& \ [\forall(\varphi \supset \psi)E\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A)]\}$.

*Proof:* Let $A = \{x; x = x\}$. Suppose $\ulcorner\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)\urcorner$ is true in $\alpha$. Suppose $\beta \mathbf{M}_\alpha \ulcorner\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A)\urcorner$. Then in constructing $\beta$, We have added a new object or sequence of objects to the domain and made it satisfy $\varphi$. A new object or sequence of objects can have any set of simple attributes consistent with $W_\alpha$ because there are no simple truths in $\alpha$ about new objects which have to be preserved. Thus $\ulcorner(\forall\varphi \supset \psi)\urcorner$ will be true in every such $\beta$ only if $W_\alpha$ entails $\ulcorner\forall(\varphi \supset \psi)\urcorner$, i.e., $\ulcorner\underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$ is true in $\alpha$. Thus $\ulcorner\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)\urcorner$ entails $\ulcorner\underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$. Hence by 5.20, it also entails $\ulcorner\underset{a}{\Diamond}\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$. By 5.18, $\ulcorner\underset{a}{\Diamond}\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A)\urcorner$ is equivalent to $\ulcorner\underset{a}{\Diamond}\exists\varphi\urcorner$. Consequently, $\ulcorner\underset{a}{\Diamond}\exists\varphi > \underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$is true in $\alpha$, i.e., $\ulcorner(\varphi \Rightarrow \psi)\urcorner$ is true in $\alpha$.

Conversely, suppose $\ulcorner(\varphi \Rightarrow \psi)\urcorner$ is true in $\alpha$. Then $\ulcorner\underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$ is true in $\alpha$. By 5.18 and 5.12, $\ulcorner\underset{a}{\Box}\forall(\varphi \supset \psi) \ \& \ \underset{a}{\Diamond}\exists\varphi\urcorner$ entails $\ulcorner[\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)]\urcorner$. Thus $\ulcorner\underset{a}{\Diamond}\exists\varphi > \underset{a}{\Box}\forall(\varphi \supset \psi)\urcorner$ entails $\ulcorner\underset{a}{\Diamond}\exists\varphi > [\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)]\urcorner$, which is equivalent to $\ulcorner\underset{a}{\Diamond}\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > [\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)]\urcorner$, which by 5.19 is equivalent to $\ulcorner\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A) > \forall(\varphi \supset \psi)\urcorner$. Thus $\ulcorner\forall(\varphi \supset \psi)E\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A)\urcorner$ is true in $\alpha$.

As a fairly easy corollary of 5.27, interpreting the sentential quantifier in the obvious way, we obtain:

(5.28) If $x_1, \ldots, x_n$ are the variables free in $\varphi$, then $\vdash(\varphi \Rightarrow \psi) \equiv (\exists P)(\exists A)\{(x)(x \in A \equiv x = x) \ \& \ (\varphi \ \& \ P. \Rightarrow \psi) \ \& \ [PE\exists(\varphi \ \& \ x_1, \ldots, x_n \notin A)]\}$.

Consequently, the conjectured characterizations of weak subjunctive generalizations turn out to be correct.

Our language contains five kinds of conditionals. They form a sort of heirarchy as follows:

(5.29) $\vdash(\varphi \rightarrow \psi) \supset (\varphi \Rrightarrow \psi)$.

$\vdash(\varphi \Rrightarrow \psi) \supset (\varphi \Rightarrow \psi)$.

$\vdash[\underset{p}{\Diamond}(\sim\varphi \ \& \ \sim\psi) \ \& \ (\varphi \Rrightarrow \psi)] \supset (\varphi \gg \psi)$.

$\vdash[\underset{a}{\Diamond}(\sim\varphi \ \& \ \sim\psi) \ \& \ (\varphi \Rightarrow \psi)] \supset (\varphi \gg \psi)$.

$\vdash(\varphi \Rrightarrow \psi) \supset (\psi > \psi)$.

$\vdash(\varphi \gg \psi) \supset (\varphi > \psi)$.

$\vdash(\varphi > \psi) \supset (\varphi \supset \psi)$.

## NOTE

[1] If $\varphi$ is an open formula, the interpretation of $\ulcorner\underset{p}{\Box}\varphi\urcorner$ is to be $\ulcorner\underset{p}{\Box}\forall\varphi\urcorner$.

CHAPTER VII

# CAUSES

## 1. Introduction

The concept of a cause pervades the entire framework of concepts in terms of which we think of the world. As Ducasse points out, this "is made evident by the very large number of verbs of causation in the language; e.g., to push, to bend, to corrode, to cut, to make, to ignite, to transport, to convince, to compel, to remind, to irritate, to influence, to create, to motivate, to stimulate, to incite, to mislead, to induce, to offend, to effect, to prevent, to facilitate, to produce, etc." (Ducasse, 1966, p. 141). However, despite intensive philosophical labors, the concept of a cause has stubbornly resisted analysis. The major difficulty has been that talk of causes seems to involve us in a mysterious metaphysical kind of contingent necessity, and hence an account of causation would seem to call for the kind of metaphysical theory which is in disfavor in contemporary philosophy. However, philosophers have been equally suspicious about the metaphysical underpinnings of subjunctive conditionals, and as we have seen, it is possible to give a straightforward non-metaphysical account of them. Furthermore, it seems initially plausible to suspect that causes are intimately bound up with laws and subjunctive conditionals, and as such contain a subjunctive element. Perhaps there is hope that we can analyze the notion of a cause with the help of our newfound understanding of subjunctive conditionals. Such an attempt will be made in this chapter.

## 2. The ontology of causes

In recent years, Donald Davidson, Jaegwon Kim, Zeno Vendler, and others have urged repeatedly that we should get clear on the ontology of causes before we undertake an analysis of the notion of a cause. The traditional supposition has been that causation is a relation between

events, and so the ontology of causes has been supposed to be the ontology of events. Kim's (1971) review of Mackie (1965) makes it clear how much trouble one can unwittingly get into by ignoring questions about the nature and individuation of events. For example, in analyzing the concept of a cause philosophers have repeatedly found themselves talking about conjunctions and disjunctions of events. But the relations of conjunction and disjunction are relations between sentences or propositions. If these relations are to be used in some derivative sense in connection with events, this derivative sense must be explained, and such an explanation will very likely presuppose facts about the individuation of events. Thus Kim and Davidson have been led to propose analyses of the concept of an event.

However, I think there is a fundamental confusion here. Why do we believe that causation is a relation between events? Most philosophers take this as just obvious, but I do not think that it is at all obvious. Let us consider this question objectively.

To begin with, most philosophers are quick to point out that they are extending the use of 'event' somewhat when they assert that causation is a relation between events. Our ordinary understanding of events (it is asserted) is that they involve changes, but a cause can sometimes be the absence of a change. For example: 'The bell's not ringing caused us to remain seated'. But, it is supposed, with this minor extension of the concept we can make it true that causes are events. However, the extension involved is not really quite so minor as philosophers have been apt to suppose. The difficulty is that the dictum 'Causes are events' makes one prone to overlook the great variety of antecedents that are possible in causal statements. We can have conjunctive causes: 'That switch $A$ was closed and switch $B$ was opened caused the light to go on'; we can have negative causes: 'The switch's not being closed caused the light to remain off'; and (hence) we can have disjunctive causes: 'That either switch $A$ was open or switch $B$ was open caused the light to remain on'. We can also have existential causes: 'That someone entered the room caused the buzzer to sound'. In general, we can use logical operators to construct causal antecedents of arbitrary complexity. We must not understand 'event' so narrowly that we rule out these logically complex causes.

Given the required extension of the concept of an event, what sorts

of things are events? Davidson (1971, p. 217) gives a strangely mixed list in answer to this question: Sally's third birthday party, the eruption of Vesuvius in 1906 A.D., my eating breakfast this morning, the first performance of *Lulu* in Chicago. What is odd about this list is that the terms it contains are not all of the same grammatical category. 'My eating breakfast this morning' is a nominalized gerund. This is typical of the first term in causal statements, which tend to be expressions like 'The bell's not ringing', or 'The eight ball's striking the five ball'. But 'Sally's third birthday party' and 'the first performance of *Lulu* in Chicago' are not of this form. Furthermore, taking all of these terms at face value as really denoting things, they must denote different sorts of things. For example, there is no way to replace 'Sally's third birthday party' with a nominalized gerund which denotes the same thing. We cannot say, e.g., that Sally's third birthday party was the same thing as Sally's having her third birthday party. This identity sentence is either logically false or grammatically illformed. This can be made clearer by noting that the birthday party may have attributes inconsistent with attributes possessed by Sally's having the party. Perhaps the party was horrible – no one enjoyed it, everyone felt self-conscious, but they all came and pretended to have fun because Sally had just recovered from a serious illness. On the other hand, Sally's having the party was wonderful because this signified her recovery from the illness that everyone expected to be fatal. If the party was horrible, but Sally's having the party was wonderful, then the party cannot be the same thing as Sally's having the party. Or to take another attribute, the party may have been long and drawn out, but this is not something that can even be meaningfully said of Sally's having the party. Switching examples, the first performance of *Lulu* in Chicago is not the same sort of thing as the play's being performed for the first time in Chicago. Performances of plays cannot be identified with the plays' being performed. For example, the performance might have been brilliant, but the play's being performed was not brilliant – it was stupid, it caused a race riot.

It seems then that there are two logically different sorts of things that philosophers have called 'events'. On the one hand there are the referents of these nominalized gerunds (supposing them to really have referents), and on the other hand there are things like birthday parties

and performances of plays. This disparity becomes particularly important when we realize two things. First, only items of the second grammatical category (birthday parties, performances of plays, baseball games, etc.) would ordinarily be called events. Second, only items of the first grammatical category enter directly into causal statements. Our most fundamental causal statements have forms like 'My eating breakfast this morning caused me to fall asleep while lecturing'. This second point requires elaboration. My claim is that the most fundamental kind of causal statement has the form

(2.1) ger($P$) caused inf($Q$).

where ⌜ger($P$)⌝ is the appropriate nominalized gerund constructed from $P$, and ⌜inf($Q$)⌝ is the appropriate infinitive construction on $Q$. I will now explain what I mean here by 'most fundamental'.

We sometimes speak as if physical objects were causes. We say things like 'The tree caused the accident (by being too close to the road)'. However, it is rather obvious that an object can cause something only by being a certain way or having a certain property. In other words,

(2.2) $x$ caused it to be the case that $Q$.

is analyzable as:

(2.3) $(\exists F)[x$'s being $F$ caused it to be the case that $Q]$.

This is a rather obvious point which, I think, will be granted by everyone. Furthermore, causal statements of the form of 2.1 are more fundamental than causal statements of the form 2.2 in the sense that the latter can be analyzed in terms of the former, but not conversely. The former cannot be analyzed in terms of statements of the form of 2.2 because the former contain more information. They tell us *how* $x$ caused it to be the case that $Q$, and that information is not contained in statements of the form of 2.2.

We also speak of events as being causes; 'Sally's third birthday party caused her sister to be jealous', 'The baseball game caused a traffic jam'. But it is my contention that these causal statements are parallel to those of form 2.2 and are to be analyzed analogously. That is, an event can only cause something by virtue of having a certain property,

and then it is true that the event's having that property was the cause. For example, consider:

(2.4) The duel caused women to faint.

This might be true simply because the duel's occurring caused women to faint. But on the other hand, duels might be everyday events which women take in their stride, and it was not the occurrence of the duel that caused women to faint, but rather it's being so bloody. This would equally make 2.4 true. This indicates that 2.4 is to be analyzed as:

(2.5) $(\exists F)$[the duel's being $F$ caused women to faint].

In general, where '$E$' is a term denoting an event,

(2.6) $E$ caused it to be the case that $Q$.

is analyzable as:

(2.7) $(\exists F)$[$E$'s being $F$ caused it to be the case that $Q$].

Thus causal statements of the form of 2.6 are analyzable in terms of those of the form 2.1. And once again, causal statements of form 2.1 do not seem to be analyzable in terms of those of the form 2.6, because the former (if they mention an event) tell us *how* the event caused what it caused, but the latter do not. In this sense, causal statements of form 2.1 are more basic or fundamental than those of form 2.6.

What this all indicates is that although we do talk about events being causes, events are not causes in the most fundamental sense of 'cause'. The sense in which events can be causes is precisely the same as the sense in which physical objects can be causes. Events play no more basic or interesting a role in causal statements than do physical objects. If we are to understand causal statements, we must look directly at those of the form ⌜ger($P$) caused inf($Q$)⌝, and these do not appear to say anything directly about events.

In order to have a simple way of referring to them, let us call statements of the form ⌜ger($P$) caused inf($Q$)⌝ *basic causal statements.* We can express basic causal statements in a kind of canonical form as ⌜Its being the case that $P$ caused it to be the case that $Q$⌝. I have argued that basic causal statements are not to be construed as expressing a relation between events. One is inclined to ask, then, what sorts

of entities are related by basic causal statements. But this question is premature. It is not obvious that basic causal statements are relational statements at all. Their surface grammar is not that of a relational statement. The difficulty is that ⌜inf($Q$)⌝ is not a denoting expression. Expressions like 'women to faint' or 'us to remain seated' do not even purport to denote anything. In light of this surface grammar, it is a little bit mysterious that philosophers have been so quick to take basic causal statements as relational statements.

But perhaps this is too fast. We do talk about causing events: 'Throwing the switch caused the explosion'; 'Seeding the clouds caused the rainstorm'. These causal statements do not have form of basic causal statements. However, they are readily seen to be equivalent to basic causal statements, viz.: 'Throwing the switch caused the explosion to occur'; 'Seeding the clouds caused the rainstorm to occur'. In general, talk about causing events is equivalent to talk about causing the events to occur. Conversely, can basic causal statements always be reformulated as statements about the causes of events? If so, then it seems we could, after all, regard the second term of a causal relation as being an event. But there are difficulties for this proposal. First, there are infinitive constructions for which there do not seem to correspond events. E.g., there does not seem to be any 'event term' $E$ such that 'The oven's being turned on caused the pie to bake' is equivalent to 'The oven's being turned on caused event $E$ to occur'. One is apt to propose that 'the baking of the pie' is such an event term, but this is a nominalized gerund, and as we have seen, nominalized gerunds do not designate events. But perhaps all that this shows is that there is no term in our language which designates the event and not that there is no such event. I do not see how to resolve that question. However, there is a more fundamental difficulty. Even when there are natural event terms corresponding to infinitive constructions, we do not get the simple equivalence we might expect. Consider, for example:

(2.8) Seeding the clouds caused it to rain today.

Corresponding to 'it to rain today' is an event – today's rain. Thus it might seem that 2.8 is equivalent to:

(2.9) Seeding the clouds caused today's rain.

However, 2.8 and 2.9 are not equivalent. 2.8 entails 2.9, but not conversely. For example, suppose the clouds were seeded at 3:00 PM, and this caused it to rain at 4:00 PM. However, meteorological conditions were such that if it hadn't rained at 4:00 because of the cloud seeding, it would have rained naturally at 5:00 PM. Then it is not true that seeding the clouds caused it to rain today – it would have rained anyway. But seeding the clouds did cause today's rain – that is, it caused that very rain we actually had today, because if the clouds had not been seeded we would have had a different rain (one at 5:00 rather than at 4:00, involving different clouds, different raindrops, etc.). Thus 2.8 and 2.9 are not equivalent. It does not seem that basic causal statements can be construed as expressing a relation the second term of which is an event.

As I remarked above, the surface grammar of basic causal statements is not that of relational statements at all, because the consequents are not substantival expressions. Still, surface grammar does not tell the whole story, and it may be possible ultimately to analyze basic causal statements in terms of some relation between entities of some sort. In fact, there seems to be one rather simple way of doing this. It seems we can always construe basic causal statements as expressing a relation between *propositions*. That is, we can think of ⌜Its being the case that $P$ caused it to be the case that $Q$⌝ as expressing a relation between the proposition-that-$P$ and the proposition-that-$Q$. This suggests that we symbolize basic causal statements using a sentential connective, ⌜$PCQ$⌝, and then interpret this sentential connective as expressing a relation between propositions.

It must be pointed out that the symbolization of basic causal statements using a sentential connective does not commit us to interpreting them in terms of a relation between propositions. The use of the sentential connective is nothing more than a shorthand device for expressing basic causal statements. On the other hand, the assumption that this sentential connective can be interpreted as a relation between propositions is tantamount to assuming the validity of the following two principles:

(2.10) If the proposition-that-$P$ = the proposition-that-$R$, then $PCQ$ iff $RCQ$.

(2.11) If the proposition-that-$Q$ = the proposition-that-$R$, then $PCQ$ iff $PCR$.

Unfortunately, the individuation of propositions is almost as much of a problem as is the individuation of events. Philosophers often pretend that logical equivalence is a necessary and sufficient condition for propositional identity, but it is pretty obvious that this is really too weak a criterion. However, it is at least clear that logical equivalence is a necessary condition for event identity, and hence 2.10 and 2.11 are implied by the following two principles:

(2.12) If $P \leftrightarrow R$, then $PCQ$ iff $RCQ$.

(2.13) If $Q \leftrightarrow R$, then $PCQ$ iff $PCR$.

If we can agree that 2.12 and 2.13 are true, then we can safely interpret basic causal statements as expressing a relation between propositions.

That 2.12 and 2.13 are true seems almost obvious to me. However, Davidson (1967) has given an argument which purports to establish their invalidity. By adapting an argument of Frege, Davidson shows that principles 2.12 and 2.13 together with the following extensionality principle:

(2.14) If $t_1$ and $t_2$ are individual terms and $t_1 = t_2$, then $(Ft_1)CQ$ iff $(Ft_2)CQ$, and $PC(Ft_1)$ iff $PC(Ft_2)$.

lead to the absurd result that the connective 'C' is truth-functional. In particular, they lead to the result that if any basic causal statements are true, then whenever $P$ and $Q$ are true, so is $\ulcorner PCQ \urcorner$. The argument is as follows. Suppose we have some true causal statement $\ulcorner RCS \urcorner$. Then $R$ and $S$ are true. But $R$ and $S$ are logically equivalent, respectively, to the statements $\ulcorner \{x; x \in \omega \ \& \ R\} = \omega \urcorner$ and $\ulcorner \{x; x \in \omega \ \& \ S\} = \omega \urcorner$ (where $\omega$ is the set of all natural numbers). Thus by 2.12 and 2.13, $(\{x; x \in \omega \ \& \ R\} = \omega)C(\{x; x \in \omega \ \& \ S\} = \omega)$. If $P$ and $Q$ are true, then $\{x; x \in \omega \ \& \ P\} = \{x; x \in \omega \ \& \ R\}$, and $\{x; x \in \omega \ \& \ Q\} = \{x; x \in \omega \ \& \ S\}$. Hence by 2.14, $(\{x; x \in \omega \ \& \ P\} = \omega)C(\{x; x \in \omega \ \& \ Q\} = \omega)$. But the antecedent and consequent of this basic causal statement are equivalent to $P$ and $Q$ respectively, so by 2.12 and 2.13, $\ulcorner PCQ \urcorner$ is true. Davidson takes this absurd conclusion as establishing the incorrectness of 2.12 and

2.13, because he thinks that 2.14 is obviously correct. But strictly speaking, all the argument shows is that we must reject either 2.12 and 2.13, or principle 2.14, and in the abstract I do not think that either choice is obviously the correct one.

Davidson thinks that 2.14 is obviously correct because he accepts the dogma that basic causal statements express a relation between events, and he thinks that events are extensional in the sense of 2.14. Personally, I do not think it is all that obvious that events are extensional in the appropriate sense, but whether they are or not is irrelevant to the validity or invalidity of 2.14 because, as we have seen, basic causal statements do not express relations between events. The assumption that they do has been responsible for a great deal of confusion regarding the logical properties of basic causal statements. In particular, this assumption has led a number of authors to endorse 2.14 and on that basis to reject 2.12 and 2.13.

The decision whether to reject 2.12 and 2.13 or to reject 2.14 must rest upon an appeal to actual examples. At first glance, the appeal to actual examples appears to support 2.14, but as we will see on closer examination, this appearance is illusory. Suppose we have a group of men in a room, just one of whom is red-headed, and the redhead is also the most powerful man in the room. Suppose the following is true:

(2.15) That the most powerful man in the room gave the order caused it to be obeyed.

Does the following follow from 2.15?

(2.16) That the only redhead in the room gave the order caused it to be obeyed.

It seems that it does. But we must be careful. 2.15 and 2.16 involve definite descriptions, and definite descriptions are subject to scope ambiguities. In general, $\ulcorner(F\imath xGx)\mathsf{C}Q\urcorner$ is ambiguous between the narrow scope reading $\ulcorner[Fx](\imath xGx)\mathsf{C}Q\urcorner$ and the wide scope reading $\ulcorner[Fx\mathsf{C}Q](\imath xGx)\urcorner$. In general, the scope notation is defined by stipulating that:

(2.17) $\ulcorner[Fx](\imath xGx)\urcorner$ is equivalent to $\ulcorner(\exists!x)Gx \;\&\; (\exists x)(Fx \;\&\; Gx)\urcorner$.

Applying this to our causal statements, the narrow scope reading

$\ulcorner[Fx](\imath xGx)\mathsf{C}Q\urcorner$ is defined as:

(2.18) $[(\exists!)Gx \;\&\; (\exists x)(Fx \;\&\; Gx)]\mathsf{C}Q$

(idiomatically: ⌜that there is a unique $G$ and it is $F$ caused it to be the case that $Q$⌝). The wide scope reading $\ulcorner[Fx\mathsf{C}Q](\imath xGx)\urcorner$ is defined as:

(2.19) $(\exists!x)Gx \;\&\; (\exists x)[Gx \;\&\; Fx\mathsf{C}Q]$

(idiomatically: ⌜there is a unique $G$, and *its* being $F$ caused it to be the case that $Q$⌝).

Corresponding to the scope ambiguity, where $t_1$ and $t_2$ are definite descriptions, 2.14 is ambiguous between two principles:

(2.14*) If $t_1 = t_2$, then $[Fx](t_1)\mathsf{C}Q$ iff $[Fx](t_2)\mathsf{C}Q$, and $P\mathsf{C}[Fx](t_1)$ iff $P\mathsf{C}[Fx](t_2)$.

(2.14**) If $t_1 = t_2$, then $[Fx\mathsf{C}Q](t_1)$ iff $[Fx\mathsf{C}Q](t_2)$, and $[P\mathsf{C}Fx](t_1)$ iff $[P\,\mathsf{C}Fx](t_2)$.

As examination of Davidson's argument shows that in order to make the substitution steps which enable him to obtain his disastrous conclusions he must have the narrow-scope version of 2.14, viz., 2.14*. This is because the term $\ulcorner\{x; \varphi x\}\urcorner$ has the same meaning as the definite description $\ulcorner\imath x(y)(y \in x \equiv \varphi y)\urcorner$. But it is my contention that only the wide-scope version, 2.14**, is true. This can be seen by considering what happens when we resolve the scope ambiguity in 2.15 and 2.16. I believe that our intuition that 2.16 follows from 2.15 concerns only the wide scope reading of these sentences, and hence supports only 2.14**. The narrow- and wide-scope readings respectively of 2.15 and 2.16 are:

(2.15*) That there was a unique most powerful man in the room and the person who gave the order was a most powerful man in the room, caused the order to be obeyed.

(2.16*) That there was a unique redhead in the room and the person who gave the order was a redhead in the room, caused the order to be obeyed.

(2.15**) There was a unique most powerful man in the room, and there was someone such that he was the most powerful man in the room and his giving the order caused it to be obeyed.

(2.16**) There was a unique redhead in the room, and there was someone such that he was a redhead in the room and his giving the order caused it to be obeyed.

Given this disambiguation,[1] it seems to me that 2.16* is unequivocally false and 2.16** is unequivocally true. Thus the example supports only 2.14** and actually constitutes a counter-example to 2.14*.

This diagnosis indicates that Davidson's argument does not, after all, establish the invalidity to 2.12 and 2.13, and hence we are free to interpret 'C' as expressing a relation between propositions. However, such an interpretation does not preclude our also interpreting 'C' in terms of a relation between some other entities having less stringent criteria for individuation. I have argued that there is no reason to think causation to be a relation between events, but given the distinctions I have made this is probably not what most recent philosophers have wanted to maintain anyway. They have really wanted to say that causation is a relation between the entities that are denoted by nominalized gerunds, and they mistakenly thought that those entities are what are commonly called 'events'. An initial obstacle to this position would be that the consequents of basic causal statements are not nominalized gerunds, but notice that the basic causal statement ⌜Its being the case that $P$ caused it to be the case that $Q$⌝ is convertible into the equivalent statement ⌜Its being the case that $Q$ was caused by its being the case that $P$⌝, and the latter does seem to express a relation between the referents of two nominalized gerunds. There are going to be philosophers who question whether nominalized gerunds denote anything at all, but this is not a difficulty I am inclined to take seriously. If nothing else, we could always artificially construct suitable entities somewhat on the order of Kim's ordered triples (Kim (1970)). I think we might reasonably say that nominalized gerunds denote 'states of affairs', taking the latter as an explicitly technical term introduced for the sole purpose of having a name for these entities. Then it seems quite reasonable to say that basic causal statements express a relation between states of affairs. However, even if we grant this, it is not at all obvious how states of affairs are individuated. In particular, states of affairs are much more 'logically strict' entities than are events, and it is not obvious that states of affairs have coarser criteria for individuation than propositions do.

I think this whole attempt to resolve the ontological problems about causation before attempting to give an analysis of causal statements is essentially backwards. Given a kind $E$ of entity, the question whether basic causal statements can be regarded as expressing a relation between entities of kind $E$ is equivalent to the question whether, whenever $P$ and $R$ 'correspond to' the same entity of kind $E$ and $Q$ and $S$ 'correspond to', the same entity of kind $E$, $\ulcorner PCQ \urcorner$ is true iff $\ulcorner RCS \urcorner$ is true. The criteria for individuating entities of kind $E$ can be regarded as telling us when two propositions 'correspond to' the same entity of kind $E$. Consequently, the question of what kinds of entities can be regarded as the relata of basic causal statements is just the question of what relations between propositions make them interchangeable in the context of basic causal statements. In other words, what relations $\mathfrak{R}$ make the following true:

(2.20) If $P\mathfrak{R}Q$ then $PCR$ iff $QCR$, and $RCP$ iff $RCQ$.

Putting the question in this way indicates that to attempt to answer the ontological question before analyzing basic causal statements is to put the cart before the horse. There are really two distinguishable problems regarding the concept of a cause. First, there is the ontological question, which we have been considering. This is the question what sorts of entities are related by causal relations. Distinct from the ontological question is the second, and more traditional, problem regarding the concept of a cause. This is to give an analysis of statements like 'John's striking the match caused there to be an explosion' in terms of other notions we already understand. There is no reason why we should have to solve the ontological problem before we can solve this latter problem. And as we have just seen, the ontological question is equivalent to a question about the logical properties of basic causal statements (i.e., the question what relations $\mathfrak{R}$ satisfy 2.20), so any adequate answer to the ontological question may actually presuppose a prior analysis of basic causal statements.

In fact, I have been unable to find any interesting relations $\mathfrak{R}$ (other than logical equivalence) which satisfy condition 2.20. This suggests that basic causal statements cannot be analyzed as expressing a relation between entities having coarser criteria for individuation than logical equivalence, and hence that basic causal statements can only be

regarded as expressing a relation between propositions or something with equally stringent criteria for individuation. I am not sure that this is the case, but for lack of any positive results to the contrary I do not propose to return to the ontological question. The remainder of this chapter will be concerned with an analysis of basic causal statements in terms of a relation between propositions.

## 3. Some causal relations

I have stated our problem as being that of analyzing the causal relation 'its being the case that . . . caused it to be the case that . . .'. However, this is not the only causal relation, and there is reason to believe that it is not even the most fundamental or most important causal relation. To begin with, this relation only holds between *true* propositions, but we often have occasion to say that something which did not happen would have caused something else. This subjunctive notion of cause is interdefinable with our indicative notion of cause:

(3.1) Its being the case that $P$ would cause it to be the case that $Q$, iff, $[P > (P \mathsf{C} Q)] \,\&\, \Diamond P$.[2]

(3.2) $P \mathsf{C} Q$ iff, its being the case that $P$ would cause it to be the case that $Q$, and it is true that $P$.[3]

I will frequently abbreviate the long construction ⌜Its being the case that $P$ would cause it to be the case that $Q$⌝ as ⌜$P$ would cause $Q$⌝. Similarly, I will abbreviate ⌜Its being the case that $P$ caused it to be the case that $Q$⌝ as ⌜$P$ caused $Q$⌝. These shortened forms are convenient, but they are grammatically illformed and must be understood as elliptical for the more complicated constructions.

There is another causal relation different from either of these two relations. We might call this the relation of causal sufficiency. Frequently, its being the case that $P$ may fail to cause it to be the case that $Q$ only because there is some third proposition $R$ such that its being the case that $R$ has already caused it to be the case that $Q$. For example, suppose we shoot a man twice. Shooting him the first time causes his death. Then shooting him the second time may only fail to cause his death because something else has already caused it. When

this happens, we say that its being the case that $P$ would be causally sufficient for it to be the case that $Q$ even though it would not or did not cause it to be the case that $Q$. This notion of causal sufficiency seems, in some sense, to be the most fundamental causal concept. The relation between one proposition and a second proposition which it would cause is that of causal sufficiency. Whether the first proposition also causes the second depends not just on the relation between the two propositions, but also on what other true propositions there are that would be causally sufficient for the second proposition.

The notion of causal sufficiency is also interdefinable with the notion of a cause. To say that $P$ would be causally sufficient for $Q$ (symbolized $\ulcorner PCSQ \urcorner$) is to say, roughly that $P$ would cause $Q$ if $Q$ weren't already true for some other reason. This suggests:

(3.3) $PCSQ$ iff $[\sim Q > (P > PCQ)]$.

However, this does not adequately handle the case in which $P$ actually does cause $Q$. In such a case, if $Q$ were false, then $P$ might no longer be sufficient to cause $Q$. For example, suppose we have a light operated by two switches in series. If both switches are closed, the light is on. Switch $A$ was closed initially, and then switch $B$ was closed. Switch $B$'s being closed caused the light to be on. But if the light weren't now on, then one of the two switches would be open, but there is nothing that determines which switch would be open. In particular, switch $A$ might be open. But in that case, switch $B$'s being closed would not cause the light to be on. So it is not true that if the light weren't on, then switch $B$'s being closed would cause the light to be on. The difficulty here is that if $P$ actually causes $Q$, it characteristically does so in conjunction with certain other collateral truths, so if $Q$ were false, we can only conclude that either $P$ would be false or one of those collateral truths would be false. If one of the collateral truths were false, then $P$ would no longer cause $Q$. We can take care of this just as we did in the analysis of the necessitation conditional by saying not just that $Q$ is false, but also specifying that it is the failure of $P$ which is involved in $Q$'s being false rather than the failure of one of the collateral truths. In other words, our analysis should be:

(3.4) $PCSQ$ iff $[(\sim P \ \& \ \sim Q) > (P > PCQ)] \ \& \ \Diamond(\sim P \ \& \ \sim Q) \ \& \ \Diamond P$.

We include the final conjuncts to exclude the trivial cases.

It is equally simple to define 'cause' in terms of causal sufficiency. The most common case in which we want to say that $P$ was causally sufficient for $Q$ without causing $Q$ is where there is more than one true proposition causally sufficient (in different ways) for $Q$. The classical example is that of two members of a firing squad who fire their weapons simultaneously with the result that their bullets hit the victim at the same instant and are each causally sufficient to cause death at the same instant. We will not happily say that either man's firing his weapon individually caused the death of the victim (although, of course, the death was caused by the fact that at least one of them fired his weapon). This seems to be because, had either man not fired his weapon, the victim would have died anyway. This suggests that we can analyze 'cause' as:

(3.5) $PCQ$ iff $[P \ \& \ PCSQ \ \& \ (\sim P > \sim Q)]$.

However, counterexamples have been proposed to the condition $\ulcorner(\sim P > \sim Q)\urcorner$. Scriven (1964) gives the example of a man whose state of excitation at 4:00 PM is such that he would suffer a stroke at 4:55 PM which would cause his death at 5:00 PM were it not that a heart attack intervenes at 4:50 PM which causes his death at 5:00 PM. The heart attack prevents the stroke, and hence is the cause of death. But it is not true that if he had not suffered the heart attack he would not have died at 5:00 PM. On the contrary, he would have died of the stroke. Thus it seems that the condition $\ulcorner(\sim P > \sim Q)\urcorner$ is too strong.

I do not think that this is a genuine counterexample. It involves the same confusion as was involved earlier in overlooking the distinction between causing it to rain today and causing today's rain. To make it easier to see this, let us first consider a modified example in which the stroke would have caused death at 5:10 PM rather than at 5:00 PM. Then it is true both that the heart attack caused the man's death and that the heart attack caused the man to die at 5:00 PM. But for both of these, the condition $\ulcorner(\sim P > \sim Q)\urcorner$ is satisfied. First, if he had not had the heart attack, the man would still have died, but it would have been a different death. It would have been a death occurring at a different time and involving quite different processes. Second, if the man had not had the heart attack he would not have died at 5:00 PM. Contrast these two true causal statements with the false causal statement that

the heart attack caused the man to die today, i.e., caused it to be the cause that there was a time today when he died.[4] This statement is false *precisely because* he would have died today anyway.

We have made the example easier to deal with by supposing that the stroke would have caused death at a different time than the heart attack. Let us return to the original example in which the stroke would also cause death at 5:00 PM. We must still distinguish between causing the man's death (that very death that actually occurred) and causing the man to die at 5:00. It is still true that the heart attack caused the man's death, and that the death would not have occurred had he not had the heart attack. The man would still have died, but it would have been a different death – one involving quite different processes. On the other hand, it is no longer true that the heart attack caused the man to die at 5:00 – and this is because he would have died at 5:00 anyway. Thus, rather than constituting a counterexample to 3.5, I think that 3.5 explains the fine structure of the judgments that we make regarding this example. Of course, my resolution of this putative counterexample presupposes certain judgments about the individuation of events, but I think that those judgments are correct. Consequently, I believe that 3.5 constitutes a correct analysis of 'causes' in terms of causal sufficiency.

Because of the interdefinability of 'cause', 'would cause', and 'causally sufficient', if we can provide an analysis of one of these notions, analyses of the others will follow. My strategy below will be to give an analysis of causal sufficiency.

## 4. CAUSAL SUFFICIENCY

We turn now to the analysis of causal sufficiency. We will build up to our final analysis by considering sequentially some plausible suggestions that have been made in the literature.

### 4.1. *Nomic Subsumption*

The simplest approach to the analysis of causal sufficiency is the nomic subsumption model. This arises out of the traditional regularity theory

or the constant conjunction theory adumbrated by Hume. On the nomic subsumption model, $P$ is causally sufficient for $Q$ just in case $Q$ 'can be obtained from' $P$ together with some physical laws. Kim (1973) has shown how difficult it is to make clear this notion of 'can be obtained from'. But this is a problem we have already discussed in talking about how laws are derivable from one another. We can formulate the nomic subsumption model quite simply as follows:

(4.1) Nomic Subsumption Model: $P$ is causally sufficient for $Q$ iff $P \Rrightarrow Q$.

There is a standard objection to the nomic subsumption model. This is that it only deals with what John Stuart Mill called 'total causes'. For example, consider a house fire that results from faulty insulation. We would say that the faulty insulation is the cause of the fire, but there is no physical law to the effect that faulty insulation is always followed by a fire. In formulating the physical law that is operative here, we must mention the presence of oxygen, the temperature of the air, the current in the wires, etc. According to the nomic subsumption model, all of these together would be causally sufficient for the fire, but the faulty insulation would not be. But this is incorrect if taken as a remark about the concept of 'cause' that we have set out to analyze. The notion of the 'total cause' is probably a useful notion (although, as we will see, the nomic subsumption model does not capture it correctly), but it is not our ordinary concept of 'cause'. The ordinary concept of a cause has much in common with subjunctive conditionals. What makes subjunctive conditionals useful, and also difficult to analyze, is that under certain circumstances we are allowed to delete true conjuncts from the antecedent. The whole problem of analyzing subjunctive conditionals was to say when that can be done. Similarly, in stating causes, we are not required to state the total cause. Under certain circumstances we can delete mention of most of the causal factors, reporting simply that the remaining causal factors caused the effect. An important part of the problem of analyzing causal sufficiency is to say when such causal factors need not be mentioned in stating the cause.

I have been urging that the nomic subsumption model is at best an analysis of 'total cause', and not an analysis of causal sufficiency. But

there is a more fundamental difficulty with the nomic subsumption model which shows that it does not even capture the notion of 'total cause'. This is the difficulty, which we will encounter again, of 'epiphenomena'. When two propositions share a common cause, without either being causally sufficient for the other, we say that they are *epiphenomena*. We might well have a law statement $\ulcorner P \Rrightarrow Q \urcorner$ which is true not because $P$ is causally sufficient for $Q$, but rather because there are physical laws which ensure that anything which would be causally sufficient for $P$ would also be causally sufficient for $Q$. For example, $P$ and $Q$ might report the occurrence of astronomical events on a star, and we might have cosmological laws which tell us that there is only one way that $P$ could be true, and that would also result in $Q$'s being true. Under these circumstances, $P$ and $Q$ are epiphenomena instead of $P$ being the total cause of $Q$.

Apparently the notion of a total cause cannot be defined so simply as proposed by the nomic subsumption model. Just because $\ulcorner P \Rrightarrow Q \urcorner$ is true, it does not follow that $P$ is the total cause of $Q$. What is required in addition is that $P$ *cause* $Q$. This suggests that we must define 'total cause' as follows:

(4.2) $P$ is a total cause of $Q$ iff $P \mathbf{C} Q$ and $P \Rrightarrow Q$.

The difference between a cause and a total cause is that the latter is a cause which is connected to its effect directly by a physical law.

Of course, if we define 'total cause' as in 4.2, it will not be of much help in analyzing causal sufficiency. In particular, if we try to define causal sufficiency directly in terms of 'total cause', our definitions will be circular. It might be possible to avoid this circle by defining 'total cause' in some other way, but, as we will see, it is possible to break the circle anyway at the other end by providing an analysis of causal sufficiency which does not use the notion of a total cause.

## 4.2. *Contingently Sufficient Conditions*

Scriven (1964) and Mackie (1965) have proposed basically similar analyses which explicitly recognize that in stating the cause of something, we do not have to state the total cause. Scriven formulates the analysis succinctly by saying that causes are "contingently sufficient

conditions". More precisely:

(4.3) $P$ is causally sufficient. for $Q$ iff there is a set of true propositions $R_1, \ldots, R_n$ such that $P$ together with $R_1, \ldots, R_n$ is sufficient for $Q$ (i.e., $(P \ \& \ R_1 \ \& \ \ldots \ \& \ R_n) \Rightarrow Q$), but $R_1, \ldots, R_n$ alone are not sufficient for $Q$ (i.e., $(R_1 \ \& \ \ldots \ \& \ R_n) \not\Rightarrow Q$).

By taking the conjunction of the $R$'s, we can simplify this proposal to:

(4.4) $P$ is causally sufficient for $Q$ iff there is a true proposition $R$ such that $(P \ \& \ R) \Rightarrow Q$, but $R \not\Rightarrow Q$.

This analysis is clearly inadequate as it stands. The simplest way of demonstrating the inadequacy is to note that it entails that if $Q$ is true, then any proposition at all is causally sufficient for $Q$. To get this result, we simply let $R$ by $\ulcorner(P \supset Q)\urcorner$. For the same reason, a false proposition would be causally sufficient for anything at all. This difficulty seems to be related to the analogous difficulty that arose in connection with subjunctive conditionals. There the difficulty was to explain under what circumstances we can detach a true proposition $R$ from a conditional $\ulcorner[(P \ \& \ R) \Rightarrow Q]\urcorner$ to obtain a true simple subjunctive $\ulcorner(P > Q)\urcorner$. In that case, the condition required for detachment was $\ulcorner REP\urcorner$. This condition also seems to be required in the case of causal sufficiency. When a condition $R$ is already true which, when conjoined with $P$ would be causally sufficient for $Q$, in order for us to delete mention of $R$ and say simply that $P$ itself would be causally sufficient for $Q$ it is at least necessary that $R$ would still be true if $P$ were true. If this condition is not satisfied, then $P$'s being true would be no guarantee of the truth of $Q$, and hence $P$ is not causally sufficient for $Q$. For example, it is true that natural gas has been flowing from the orifice of the pilot light on my furnace for the past hour. That, taken in conjunction with the flame on the pilot light's having blown out an hour ago and my now lighting a match in the vicinity of the water heater, would be causally sufficient for an explosion. But we cannot conclude, in accordance with 4.4, that the flame on my water heater having gone out an hour ago and my now lighting a match in its vicinity, would be causally sufficient for an explosion. This is because it is not true that gas would have been flowing from the pilot light for the

last hour had it blown out. There is a safety mechanism designed to prevent that. This indicates that we must at least add one further condition to 4.4:

(4.5) $P$ is causally sufficient for $Q$ iff there is a true proposition $R$ such that $REP$ and $[(P \ \& \ R) \Rightarrow Q]$ and $R \not\Rightarrow Q$.

But now something remarkable has happened. The analysans of 4.5 is equivalent to the condition that the simple subjunctive $\ulcorner(P > Q)\urcorner$ be true. This leads naturally into the attempt to analyze causal sufficiency in terms of subjunctive conditionals.

### 4.3. *Causal Sufficiency and Subjunctive Conditionals*

That $\ulcorner(P > Q)\urcorner$ be true is clearly a necessary condition for $P$ to be causally sufficient for $Q$, but I think it is equally clear that it is not a sufficient condition. For example, $\ulcorner(P \ \& \ Q)\urcorner$ entails $\ulcorner(P > Q)\urcorner$, but we certainly do not want to conclude that any two true propositions are causally sufficient for one another. We must add additional requirements if we are to obtain a satisfactory analysis of causal sufficiency.

To begin with, it seems clear that causal sufficiency is a special case of necessitation:

(4.6) If $P$ is causally sufficient for $Q$, then $(P \gg Q)$.

If $P$ is causally sufficient for $Q$, then the truth of $P$ must be sufficient to 'bring it about' that $Q$ is true, in the sense that $(P \gg Q)$. Although this is still not a sufficient condition for causal sufficiency, I think we are now on the right track. From this point the analysis will proceed by accumulating additional necessary conditions until finally we have a list which is also sufficient.

The basic connection between cause and effect seems to be that of necessitation. But not all necessitations give rise to causal relations. If $P \rightarrow Q$, then $P \gg Q$, but it is not usually true in such a case that $P$ is causally sufficient for $Q$. Kim (1973) gives the following two examples:

(i) Yesterday's being Monday does not *cause* today to be Tuesday.

(ii) George's being born in 1950 and still being alive in 1971 does not *cause* him to have reached the age of 21 in 1971.

It seems reasonable to suppose that causation is a species of *contingent* necessitation, and hence the cause can never entail the effect. However, there are what appear to be counterexamples to this supposition:

(1) If you heat the metal to a sufficiently high temperature, that will cause it to melt.

(2) His bending the rod over his knee until it broke caused the rod to break.

(3) His being fatally stabbed caused him to die.

These examples seem a little odd, but I think the oddness is due to the triviality of the connection between the antecedent and the consequent rather than it being the oddness of logical impossibility. It is quite possible for statements like (1), (2), and (3) to be true, and they seem to be examples of the cause entailing the effect. However, upon closer examination this becomes less obvious. The only examples I have been able to find in which the cause seems to entail the effect are of the same general sort as (1), (2), and (3), and it seems that these statements can be paraphrased as follows:

(1*) There is a temperature such that the metal's being heated to that temperature is causally sufficient for it to melt, and if you heat it to that temperature now, that will cause it to melt.

(2*) He bent the rod to a degree which was causally sufficient for it to break, and his bending it to that degree caused it to break.

(3*) He was stabbed in a way which was causally sufficient for him to die, and his being stabbed in that way caused him to die.

Thus (1), (2), and (3) can be regarded as not having the form $\ulcorner PCQ \urcorner$, but instead a more complicated form involving quantification into causal contexts. So construed, they do not constitute counter-examples to the requirement that causation involves contingent necessitation. Consequently, I think that we can build this requirement into the analysis of causal sufficiency:

(4.7) If $P$ is causally sufficient for $Q$, then $P$ does not entail $Q$.

We have two necessary conditions for causal sufficiency, but these conditions are not also sufficient. So far we have overlooked a well known problem. This is the problem of epiphenomena. Frequently we will have two propositions $P$ and $Q$ such that $P$ contingently necessitates $Q$, but does so because $P$'s being true necessitates the truth of a third proposition $R$ which would cause $Q$ without the causal chain passing through $P$. In such a case, we say that $Q$ is an 'epiphenomenal effect' of $P$. To take a simple example, suppose we have a black box with a switch and two lights. The box is wired in such a way if the switch is thrown, light $A$ comes on and is followed two seconds later by light $B$. In an actual situation in which both lights are off, we might well know that if light $A$ were to come on this would be because the switch was thrown. Letting $P$ be 'Light $A$ comes on' and $Q$ be 'Light $B$ comes on', we see that $P$ contingently necessitates $Q$. But it is clear that $P$ is not causally sufficient for $Q$. The connection between $P$ and $Q$ is rather that what would cause $P$ to be true would also cause $Q$ to be true.

The case of epiphenomena is to be distinguished from the case in which $P$ and $Q$ have a common cause $R$, but $R$ causes $Q$ *by* causing $P$. This kind of case gives rise to talk of 'causal chains'. If in the previous example, light $B$ came on not as a direct result of throwing the switch but rather in response to a photoelectric cell sensing the light output of light $A$, then although throwing the switch would cause both lights to come on, we would not have a case of epiphenomena. This is because light $A$'s coming on would itself cause light $B$ to come on. The difference between these two cases is that in the epiphenomenal case light $A$'s coming on *without* the switch being thrown would not necessitate light $B$'s coming on, whereas in the non-epiphenomenal case light $A$'s coming on would necessitate light $B$'s coming on regardless of whether the switch is thrown. This suggests a way of characterizing the epiphenomenal case. As a first approximation we might try:

(4.8) $Q$ is an epiphenomenal effect of $P$ iff there is a proposition $R$ such that $(P \gg R)$ and $(R \gg Q)$ and $\sim[(P \ \& \ \sim R) \gg Q]$.

For example, in the epiphenomenal case of the black box, $R$ is 'The switch is thrown'.

4.8 embodies the right idea, but it is not entirely adequate as it stands. The difficulties are those of 'underdetermination' and 'overdetermination'. First underdetermination: rather than knowing of a particular $R$ which would necessitate $Q$, we may know of a set $\Gamma$ such that *one* of the members of $\Gamma$ would be true if $P$ were true, but it is not determined which member of $\Gamma$ it would be, and we may know that no matter which member of $\Gamma$ would be true, it would also necessitate $Q$. This gives us:

> $Q$ is an epiphenomenal effect of $P$ iff there is a set $\Gamma$ of propositions such that $(P \gg \Sigma\Gamma)$ and $\sim[(P \ \& \ \sim\Sigma\Gamma) \gg Q]$, and for each $R$ in $\Gamma$, $R \gg Q$.

There remains the problem of overdetermination. We might know that if $P$ were true then some member of $\Gamma$ would be true, but we might also know that if $P$ were true and no member of $\Gamma$ were true, then some member of a second set $\Gamma^*$ would be true, and each member of $\Gamma^*$ necessitates $Q$. In this case we might have $[(P \ \& \ \sim\Sigma\Gamma) \gg Q]$ true in spite of the fact that $Q$ is an epiphenomenal effect of $P$. In this case, the reason $Q$ is an epiphenomenal effect of $P$ is that $[(P \ \& \ \sim\Sigma\Gamma) \gg \Sigma\Gamma^*]$ and $\sim[(P \ \& \ \sim\Sigma\Gamma \ \& \ \sim\Sigma\Gamma^*) \gg Q]$, and for each $R$ in $\Gamma^*$, $R \gg Q$. Generalizing this, we may have a whole sequence $\Gamma_0, \ldots, \Gamma_\beta, \ldots (\beta < \alpha)$ of sets of propositions such that if $P$ were true then some member of $\Gamma_0$ would be true, but if $P$ were true and no member of $\Gamma_0$ were true then some member of $\Gamma_1$ would be true, and so on. This suggests the following definition:

(4.9) $Q$ is an epiphenomenal effect of $P$ iff there is a sequence $\Gamma_\beta (\beta < \alpha)$ of sets of propositions such that for each $\beta < \alpha$, $[(P \ \& \ \prod_{\gamma<\beta} \sim\Sigma\Gamma_\gamma) \gg \Sigma\Gamma_\beta]$, and for each $R$ in $\Gamma_\beta$, $(R \gg Q)$, and $\sim[(P \ \& \ \prod_{\gamma<\alpha} \sim\Sigma\Gamma_\gamma) \gg Q]$.

We now have three requirements for causal sufficiency. Let us use them to define a notion of 'almost causal sufficiency':

(4.10) $P$ is almost causally sufficient for $Q$ iff $(P \gg Q)$ and $P$ does not entail $Q$, and $Q$ is not an epiphenomenal effect of $P$, and $\Diamond(\sim P \ \& \ \sim Q) \ \& \ \Diamond P$.

What, if anything, must be added to almost causal sufficiency to obtain causal sufficiency?

I believe that the one outstanding problem for the relation of almost causal sufficiency is that of the direction of causation – in many cases the relation of almost causal sufficiency does not discriminate between cause and effect. For example, consider a simple electrical circuit with a switch which operates a light. Let $P$ be 'The switch is thrown' and $Q$ be 'The light comes on'. Clearly, $P$ is almost causally sufficient for $Q$. Unfortunately, this goes the other way too: $Q$ is almost causally sufficient for $P$. That is, if the light were to go on then the switch would have been thrown, there is no entailment, and $P$ is not an epiphenomenal effect of $Q$.

It is considerations like this that have led philosophers to build temporal relations into the analysis of causation. The idea is that we can distinguish cause from effect by seeing which comes first. Basically, I think that this idea is correct, but there are numerous details to be worked out. The first difficulty is that it is often unclear what it means to talk about temporal relations between a particular causal antecedent and causal consequent. Talk of temporal order would seem to presuppose a notion of the *date* of a proposition, but this is a very problematic notion. If the content of a proposition is that some object is in a simple state at a particular time, then it is clear what to count as the date of the proposition. In this case, the date is the time mentioned. But it is not so clear what to count as the date of more complex propositions. One might suppose that if a proposition asserts that something is true at a particular time, then that time should be taken as the date of the proposition. But this prima facie reasonable supposition leads to the embarrassing conclusion that logically equivalent propositions can have different dates. Consider, for example:

> At 3:00 PM, $x$ became $\varphi$.
>
> At 2:00 PM, it was one hour before $x$ was to become $\varphi$.

In general, I doubt that it makes much sense to talk about the date of a proposition even when it is in some sense a temporal proposition. For example, what is the date of the proposition that John gave Joe a black eye last week? Is the date all of last week, or is it the time when he gave him the black eye, or what? Or consider the proposition that one of my ancestors was a horse thief. What on earth could count as the date of this proposition?

If we attempt to date propositions in terms of what dates they *mention*, we will get into all of the above difficulties. But such a notion of the date of a proposition will not help us anyway. Consider, for example:

(4.11) The clouds being seeded today caused it to rain today.

The *mentioned* dates of this causal antecedent and consequent would seem to be all of today. But so construed, this cause and effect have the same date and so cannot be discriminated on that basis. However, reflection indicates that the dates that are relevant to this example are not the dates mentioned by the causal antecedent and consequent, but are rather the precise time when the clouds were seeded and the precise time when it rained. The clouds were indeed seeded before it began to rain.

The notion of the date of a proposition that is relevant here is, roughly, the date of 'what makes the proposition true'. For example, what makes it true that the clouds were seeded today is that potassium iodide crystals were released into the clouds at certain times today, and it is that collection of times which constitutes the date relevant to assessing the causal relation. In the case of a causal statement like 4.11, it seems that what is required is that 'what makes the antecedent true' caused 'what makes the consequent true', and that a temporal relation is built into this latter causal relation which thereby discriminates between cause and effect. For example, in the case of 4.11, the clouds being seeded when they were caused it to rain when it did.

To turn this into a general account, we have to get clear on 'what makes the antecedent and consequent true'. If, as in 4.11, a statement is an existential generalization, then what makes it true is whatever makes some of its instances true. Similarly, if a statement is a disjunction, what makes it true is whatever makes one of its disjuncts true. If a statement is a conjunction, what makes it true is whatever makes both conjuncts true. And so on for each kind of logical compound until we get down to propositions which are not logical compounds, i.e., propositions which are simple in the sense of Chapter IV. What makes a simple proposition true is simply it itself. And simple propositions have precise dates built into them. This suggests that we proceed as follows. First, we define a notion of 'direct causal sufficiency' which

holds between simple propositions and which, by making reference to time, allows us to discriminate between cause and effect. Then we define causal sufficiency something like:

(4.12) $P$ is causally sufficient for $Q$ iff $P$ is almost' causally sufficient for $Q$, and if $P$ were true then there would be sets $A$ and $B$ of simple propositions such that $A$ would be what makes $P$ true, and $B$ would be what makes $Q$ true, and the propositions in $A$ would be directly causally sufficient for the propositions in $B$.

As it is stated, 4.12 has a great many deficiencies. Let us take them one at a time. First, what does it mean to say that $A$ would be what makes $P$ true? As a first approximation, it seems reasonable to take this as meaning that $A$ is a minimal set of true simple propositions adequate to entail $P$. More precisely:

> $A$ makes $P$ true iff $A$ is a set of true simple propositions, $A$ entails $P$, and no proper subset of $A$ entails $P$.

However, the requirement that $A$ entail $P$ is too strong. The difficulty arises in the case where $P$ is a universal generalization. Consider:

> That all of the buttons on the console are pushed is causally sufficient for the light to go on.

In this case, what makes it true that all of the buttons on the console are pushed is the collection of propositions reporting that each button individually is pushed: {'Button 1 is pushed', 'Button 2 is pushed', . . . , 'Button $n$ is pushed'}. But this collection does not entail that all of the buttons are pushed without the addition of a premise telling us that these are all the buttons there are on the console. Let us call a proposition of the latter form an 'enumerative proposition':

(4.13) An enumerative proposition is a proposition of the form $(x)[Fx \supset .x = a_1 \vee \ldots \vee x = a_n]$.

Then our amended definition of 'makes true' is:

(4.14) $A$ makes $P$ true in a world $\beta$ iff $A$ is a set of simple propositions true in $\beta$, and there is a set of $B$ of enumerative propositions true in $\beta$ such that $A \cup B$ entails $P$, and

there is no proper subset $A_0$ of $A$ for which there exists a set $B_0$ of enumerative propositions true in $\beta$ such that $(A_0 \cup B_0)$ entails $P$.

Notice that in plugging 4.14 into 4.12, we do not get the result from 4.12 that the propositions in $A$ and $B$ must be true in the real world. All that is required by 4.12 is that some such sets $A$ and $B$ would contain true propositions *if* $P$ were true. We can make this quite precise by reformulating 4.12 as follows:

(4.15) $P$ is causally sufficient for $Q$ iff $P$ is almost causally sufficient for $Q$, and for each world $\beta$ such that $\beta \mathbf{M} P$, there are sets $A$ and $B$ of simple propositions such that:
(i) $A$ makes $P$ true in $\beta$;
(ii) $B$ makes $Q$ true in $\beta$;
(iii) in $\beta$, the propositions in $A$ are directly causally sufficient for the propositions in $B$.

However, 4.15 runs into difficulty with a kind of 'causal overkill'. Consider:

(4.16) That someone was in the room was causally sufficient for the warning buzzer to sound.

Let us suppose in fact that both Brown and Robinson were in the room. Then there are two distinct minimal sets $A$ making it true that someone was in the room, viz., {'Jones was in the room'} and {'Robinson was in the room'}. All that 4.15 requires is that at least one of these sets be directly causally sufficient for the warning buzzer to sound. Suppose, then, that Jones' presence is causally sufficient to set off the buzzer, but Robinson's presence is not. In such a case, we would not agree that 4.16 is true. What is required for causal sufficiency is not just that *some* minimal set $A$ be directly causally sufficient for the causal consequent, but that *every* minimal set $A$ be directly causally sufficient.

On the other hand, if we turn our attention to the causal consequent, we only want to require that $A$ be directly causally sufficient for the propositions in *some* minimal set $B$. If $A$ is directly causally sufficient for the propositions in some minimal set $B$ which makes $Q$ true, then $A$ is causally sufficient to make $Q$ true in some way or other, and that

is all we want. For example, suppose we have a room lit by three lights numbered 1, 2, 3. Suppose the lights are operated by correspondingly numbered switches, with switches 1 and 2 being in the room and switch 3 being at some remote place outside the room. Suppose lights 1 and 3 are on. Then the following is true:

> That one of the switches in the room is on is causally sufficient for there to be a light on in the room.

What makes it true that there is a switch on in the room is the unit set $A$: {'Switch 1 in on'}. There are two minimal sets which make it true that one of the lights is on, viz., {'Light 1 is on'} and {'Light 3 is on'}. $A$ is only directly causally sufficient for the first of these sets $B$, but clearly that is all that is required. This indicates that we must modify 4.15 to read:

(4.17) $P$ is causally sufficient for $Q$ iff $P$ is almost causally sufficient for $Q$, and for each world $\beta$ such that $\beta \mathbf{M} P$:
(i) there is a set $A$ of simple propositions such that $A$ makes $P$ true in $\beta$;
(ii) for every set $A$ of simple propositions such that $A$ makes $P$ true in $\beta$, there is a set $B$ of simple propositions such that $B$ makes $Q$ true in $\beta$; and the propositions in $A$ are directly causally sufficient in $\beta$ for the propositions in $B$.

Now what remains is to give an analysis of direct causal sufficiency. This is to be a relation of causal sufficiency between sets of simple propositions and simple propositions, and is supposed to provide the means for discriminating between cause and effect by appealing to temporal relations.

An obvious first attempt at defining direct causal sufficiency would be the following:

(4.18) If $A$ is a set of simple propositions and for each simple proposition $Q$, $\sigma(Q)$ is the date of $Q$, then $A$ is directly causally sufficient for the simple proposition $P$ iff $A$ is almost causally sufficient for $P$ and $(Q)(Q \in A \supset \sigma(Q) < \sigma(P))$.[5]

Definition 4.18 embodies the traditional view that a cause must precede its effect. This traditional view is connected with the almost universally held view that causal relations are asymmetric – if $P$ caused $Q$, then $Q$ could not cause $P$. The only obvious way to ensure that causal relations are asymmetric is to build into them the requirement that the cause precede the effect. However, it takes little reflection to see that this requirement is unreasonable. At the very least, cause and effect can occur simultaneously. For example, the five ball's striking the eight ball caused the eight ball to accelerate, and these two events occurred simultaneously. Frequently causes do precede their effects, but not invariably.

Apparently, the most we can require is that the effect not precede the cause. But is even this reasonable? We do say things like, 'The baseball game's being held this afternoon caused the soccer game to be cancelled this morning'. However, it seems wrong to say that it is literally the occurrence of the baseball game that caused the cancellation of the soccer game. Rather, it was the prior *decision* to hold the baseball game that caused the cancellation. Other putative examples of effects preceding causes can generally be handled in this same way. The very idea of an effect preceding its cause seems mindboggling. The only argument I have for this is that it seems required in order to establish the direction of causation. Without this assumption, it does not seem possible to distinguish between cause and effect in many causal contexts. Thus I propose to build this into the analysis of direct causal sufficiency. But we cannot require that the cause *precede* the effect.

But what does this do to the supposed asymmetry of causal sufficiency? It will indeed have the effect, at least on the analysis given here, that causal sufficiency is not asymmetric. But I am not convinced that it should be asymmetric. It seems to me that there are 'feedback examples' in which each of two propositions causes the other. For example, consider two boards arranged in an inverted 'V' so that their tops are leaning against one another. Each board is holding the other one up. What this comes to is that either board's not starting to fall at time $t$ prevents the other board from beginning to fall at time $t$, and hence we have a case of symmetric causation. Consequently, the fact that asymmetry will not result from my analysis of causal sufficiency does not bother me.

If we are agreed that cases of symmetric causation are possible, can we simply modify 4.18 by requiring that the effect not precede the cause, and then accept temporally symmetric cases as cases of symmetric causation? Unfortunately, things are not that simple. Consider the switch and light again, and suppose for the sake of the example that closing the switch *instantaneously* causes the light to come on. Letting $P$ be 'The switch is closed' and $Q$ be 'The light comes on', we have, as before, that $P$ and $Q$ are each almost causally sufficient for the other, and we have that $\sigma(P) = \sigma(Q)$ and hence that cause and effect cannot be distinguished on temporal grounds. But we most assuredly do not want to say that the light's coming on causes the switch to be closed.

In the case in which $\sigma(P) < \sigma(Q)$, the temporal ordering does enable us to distinguish between cause and effect, but as we have seen, it is possible for the cause and effect to be simultaneous, in which case the temporal ordering by itself is not sufficient to make the distinction. However, it will turn out that the distinction in the case of simultaneous cause and effect can be made by making it parasitic on the distinction in the non-simultaneous case. First, let us introduce a technical term to refer to the non-simultaneous case which we already know how to handle:

(4.19) If $A$ is a set of simple propositions and $P$ is a simple proposition, then $A\,\mathsf{SCS}P$ ($A$ is strictly causally sufficient for $P$) iff $A$ is almost causally sufficient for $P$ and $(Q)(Q \in A \supset \sigma(Q) < \sigma(P))$.

Now we consider the case of simultaneous cause and effect. On what basis do we discriminate between the cause $P$ and effect $Q$? I think we do this on the basis of the role $P$ and $Q$ are able to play in temporally extended causal chains. Roughly, in order for $P$ to be causally sufficient for $Q$, it must be possible to cause $Q$ *by* causing $P$, where these latter causal relations involve non-simultaneous cause/effect pairs. For example, it is possible to cause the light to go on by causing the switch to be closed, but it is not possible to cause the switch to be closed by causing the light to go on. This is the basis upon which we say that the closing of the switch causes the light to go on but not vice versa.

But what exactly does it mean to say that it is possible to cause $Q$ by causing $P$. To begin with, the 'possible' in this formula is an existential

quantifier. To say that it is possible to cause $Q$ by causing $P$ is to say that there is something which would be causally sufficient for $Q$ *by* being causally sufficient for $P$. As we are working on the level of simple propositions, the 'something' must be a set of simple propositions, and the causal sufficiency in question should be strict causal sufficiency. But still, what does it mean to say that a set of simple propositions is strictly causally sufficient for $Q$ *by* being strictly causally sufficient for $P$.

To simplify things initially, let us suppose that the set of simple propositions has a single member $R$. Then we want to explain what it means to say that $R$ is strictly causally sufficient for $Q$ by being strictly causally sufficient for $P$. Clearly this requires at least that $R$ is strictly causally sufficient for $P$ and $R$ is strictly causally sufficient for $Q$. In addition it would seem to require that we do not have $\ulcorner(R \ \& \sim P) > Q\urcorner$. That is, although $R$ is sufficient to bring about $Q$, if $\ulcorner(R \ \& \sim P)\urcorner$ were true, then the way in which $R$ would bring about $Q$ has been undermined. The causal chain from $R$ to $Q$ passes through $P$, so if we have $\ulcorner(R \ \& \sim P)\urcorner$, then the causal chain has been broken and there is no reason to expect $Q$ to be true. Can we perhaps take this as our analysis and say that $R$ is strictly causally sufficient for $Q$ by being strictly causally sufficient for $P$ iff $R\mathsf{SCS}P \ \& \ R\mathsf{SCS}Q \ \& \sim[(R \ \& \sim P) > Q]$? Unfortunately, this is still too weak to discriminate between $P$ and $Q$. This is because if $R$ is SCS for $Q$ by being SCS for $P$, then we also have $\ulcorner\sim[(R \ \& \sim Q) > P]\urcorner$. This is because if we had $\ulcorner(R \ \& \sim Q)\urcorner$, this would tell us that the causal chain between $R$ and $Q$ had broken down somewhere, but we would not know whether it had broken down between $R$ and $P$ or between $P$ and $Q$. Consequently, if we had $\ulcorner(R \ \& \sim Q)\urcorner$. then $P$ might be false. For example, in the case of the switch and the light, suppose $R$ reports the setting in motion of a mechanical contrivance which closes the switch by the activation of certain motors and levers. If the mechanical contrivance were set in motion but the light did not come on, then something would have gone wrong, but the malfunction might have been either in the mechanical contrivance or in the wiring between the switch and the light, so we cannot conclude that the switch would have been closed.

Evidently a stronger condition is required to capture what we mean by saying that $R$ is SCS for $Q$ by being SCS for $P$. The nature of this

stronger condition can be seen by realizing that it is always possible to choose $R$ such that we have not just $\ulcorner\sim[(R\ \&\ \sim P)>Q]\urcorner$, but $\ulcorner[(R\ \&\ \sim P)>\sim Q]\urcorner$. Again, if $R$ reports the setting in motion of the mechanical contrivance, then if $R$ is true but the switch is not closed then the light will not go on. But we do not have that if $R$ is true but the light does not go on then the switch is not closed. On the contrary, if $R$ is true but the light does not go on, this might be because something was wrong with the circuit rather than with the mechanical contrivance.

In general, if $P$ and $Q$ are simultaneous and $P$ is causally sufficient for $Q$, then it seems to be the case that we can construct a set $A$ of simple propositions which is SCS for both $P$ and $Q$ and such that $[(\Pi A\ \&\ \sim P)>\sim Q]$. To have the latter condition satisfied, we must include in $A$ both propositions sufficient to bring about $P$ and propositions sufficient to rule out any other ways of bringing about $P$. For example, suppose our light is connected to two switches 1 and 2, and closing either is causally sufficient for the light to be on. Suppose further that switch 2 is closed. Then closing switch 1 is causally sufficient for the light to be on, although it would not cause the light to be on. In this case, in constructing our set $A$ we must include a mechanism for closing switch 1, and we must also include propositions precluding switch 2's being closed. Thus my proposal becomes:

(4.20) If $P$ and $Q$ are simple propositions and $\sigma(P)=\sigma(Q)$, then $P$ is directly causally sufficient for $Q$ iff $P$ is almost causally sufficient for $Q$ and there is a set $A$ of simple propositions such that $A\mathsf{SCS}P\ \&\ A\mathsf{SCS}Q\ \&\ [(\Pi A\ \&\ \sim P)>Q]$.

Notice that in cases of genuinely symmetric causation, like that of the two boards leaning against one another, this gives the result that each is causally sufficient for the other. Let us call the boards 'board 1' and 'board 2'. $P$ reports that board 1 does not begin to fall, and $Q$ reports that board 2 does not begin to fall. Then we want to say that $P$ is causally sufficient for $Q$, and also $Q$ is causally sufficient for $P$. To see that it results from 4.20 that $P$ is causally sufficient for $Q$, we note that we can construct an $A$ describing, e.g., building certain braces around board 1, such that $A$ is causally sufficient for $P$ and thereby for $Q$, but $A$ is such that if we did build the braces but board 1 began to fall anyway, then board 2 would begin to fall. To see that we also have $Q$ causally sufficient for $P$, we note that we can construct another set $A^*$

which is analogous to $A$ but involves bracing board 2 rather than board 1, and then we have $[(\Pi A^* \ \& \ \sim Q) > \sim P]$.

We must generalize 4.20 to include the case in which the causal antecedent is a set of simple propositions rather than a single proposition $P$. This involves a few new difficulties. First, when do we apply the test? The general case not covered by 4.19 is that in which $(Q)(Q \in A \supset \sigma(Q) \leq \sigma(P))$ and $(\exists Q)(Q \in A \ \& \ \sigma(Q) = \sigma(P))$. But it might seem that as long as $Q$ contains a member having a date earlier than $P$, this is sufficient to distinguish cause from effect and hence we need not apply our complex test. However, this is an illusion. For example, $A$ might contain a list of conditions under which one member $Q$ of $A$ would be causally sufficient for $P$, where $\sigma(Q) = \sigma(P)$. We might have that under those conditions, $Q$ would also be almost causally sufficient for $P$ (although not really causally sufficient for $P$). Then letting $A^*$ be the result of replacing $Q$ by $P$ in $A$, we would have that $A^*$ is almost causally sufficient for $Q$. But we would not want to conclude that $A^*$ is causally sufficient for $Q$, even though $A^*$ contains members which predate $Q$. Thus our test must be applied in general if $A$ contains any members having the same date as $P$.

But what precisely does our test amount to when the causal antecedent is a set rather than a single proposition? We might suppose that it requires there to be a set $B$ of simple propositions such that:

(i) $(Q)[Q \bullet A \supset B\mathsf{SCS}Q]$;
(ii) $B\mathsf{SCS}P$;
(iii) $[(\Pi B \ \& \ \sim \Pi A) > \sim P]$.

The difficulty is with requirement (i). That requirement implies that $(Q)(R)[(Q \in B \ \& \ R \bullet A) \supset \sigma(Q) < \sigma(P)]$. But there is no reason to expect this to be possible. There might be no lower bound to the dates of the propositions in $A$. $A$ might contain propositions with dates going back arbitrarily far in time. The solution to this is that in deciding whether $B$ is strictly causally sufficient for $A$, we do not require that all of $B$ be SCS for each member $Q$ of $A$, but rather we require that that part of $B$ which precedes $Q$ be SCS for $Q$. Let us define:

(4.21) If $B$ is a set of simple propositions and $t$ is a time, then $B_t = \{Q;\ Q \in B \ \& \ \sigma(Q) < t\}$.

Then it seems we should replace (i) by:

(i*) $(Q)[Q \in A \supset B_{\sigma(Q)}\mathbf{SCS}Q]$.

Putting all of this together, we obtain our final analysis of direct causal sufficiency:

(4.22) If $A$ is a set of simple propositions and $P$ is a simple proposition, then $A\mathbf{SCS}P$ iff $A$ is almost causally sufficient for $P$ and either;

(i) $(Q)(Q \in A \supset \sigma(Q) < \sigma(P))$ (i.e., $A\mathbf{SCS}P$); or

(ii) $(Q)(Q \in A \supset \sigma(Q) \leq \sigma(P))$ and $(\exists Q)(Q \in A)$ & $\sigma(Q) = \sigma(P))$, and there is a set $B$ of simple propositions such that:

(a) $(Q)[Q \in A \supset B_{\sigma(Q)}\mathbf{SCS}Q]$;

(b) $B\mathbf{SCS}P$;

(c) $[(\Pi B$ & $\sim \Pi A) > \sim P]$.

Finally then, we want to incorporate our analysis of direct causal sufficiency into 4.17 to give us an analysis of causal sufficiency. In order to do this, we must first become aware of the inadequacy of the final clause of 4.17, which requires that if $A$ and $B$ are the sets of simple propositions making $P$ and $Q$ true, then (in order for $P$ to be causally sufficient for $Q$) we must have the propositions in $A$ directly causally sufficient for the propositions in $B$. The first difficulty with this requirement is that the dates of members of $A$ may overlap with the dates of members of $B$. In analogy to 4.22, all we really want to require is that the members of $A$ having dates no later than that of any particular proposition $R$ in $B$ be directly causally sufficient for $R$. That is, defining:

(4.23) If $C$ is a set of simple propositions and $t$ is a time, then $C_{[t]} = \{Q;\ Q \in C$ & $\sigma(Q) \leq t\}$

we require of each $R$ in $B$ that $A_{[\sigma(R)]}$ be directly causally sufficient for $R$. However, this is still too strong a requirement. Consider, for example:

(4.24) Its raining for forty days and forty nights would cause it to be the case that all the unicorns die.

Notice that this does not require that the rain would cause the death of each unicorn – only that the rain would cause the death of each unicorn who would not die anyway. Suppose there are just three unicorns – Mary, Charley, and Vasilly. Let us suppose that the rain could not cause Vasilly to die because he is on top of a mountain, but he will die anyway of old age. Applying our analysis of causal sufficiency to 4.24, what would be required is that if $A$ is the set of simple propositions describing the rain, then $A$ be directly causally sufficient for each member of the set $B = \{$'Mary dies', 'Charley dies', 'Vasilly dies'$\}$. But this is too strong a requirement. $A$ need not be (and is not) directly causally sufficient for 'Vasilly dies' because this is already true and would still be true even if it did rain for forty days and forty nights.

Thus it seems that in order for $P$ to be causally sufficient for $Q$, $A$ need not be directly causally sufficient for every member $R$ of $B$. There may be propositions $R$ which are already true and would still be true even if $P$ were true. Of course, this condition is automatically true if $P$, $Q$, and hence $R$, are already true, e.g., if we transform 4.24 into:

> Its raining for forty days and forty nights caused it to be the case that all the unicorns died.

So, in analogy to our analysis of necessitation, what we actually need is the requirement that, first, $REP$, and second, this would still be true even if $(\sim P \ \& \ \sim Q)$; i.e., $(REP)E(\sim P \ \& \ \sim Q)$.

Thus our final analysis of causal sufficiency becomes:

(4.25) $P$ is causally sufficient for $Q$ iff $P$ is almost causally sufficient for $Q$, and for each world $\beta$ such that $\beta \mathbf{M} P$:

(i) there is a set $A$ of simple propositions such that $A$ makes $P$ true in $\beta$;

(ii) for every set $A$ of simple propositions such that $A$ makes $P$ true in $\beta$, there are sets $B_1$ and $B_2$ of simple propositions such that $(B_1 \cup B_2)$ makes $Q$ true in $\beta$, and:

(a) $B_1 \neq \varnothing$ and $(R)[R \in B_1 \supset (A_{[\sigma(R)]}\mathbf{DCS}R \text{ in } \beta)]$;

(b) $(R)(R \in B_2 \supset [(REP)E(\sim P \ \& \ \sim Q) \text{ in } \beta])$.

Our analysis has become rather complicated, but the basic idea is really quite simple. Causal relations between simple propositions are

just contingent non-epiphenomenal necessitations with the appropriate temporal order. The causal relations between logically complex propositions are just contingent non-epiphenomenal necessitations underlain by causal relations between the simple propositions which make the complex propositions true.

## 5. Remarks on the analysis

I have developed an analysis of causal concepts in terms of subjunctive conditionals. This analysis may be regarded as a kind of regularity theory, although it is more complicated than a simple Humean-type theory. Its claim to being a regularity theory comes from the way subjunctive conditionals arise out of subjunctive generalizations. Because of its additional complexity, the present theory is not subject to the traditional difficulties regarding regularity theories. For example, consider two popular counterexamples. Many simple regularity theories entail that night is the cause of day. Its being night at $t_1$ does indeed necessitate its being day at $t_2$, but this is not a causal necessitation because these are epiphenomena. They both are the causal effects of certain facts about the earth and the sun, and its being night at $t_1$ without those facts being true would not necessitate its being day at $t_2$.

Many regularity theories also entail that birth is the cause of death. The reason this does not result from the present theory is more complicated than in the case of the previous example. In this case the necessitation fails. A person's being born does necessitate that he will sometime die, but this is not the necessitation that is required for birth to cause death. By principle 4.25, a person's being born in the way he is at a certain time $t_1$ (i.e., the conjunction of the simple propositions describing his birth) must necessitate his dying at the time he does. But there is no such necessitation. A person's being born at a certain time necessitates that he will sometime die, but it does not necessitate that he will die at any particular time. Thus birth does not cause death.

The literature is full of sophisticated counterexamples to earlier analysis (e.g. Scriven, 1966; Kim, 1971, 1973, 1973a). However, the present analysis has been designed explicitly to handle those counterexamples and I believe that it does so successfully.

It has frequently been maintained that causal concepts embody a contextual element, and that this must be included in any adequate analysis of these concepts. Scriven (1974) supposes that 20% of the people exposed to heavy doses of ultraviolet radiation develop skin cancer, with certain hereditary factors determining which people will develop cancer when exposed to the radiation. Given a man who has developed cancer upon exposure to the radiation, if we ask what caused him to develop cancer, the answer might be either 'He was exposed to heavy doses of radiation' or 'He had the hereditary factors making him susceptible to cancer when exposed to radiation'. Which answer is appropriate depends upon what we are interested in. The difference is that between the two questions, 'Why did he develop cancer *now* when he did not do so before?', and, 'Why did *this* man develop cancer when others who were exposed to the radiation did not?' Scriven and Mackie (1965) suppose this to show that a statement of the form ⌜$X$ caused $Y$⌝ is really elliptical for a more complicated statement of the form ⌜$X$ caused $Y$ relative to the contrast class $C$⌝. If we want to know why the man developed cancer *now*, the contrast class consists of moments of his history, whereas if we want to know why *he* developed cancer when others who were exposed to the radiation did not, then the contrast class consists of men exposed to the radiation.

I think that Scriven and Mackie are onto something, but we must be careful to distinguish between statements of the form ⌜$X$ caused $Y$⌝ and statements of the form ⌜$X$ was the cause of $Y$⌝. The latter locution is one I have carefully avoided up to this point. First consider statements of the form ⌜$X$ caused $Y$⌝. If we ask, 'What caused this man to develop skin cancer?' the context may make it quite inappropriate to reply, 'His being exposed to radiation caused him to develop skin cancer'. But the inappropriateness of a reply does not establish its falsehood. Frequently, a reply is inappropriate simply because it is not helpful. For example, if the question is asked in the context of a discussion of people who were exposed to a heavy dose of radiation in an industrial accident and some of whom developed skin cancer while others did not, then it would be quite unhelpful to reply that the radiation caused one of these people to develop skin cancer. But the reason it is unhelpful is that we already know that the radiation caused

the cancer. The radiation did cause cancer in some of the people, and did not cause cancer in others. But what we are looking for is another factor which explains this selectivity, i.e., which caused cancer in all and only those people (in the group) in whom the factor was present.

I think that any contextual element in statements of the form ⌜$X$ caused $Y$⌝ is not part of the meaning of the statement, but rather is part of the pragmatics of language in general. However, this changes when we turn to the peculiar statement form ⌜$X$ was the cause of $Y$⌝. The reason this is peculiar is that there can simultaneously be more than one thing which is the cause of something. For example, it may be true both that the cause of the explosion was the room's being filled with gas, and that the cause of the explosion was Jones' throwing the switch. This makes the use of the definite article rather strange. The only way I can see to make sense of this is to suppose that statements about 'the cause' of something really are elliptical for more complicated statements involving a contrast class. However, I will not pursue this further at this time.

## 6. The logic of causes

If the present analysis of causal relations is acceptable, it settles a number of questions about the logical properties of these relations. To begin with, the analysis leads immediately to interchange principles for logical equivalence:

(6.1) $P \leftrightarrow Q \;\&\; PCR. \supset QCR$

(6.2) $P \leftrightarrow Q \;\&\; RCP. \supset RCQ$

(6.3) $P \leftrightarrow Q \;\&\; PCSR. \supset QCSR$

(6.4) $P \leftrightarrow Q \;\&\; RCSP. \supset RCSQ$

These interchange principles, in turn, imply that substitutivity of identity fails for causal contexts. That is, the following two principles fail:

(6.5) $t_1 = t_2 \;\&\; (Ft_1)CSQ. \supset (Ft_2)CSQ$

(6.6) $t_1 = t_2 \;\&\; PCS(Ft_1). \supset PCS(Ft_2)$

where if $t_1$ or $t_2$ is a definite description, it is understood to have

narrow scope. The failure of 6.5 and 6.6 can be seen as follows. Suppose $\ulcorner Ft_1 \urcorner$ is causally sufficient for $Q$, where $\ulcorner Ft_1 \urcorner$ and $Q$ are both false. Then let $t_2$ be the definite description $\ulcorner \imath x(x = t_2 \ \& \sim Q) \urcorner$. Then $\ulcorner t_1 = t_2 \urcorner$ is true. Hence, assuming 6.5, $\ulcorner (Ft_2)\mathbf{CS}Q \urcorner$ is true. But $\ulcorner Ft_2 \urcorner$ is equivalent to $\ulcorner Ft_1 \ \& \sim Q \urcorner$. Hence by 6.3, $(Ft_1 \ \& \sim Q)\mathbf{CS}Q$. By our analysis, this implies $\ulcorner (Ft_1 \ \& \sim Q) > Q \urcorner$, which within *SS* implies that $\ulcorner Ft_1 \urcorner$ logically entails $Q$. This result is absurd, so principle 6.5 must fail. Principle 6.6 fails for analogous reasons.

The following two principles are equivalent to one another, and it may seem that they should hold:

(6.7) $P\mathbf{C}(Q \ \& \ R) \supset [P\mathbf{C}Q \ \& \ P\mathbf{C}R]$.

(6.8) $[P\mathbf{C}Q \ \& \ Q \rightarrow R] \supset P\mathbf{C}R$.

However, neither of these principles holds. For example, it might be true that John's hitting Joe caused Joe to have a black eye. That Joe has a black eye entails that Joe has an eye. But John's hitting Joe did not cause Joe to have an eye. Thus principle 6.8, and hence the equivalent 6.7, are both false. Slight modifications of this example yield counterexamples to the analogous principles regarding 'would cause' and causal sufficiency.

I think that the failure of 6.7 seems particularly surprising because in English, if we say, e.g., 'Closing the switch caused both lights to be on', this is ambiguous between 'Closing the switch caused it to be the case that both lights are on' and 'Closing the switch caused each light to be on'. The reason 6.7 fails is that $P$ can cause a conjunction to be true simply by causing the truth of that part of it which isn't already true. E.g., if one of the lights is already on, and the switch is wired to the other light, then closing the switch will cause it to be the case that both lights are on (where before only one was on).

There are some other distribution principles which fail for much the same reason. The following both fail:

(6.9) $P\mathbf{C}(x)(Fx \supset Gx) \supset (x)[Fx \supset P\mathbf{C}Gx]$.

(6.10) $(\exists x)[Fx \ \& \ P\mathbf{C}Gx] \supset P\mathbf{C}(\exists x)(Fx \ \& \ Gx)$.

We have already seen a counterexample to 6.9 in the distinction between 'Its raining for forty days and forty nights caused all the unicorns to die' and 'Its raining for forty days and forty nights caused it

to be the case that all the unicorns died'. The former requires that each unicorn was killed by the rain, but the latter only requires that the rain killed all the unicorns that wouldn't have died anyway. To generate a counterexample to 6.10, suppose we have a room containing two lights, one of which is on and the other of which is off and controlled by a switch. We throw the switch, thereby causing the second light to come on. Then there is a light such that throwing the switch caused it to come on, but it is not true that throwing the switch caused there to be a light that was on (because one of the lights would have been on anyway).

It has generally been supposed that causal relations are transitive. For example, David Lewis (1973b) explicitly builds this into his analysis of 'would cause'. But transitivity does not result from my analysis. Nor should it. None of the following three principles is correct:

(6.11) $[P\mathbf{CS}Q \ \& \ Q\mathbf{CS}R] \supset P\mathbf{CS}R.$

(6.12) [$P$ would cause $Q$ & $Q$ would cause $R$] $\supset$ $P$ would cause $R$.

(6.13) $[P\mathbf{C}Q \ \& \ Q\mathbf{C}R] \supset P\mathbf{C}R.$

It is really quite simple to get counterexamples to 6.11 and 6.12. These principles fail for the same reason that transitivity fails for subjunctive conditionals. For example, let us suppose that it is a chemical process $C$ which causes a match in the presence of oxygen to light upon being struck. The chemical process $C$ (we can suppose) can occur with or without there being oxygen present, but it is only when oxygen is present that the occurrence of the process causes the match to ignite. Now consider a match which is in fact in the presence of oxygen and which would still be in the presence of oxygen even if chemical process $C$ were to occur. Then the following three statements are all true:

(1) The occurrence of chemical process $C$ would cause this match to ignite.

(2) This match being struck in the absence of oxygen would cause chemical process $C$ to occur.

(3) This match being struck in the absence of oxygen would not cause this match to ignite.

Consequently, 'would cause' is not transitive. And this example is also

a counterexample to the transitivity of causal sufficiency. Transitivity fails for these two causal relations for precisely the same reason it fails for subjunctive conditionals. This is because changing causal antecedents involves 'world hopping'. That is, evaluating the truth values of causal statements with different antecedents involves evaluating the truth values of subjunctive conditionals with different antecedents, and that in turn involves our looking at different possible worlds.

Although it is quite obvious that transitivity fails for both causal sufficiency and 'would cause', it may not seem so obvious that it fails for 'caused'. This is because in the case of 'caused' we are dealing with true causal antecedents and consequents, and hence one is apt to suppose that there is no world hopping involved. However, this is a mistake. $\ulcorner P \mathbf{C} Q \urcorner$ entails $\ulcorner (\sim P \ \& \ \sim Q) > (P > Q) \urcorner$, and so in spite of appearances, world hopping is involved. This can be substantiated by seeing how our analysis actually leads us to intuitive counterexamples to 6.13. Characteristically, $Q$ only causes $R$ because some third proposition $S$ is true. In such a case, it is required of $S$ that $[(\sim Q \ \& \ \sim R) > (Q > S)]$. However, even though $P$ caused $Q$, if it is not true that $[(\sim P \ \& \ \sim R) > (P > S)]$, then there is no reason to expect $P$ to have caused $R$. Let me give two counterexamples that take this form:

(1) Suppose there are two kinds of gasoline. One kind gives an engine extra power, but causes it to overheat if run for a long time. The other kind gives less power, but allows sustained running of the engine. Suppose there is an engine which was running, and I filled the almost empty fuel tank with high-power gasoline. That I filled the fuel tank caused the engine to run for five more hours. And the engine's running for five more hours caused it to overheat (because it was burning high-power gasoline). But, that I filled the fuel tank did not cause the engine to overheat. Rather, that I filled the fuel tank *with high-power gasoline* caused the engine to overheat. Thus, 'causes' is not transitive. This counterexample arises because it is not true that if I had not filled the tank and the engine had not overheated, then had I filled the tank, the engine would have been burning high-power gasoline.

(2) Consider a 'vacuum oven'. This is a chamber which can be evacuated and then heated. Suppose this oven is operated by two buttons. Pushing button $A$ results in any gas in the oven being pumped

out, and then five minutes after the button is pushed the heating element comes on and heats the contents of the oven to a very high temperature. Pushing button $B$ results in the gas being left in the oven, so that nothing happens for five minutes, and then the heating elements come on. Suppose now that the oven is filled with pure oxygen and contains a piece of paper. I push button $B$, after five minutes the heating element comes on, and because of the oxygen environment this causes the paper to burst into flame. In this case we have the following: (1) That the heating element came on, caused the paper to ignite; (2) that there was a control button that was pushed, caused the heating element to come on; but, that there was a control button that was pushed, did not cause the paper to ignite. Rather, that button $B$ was pushed caused the paper to ignite. Thus, once again we have a failure of transitivity. The reason that there being a button that was pushed did not cause the paper to ignite is that it was not true that if no button had been pushed and the paper had not ignited, then if some button had been pushed then the oven would have been filled with oxygen. On the contrary, if it were true that no button was pushed, then had some button been pushed it might have been button $A$ that was pushed, in which case the oven would not have been filled with oxygen.

Apparently none of our causal relations are transitive. This is certainly surprising in light of the almost universally held belief that they are transitive, but the conclusion seems inescapable. This has important implications for a great many of our favorite beliefs about causes. For example, it makes nonsense of our ordinary way to thinking about causal chains. The 'received view' of causal chains is that they arise when we have, between two propositions $P$ and $Q$, a sequence $P_1, \ldots, P_n$ of propositions such that $P\mathsf{C}P_1$, $P_1\mathsf{C}P_2, \ldots, P_n\mathsf{C}Q$, and 'hence', $P\mathsf{C}Q$. But, as we have just seen, the 'hence' doesn't follow. Does this mean that all talk of causal chains is nonsense? It does not seem that it should. We can often talk about the 'way' in which one proposition $P$ causes another proposition $Q$ to be true. This 'way' involves tracing out a sequence of intermediate causal links. To take a straightforward example, consider a row of dominoes standing on end. Pushing over the first domino causes the final domino to fall over, and it does so by causing each intermediate domino to fall

in turn. How are we to understand this in light of the failure of transitivity for causes?

I suggest that we can make sense of causal chains by concentrating on the notion of $P$ causing $Q$ *by* causing $R$. I propose that what this means is that $P$ caused $Q$, $P$ caused $R$, and that $P$ without $R$ would not have been causally sufficient for $Q$:

(6.14) $P$ caused $Q$ by causing $R$ iff $PCQ$ & $PCR$ & $\sim[(P$ & $\sim R)\mathbf{CS}Q]$.

Then in a causal chain, what happens is that the first link causes each successive link by causing the preceding links:

(6.15) A sequence $\langle P_1, \ldots, P_n\rangle$ of propositions is a causal chain iff $P_1$ caused $P_2$, and for each $k$ such that $2 < k \leq n$, $P_1$ caused $P_k$ by causing $(P_2$ & ... & $P_{k-1})$.

In summary, philosophers have held a remarkable number of false beliefs about the logical properties of causal relations. However, if we are careful we are now in a position to sort out what is true and what is false.

## NOTES

[1] In point of fact, I do not think that the narrow-scope readings 2.15* and 2.16* are plausible interpretations of the English sentences 2.15 and 2.16, but that does not affect the logical point being made.

[2] The condition $\ulcorner\Diamond P\urcorner$ is included so that we are not forced to say that a contradiction would cause everything.

[3] It is worth noting that 3.2 is entailed by 3.1 together with the principle that $\ulcorner PCQ\urcorner$ entails $P$. Thus it need not be defended separately from 3.1.

[4] The statement that the heart attack caused the man to die today is ambiguous between the true statement that there was a time today such that the heart attack caused him to die at that time, and the false statement that the heart attack caused there to be a time today when he died.

[5] Strictly speaking, definition 4.18 does not make sense because almost causal sufficiency has been defined as a relation between individual propositions and not as a relation between sets of propositions and propositions. However, this is trivially rectified by simply replacing '$P$' by '$\Pi A$' in definition 4.10.

CHAPTER VIII

# PROBABILITIES

## 1. Introduction

At this point I must be candid and admit that everything that has come before may be just so much science-fiction. This is because there may be no true, exceptionless, laws. It may be that all putative laws are really just approximations to general probability statements. For example, this is the way much of quantum mechanics has gone. Should it come to pass that there really are no true general laws, then the counterfactual conditionals, necessitation conditionals, and causal statements that we base upon putative general laws will all be false. All of the logical framework developed above will become of only theoretical interest. In this eventuality, our law statements and counterfactuals must be replaced by various kinds of probability statements.

Many philosophers are prone to suppose that the situation I have just described in which there are no general laws is almost certainly the situation we are really in. They defend this view by pointing in an off-hand way to quantum mechanics and observing that the probabilistically governed behavior of elementary particles must underly everything else. However, this involves at least a partial misconception of the nature of quantum mechanics. It is certainly true that some of the fundamental laws of quantum mechanics are probabilistic, but it is often overlooked that quantum mechanics embodies lots of other laws that are not probabilistic. Among these are conservation laws (the conservation of energy, momentum, angular momentum, spin, strangeness, etc.), laws regarding relativistic transformations, laws regarding the charges, masses, spins, etc., of elementary particles, laws governing weak, strong, and Coulomb forces, and so on. There are actually more non-probabilistic laws in quantum mechanics then there are probabilistic laws, and the prospects of this changing seem remote. It would alter the entire character of quantum mechanics if it were decided, e.g., that the charge on an electron is not fixed but is instead represented by some sort of probabilistic distribution centered around what is presently believed to be the charge of an electron.

The situation I have just described in quantum mechanics, in which we have a mixture of probabilistic and non-probabilistic laws, seems rather likely to be representative of the true state of the world (although I am not totally convinced that the world is not governed instead by strictly deterministic laws). It will turn out that even in this case some changes are required in our account of subjunctive conditionals. In Chapters IV and VI I implicitly made the simplistic assumption that there were no basic probabilistic laws. In this chapter we will investigate what happens if we reject that assumption.

Probability enters into our account of subjunctive conditionals in two ways. First, countenancing basic probabilistic laws will force us to modify our definition of the relation '$\beta\mathbf{M}_{\alpha}\varphi$'. We must get clear on the logical nature of basic probabilistic laws and see how they are involved in the relation **M**. Second, it will turn out that many subjunctive conditionals that we are apt to assert are not really true. What are true instead are a variety of probabilistic statements like ⌜If it were true that $P$, then it would probably be true that $Q$⌝, ⌜If it were true that $P$, it would almost certainly be true that $Q$⌝, etc. We must get clear on the nature of these probability statements and how they are related to subjunctive conditionals. These will be the two tasks undertaken in this chapter.

## 2. Indefinite probabilities

Basic probabilistic laws report 'indefinite probabilities'. An *indefinite* probability statement has the form ⌜The probability of *an* $A$ being a $B$ is $r$⌝. Indefinite probability statements do not report the probability of a proposition, but rather concern predicates or open formulas. The probability of an $A$ being a $B$ might reasonably be symbolized as ⌜prob$(Bx/Ax)$⌝.

### 2.1. *Relative Frequencies*

The traditional view on indefinite probabilities is that they are relative frequencies or limits of relative frequencies. If there are just $n$ $A$'s, and $m$ of them are $B$'s, then prob$(Bx/Ax) = m/n$. In this case, the indefinite

probability is identified with the relative frequency of $B$'s in $A$'s. If there are infinitely many $A$'s, and infinitely many of them are $B$'s, then the relative frequency of $B$'s in $A$'s is no longer well-defined, so instead the traditional view identifies $\text{prob}(Bx/Ax)$ with the limit of the relative frequency of $B$'s in larger and larger finite subsets of the set of all $A$'s. In this case, $\text{prob}(Bx/Ax)$ is said to be the limit of the relative frequencies.

I suspect that most people find the above view unobjectionable in the case in which there are only finitely many $A$'s. But there is a well-known problem for the case in which there are infinitely many $A$'s. If there are infinitely many $A$'s, then we are supposed to consider a sequence $A_0, A_1, \ldots, A_n, \ldots$ of finite subsets of $A$ such that (1) for each $i, j$, if $i<j$ then $A_i \subseteq A_j$, and (2) $\bigcup_{i\in\omega} A_i = A$, and then letting $\text{Car}(X)$ be the cardinal of a set $X$, we define:

$$\text{prob}(Bx/Ax) = \lim_{i\to\infty} \frac{\text{Car}(A_i \cap B)}{\text{Car}(A_i)}.$$

The problem is that the limit of relative frequencies that we get in this way depends upon the particular sequence $\{A_i; i \in \omega\}$ that we choose. It is quite possible to choose another sequence $\{A_i^*; i \in \omega\}$ of subsets which gives a different limit for the relative frequency. For example, consider a very durable coin which lasts forever and, over all time, is tossed infinitely many times. Letting $H$ be the set of tosses of this coin which comes up heads, and letting $A_i$ be the set of the first $i$ tosses of this coin, it is quite plausible to suppose that

$$\lim_{i\to\infty} \frac{\text{Car}(A_i \cap H)}{\text{Car}(A_i)} = 1/2.$$

But now, suppose we define $A_i^*$ to be the set of the first $i$ tosses resulting in heads and the first $2i$ tosses resulting in tails. Then we still have $\bigcup_{i\in\omega} A_i^* = A$, but now

$$\lim_{i\to\infty} \frac{\text{Car}(A_i^* \cap H)}{\text{Car}(A_i^*)} = 1/3.$$

Thus we cannot define $\text{prob}(Bx/Ax)$ in terms of just any sequence of subsets $\{A_i; i \in \omega\}$.

Perhaps it is generally supposed that the solution to this problem is

that certain sequences of subsets of $A$ are 'natural', and others are 'unnatural'. For example, it does seem that the sequence $\{A_i; i \in \omega\}$ above was natural, and the sequence $\{A_i^*; i \in \omega\}$ was unnatural. Thus it may be felt that if we can somehow characterize which sequences are natural, we can solve the problem of how to define prob$(Bx/Ax)$ in terms of limits of relative frequencies.

I think, however, that there are at least three reasons why this will not work. First, in the infinite case again, if the infinitude of $A$'s is of a temporal origin, there being only finitely many $A$'s at any one time but infinitely many $A$'s over all time, then just as in the case of the coin there is a natural sequence of subsets defined by considering all the $A$'s existing in progressively broader temporal intervals. But suppose instead that there are infinitely many $A$'s at a single instant. For example, astronomers once believed that there were infinitely many stars, from which it would also follow that there are infinitely many elementary particles, probably infinitely many atoms and molecules, perhaps infinitely many rocks, trees, clouds, etc. In this case, choosing any particular sequence of subsets over all others seems quite arbitrary. Of course, certain sequences of subsets artificially contrived for the sole purpose of getting strange limits of relative frequencies do seem unnatural, but this rules out very few sequences of subsets. Among those sequences of subsets which are not so artificially contrived, we may not be able to prove that they lead to different limits of relative frequencies, but there is no a priori reason to think that they will not.

The difficulty in defining the limit of the relative frequency in the infinite case is, I suspect, insurmountable. However, surprisingly enough, there is an even more obvious difficulty for the finite case. Consider a particular coin which is subjected to very careful physical examination and ascertained to be a completely fair coin. That is, we may conclude on the basis of our knowledge of physics that the probability of a flip of this coin resulting in heads is 1/2. But suppose the coin is never flipped. It is melted down shortly after the examination. Then the relative frequency does not exist, but we would still regard the probability statement as true. Or suppose the coin is flipped just once, resulting in a head, and then melted down. Then the relative frequency is 1, but we would still insist that the probability is 1/2. It is

just not true that in the finite case, the probability of an $A$ being a $B$ is the same thing as the relative frequency of $B$'s in $A$'s. What is happening here is that, just as in the case of non-probabilistic laws, these indefinite probability statements are subjunctive in character. They are not just about actual $A$'s, but also about physically possible $A$'s. To say that the probability that a flip of our coin will result in heads is 1/2 is to make a statement about all possible flips of the coin, and not just about those few flips that actually take place. If the coin is only flipped a few times, then the relative frequency will almost certainly not be the same as the probability. Furthermore, the probability statement supports conclusions about physically possible flips of the coin that do not actually occur. E.g., we can conclude from the probability statement that if, instead of melting the coin down, we had flipped it one hundred more times, out of those one hundred flips we would probably have gotten in the vicinity of fifty heads.

### 2.2. *Subjunctive Indefinite Probabilities*

What the relative frequency theory overlooks is that our indefinite probability statements are subjunctive in the same way subjunctive generalizations are. As such, they cannot be identified with summations of actual indicative states of the world, any more than subjunctive generalizations can be identified with material generalizations. In fact, as we will see below, subjunctive generalizations can be regarded as limiting cases of indefinite probability statements.

Just as there is a distinction between strong and weak subjunctive generalizations, there is a distinction between two kinds of indefinite probability statement. Consider our infinitely durable coin which is flipped infinitely many times. Let us suppose that physical examination shows it to be perfectly balanced and shaped, etc., so that the probability of a flip yielding a head is 1/2. Let $D$ be a description of all of the relevant physical characteristics of the coin. Thus what we have is that the probability of a flip of a coin of description $D$ yielding heads is 1/2. Now let us suppose in addition that there is just the one coin of description $D$. Furthermore, suppose that over the years a secret society has grown up which worships this particular coin. It is part of their mystical beliefs that only 1/3 of the flips of this coin should yield

heads, and so to ensure this they enclose the coin in an infinitely durable coin-flipping machine which, through the use of magnetic fields, biases the flips in such a way that the relative frequency of heads always hovers right around 1/3. What are we to say now about the probability of a flip of a coin of description $D$ yielding a head? On the one hand it is reasonable to say, on the basis of the physical description $D$, that the probability is 1/2. After all, as I have argued, the probability statement is as much about physically possible tosses of coins of description $D$ as it is about actual tosses of such coins, and it is only an accident that there is only one coin of description $D$ and it finds itself in such peculiar circumstances. On the other hand, it also seems reasonable to say that, *because* there is only one such coin and it is enclosed in the sacred machine, the probability is only 1/3. This is still a subjunctive statement because even if the coin had been flipped more times that it actually was, it would still have been in the machine and hence the relative frequency would still have remained around 1/3.

I think that what is happening here is that we must make a distinction between two kinds of indefinite probabilities – strong ones and weak ones. The distinction is parallel to the distinction between strong and weak subjunctive generalizations. The strong subjunctive generalization $\ulcorner(Fx \Rrightarrow Gx)\urcorner$ is about all physically possible circumstances in which we might encounter an $F$. On the other hand, the weak generalization $\ulcorner(Fx \Rightarrow Gx)\urcorner$ is about a more narrowly circumscribed set of circumstances – the 'actually possible' circumstances in which we might encounter an $F$. As we have seen, weak generalizations are parasitic upon strong generalizations. $\ulcorner(Fx \Rightarrow Gx)\urcorner$ is true *because of* certain physically contingent facts $P$ about the world, and what makes $\ulcorner(Fx \Rightarrow Gx)\urcorner$ true is that $\ulcorner[(Fx \;\&\; P) \Rrightarrow Gx)\urcorner$ is true. For example, given Chisholm's bottle of rat poison, what makes it true that anyone who drank from this bottle would be poisoned, is that the bottle does contain rat poison and that the strong generalization 'Anyone who drank from a bottle containing rat poison would be poisoned' is true.

Analogously, there are two kinds of subjunctive indefinite probability statements. On the one hand there are *strong indefinite probability statements*, henceforth symbolized $\ulcorner \mathrm{prob}_S(Gx/Fx) = r\urcorner$, which take into account all physically possible circumstances in which there could be

an $F$. These probability statements are physically necessary – as they take into account all physically possible circumstances, they will remain true in all worlds having the same physical laws. On the other hand, there are *weak indefinite probability statements*, symbolized $\ulcorner \text{prob}_W(Gx/Fx) = r \urcorner$, which rely for their truth on physically contingent facts about the world. Just as in the case of weak subjunctive generalizations, weak indefinite probability statements are parasitic upon strong indefinite probability statements. In general, if $P$ is the physically contingent proposition the truth of which makes the weak probability what it is, then $\text{prob}_W(Gx/Fx) = \text{prob}_S(Gx/(Fx \ \& \ P))$. For example, in the circumstances described above, the weak probability of a flip of a coin of description $D$ yielding a head is the strong probability of a flip of a coin of description $D$ yielding a head under the circumstances described. Thus in the above circumstances, the strong probability of a flip of a coin of description $D$ yielding heads is 1/2, but the weak probability is 1/3.

It must be pointed out that there is a third kind of indefinite probability statement. For example, consider an urn containing seven white balls and three black balls. We want to say that the probability of a ball in this urn being black is 3/10. However, this probability is not subjunctive. It is about a fixed set of objects (the objects actually in the urn), and as such it implies nothing about the probability of other balls being black if they were in the urn. This is a *material indefinite probability statement.* Let us symbolize these as $\ulcorner \text{prob}_M(Gx/Fx) = r \urcorner$. It is worthwhile to compare our three kinds of indefinite probability statements with the three kinds of generalizations we have encountered. There are strong subjunctive generalizations, weak subjunctive generalizations, and material generalizations, and correspondingly there are strong indefinite probability statements, weak indefinite probability statements, and material indefinite probability statements. The generalizations are, in the following sense, limiting cases of the corresponding probability statements:

(2.1) $(Fx \Rrightarrow Gx) \rightarrow \text{prob}_S(Gx/Fx) = 1.$

(2.2) $(Fx \Rightarrow Gx) \rightarrow \text{prob}_W(Gx/Fx) = 1.$

(2.3) $(x)(Fx \supset Gx) \rightarrow \text{prob}_M(Gx/Fx) = 1.$

For familiar reasons, these entailments do not go in the other direction. For example, if $h$ is a real-valued function of $F$'s, and the values of $h$ are evenly distributed over some finite interval, then for each particular $r$ in that interval, $\text{prob}_W(h(x) = r/Fx) = 0$, but this does not imply that $(Fx \Rightarrow h(x) \neq r)$. To suppose otherwise would lead to the absurd result that $(y)(Fx \Rightarrow h(x) \neq y)$.

Some unrepentant frequentists, having noted material indefinite probabilities, may assert that those are the probabilities they have been talking about all along. But they would be wrong. For example, consider the urn with seven white balls and three black balls. We can talk about the material probability of a ball in the urn being black, but that is not the probability concerning that urn which philosophers have generally wanted to talk about. It has been far more common for them to talk about the probability that drawing a ball from that urn will produce a black ball. This is an altogether different probability. To begin with, it is subjunctive. It is not just about all actual draws that have been performed, but also about what would happen if there were more draws. Furthermore, this probability need not be the same as the material probability of a ball in the urn being black. For example, if the black balls are a different size than the white balls, this may affect the probability of a black ball being drawn, but it does not affect the probability of a ball in the urn being black. Although material probabilities make sense (at least in the finite case – the infinite case is problematic), I think it must be concluded that they are not of much use and that they are not the probabilities in which we are generally interested.

### 2.3. *Strong Indefinite Probabilities*

As we will see in the next section, weak indefinite probabilities can be defined in terms of strong indefinite probabilities. So now let us examine strong indefinite probabilities. $\ulcorner\text{prob}_S(Gx/Fx)\urcorner$ will not be defined for all predicates $F$ and $G$. $\text{prob}_S(Gx/Fx)$ is intended to be an *indefinite* probability, but there are clear examples of predicates $F$ and $G$ for which $\text{prob}_S(Gx/Fx)$, if it were defined, would be a definite probability rather than an indefinite probability. For example, $\text{prob}_S(Ha/x = a)$, if it were defined, would be the probability of the

proposition *Ha*. Of course, definite probabilities make perfectly good sense, and will be discussed in due course, but the point is that they are not what the function $\text{prob}_S$ gives us. $\text{Prob}_S$ is an indefinite probability. Accordingly, there must be some restrictions placed on the arguments of $\text{prob}_S$.

Unfortunately, it is not at all obvious what the requisite restrictions are, and I will not attempt to elicit them here. But one thing at least seems clear. If there is an $E$ such that $\text{prob}_S(Ex/Fx)$ exists, and if there is an $H$ such that $\text{prob}_S(Gx/Hx)$ exists, then both $\text{prob}_S(Gx/Fx)$ and $\text{prob}_S(Fx/Gx)$ exist. This is to say that there is a set $\Pi$ of predicates such that $\text{prob}_S(\psi/\varphi)$ exists iff $\psi, \varphi \in \Pi$. It is worth noting just how broad the class $\Pi$ really is. For example, at least in the case where $\underset{p}{\Diamond}(\exists x)Fx$, it is natural to read $\ulcorner\text{prob}_S(Gx/Fx)\urcorner$ as the proportion of physically possible $F$'s that are $G$'s. This would naturally lead us to suspect that $\text{prob}_S(Gx/Fx)$ does not exist if $\sim\underset{p}{\Diamond}(\exists x)Fx$. But that would be a mistake. For example, 'black holes' are supposed to be astronomical phenomena resulting from the gravitational collapse of stars and consisting of regions of space from which nothing (not even light) can escape. Thus it is supposed to be physically impossible to retrieve something from a black hole. Nevertheless, it makes perfectly good sense to talk about the probability of heads resulting from a flip of a coin of physical description $D$ which was retrieved from a black hole. If $D$ ensures that it is a fair coin, then the probability is 1/2.

Even in the case in which $\sim\underset{p}{\Diamond}(\exists x)Fx$, $\text{prob}_S(Gx/Fx)$ exists, although this is perhaps best viewed as a convention. At least ordinarily, we want to say that the logical entailment of $Gx$ by $Fx$ ensures that $\text{prob}_S(Gx/Fx) = 1$. If $\sim\underset{p}{\Diamond}(\exists x)Fx$, then $(x)(Fx \rightarrow Gx)$, and so for any $G$ at all in $\Pi$, we should have $\text{prob}_S(Gx/Fx) = 1$.

Our indefinite probabilities are conditional probabilities. They are probabilities of certain things *given* other things. The traditional approach to conditional probabilities was to define them in terms of non-conditional or *absolute probabilities*. The absolute probability $\text{prob}(Fx)$ is supposed to be the probability of *anything at all* being an $F$. Absolute probabilities can be defined easily enough in terms of conditional probabilities:

$$\text{prob}(Fx) = \text{prob}_S(Fx/Gx \vee \sim Gx).$$

The traditional move was to somehow start with absolute probabilities and then define conditional probabilities as follows:

$$\text{prob}_S(Gx/Fx) = \frac{\text{prob}(Gx \,\&\, Fx)}{\text{prob}(Fx)}.$$

The recognized difficulty with this approach is that it does not work in the case in which $\text{prob}(Fx) = 0$. In that case, $\text{prob}_S(Gx/Fx)$ would be left undefined. Hence this approach cannot handle the case in which $F$ is logically impossible, and presumably cannot handle the case in which $F$ is physically impossible. But I think it is generally supposed that this traditional approach will work for any more normal case. In fact, however, I think that absolute probabilities really make very little sense. If we define them as above (and really mean what we say so that they are not conditional on some additional implicit and non-tautological condition like that of being a physical object), then the absolute probability for any normal predicate will be zero. For example, the probability of a thing being red given that it is anything at all (including sets, real numbers, transfinite ordinals, etc.), is surely zero. Thus I think it is best if we just forget about absolute probabilities and rest content with conditional probabilities.

Let us look more closely at what it means to say that $\text{prob}_S(Gx/Fx) = r$. As a first approximation, we want to read this as saying that the proportion of physically possible $F$'s that are $G$'s is $r$. But what does *this* mean? To talk about physically possible $F$'s is, presumably, to talk about the physically possible ways of being an $F$. We can regard a way of being an $F$ as a maximal consistent set of predicates which contains $F$. Let us define:

$$(2.4)\qquad \Omega = \{A; (D)[D \text{ is a predicate} \supset (D \in A \vee \ulcorner \sim D \urcorner \in A)] \,\&\, \underset{p}{\Diamond}(\exists x)(D)(D \in A \supset Dx)\}.$$

Then the set of all physically possible ways of being an $F$ is:

$$(2.5)\qquad [F] = \{A; A \in \Omega \;\&\; F \in A\}.$$

Then it is natural to suppose that for each $F$ in $\Pi$, there is an additive measure function $\mu$ defined on (certain) subsets of $[F]$ such that $\text{prob}_S(Gx/Fx) = \mu([F] \cap [G])$.

The above approach seems to work as long as $\underset{p}{\Diamond}(\exists x)Fx$, but if $\sim\underset{p}{\Diamond}(\exists x)Fx$ it would lead to $\text{prob}_S(Gx/Fx)$ being undefined (because $[F]=\varnothing$). If $F$ is physically possible, then instead of talking about physically possible ways of being an $F$ (there aren't any), it seems natural to talk counterfactually about what *would* be physically possible ways of being an $F$ *if* $F$ were physically possible. This is on the right track, but is still not quite right. The difficulty is that there might be different ways of making $F$ physically possible. For example, if $F$ is a disjunction $\ulcorner F_1x \vee F_2x\urcorner$ where $\ulcorner F_1x\urcorner$ contradicts one basic law and $\ulcorner F_2x\urcorner$ contradicts another basic law, then we could make $F$ physically possible by rejecting either of these laws. But which of these two laws we reject may lead to quite different maximal consistent sets of predicates being physically possible, and there may be no such sets containing $F$ which would be physically possible regardless of which law we reject. Thus we do not want to talk about what maximal consistent sets of predicates *would* be physically possible ways of being an $F$ if $F$ were physically possible; instead we want to talk about what maximal consistent sets of predicates *might* be physically possible ways of being an $F$ if $F$ were physically possible. In other words, in computing $\text{prob}_S(Gx/Fx)$, we take into account all of the physically possible way of being an $F$ that would result from *each* of the ways of making $F$ physically possible. Thus we define:

(2.6) $\Omega_F = \{A; (D)[D \text{ is a predicate} \supset (D \in A \vee \ulcorner\sim D\urcorner \in A)] \;\&$
$\Diamond(\exists x)(D)(D \in A \supset Dx) \;\&$
$\sim[\underset{p}{\Diamond}(\exists x)Fx > \sim\underset{p}{\Diamond}(\exists x)(D)(D \in A \supset Dx)]\}.$

(2.7) $[G]_F = \{A; A \in \Omega_F \;\&\; G \in A\}.$

Then $\text{prob}_S(Gx/Fx)$ is supposed to be a measure of the proportion of things in $[F]_F$ that are also in $[G]_F$. This means in particular that for each $F$ in $\Pi$, there is an additive measure function $\mu_F$ such that $\mu_F([F]_F)=1$, and for each $G$ in $\Pi$ $\text{prob}_S(Gx/Fx)=\mu_F([G]_F \cap [F]_F)$. This works for all cases except that where $[F]_F=\varnothing$. It is easily demonstrated that $[F]_F=\varnothing$ iff $\sim\Diamond(\exists)Fx$, so in this case it seems best to simply stipulate that, since $(x)(Fx \rightarrow Gx)$, $\text{prob}_S(Fx \rightarrow Gx)=1$.

We cannot get by with a single measure function $\mu$ and define

$$\mu_F([G]_F \cap [F]_F) = \frac{\mu([G]_F \cap [F]_F)}{\mu([F]_F)}$$

as has been the traditional assumption when dealing with probability measure-theoretically. This is because such a function $\mu$ would have to give a finite value to every subset of $\Omega$ in its domain, and so in particular, $\mu([G \vee \sim G]_{\ulcorner G \vee \sim G \urcorner}) = \mu(\Omega)$ must be finite. But this will have the result that for many logically contingent $F$'s, $\mu([F]_F) = 0$, and hence the above definition will not work. Instead, we must have a whole class $\mathfrak{F}$ of measure functions. For each $X \subseteq \Omega$, if $X \neq \varnothing$ and $X$ lies in the domain of some member of $\mathfrak{F}$, then there should be a unique $\mu$ in $\mathfrak{F}$ such that $X$ is in the domain of $\mu$ and $\mu(X) \neq 0$. Let $\mu_X$ be this unique $\mu$. Then we define:

$$(2.8) \qquad \mu_F(X) = \frac{\mu_{[F]_F}(X)}{\mu_{[F]_F}([F]_F)}.^{1}$$

Finally then, we can define:

$$(2.9) \qquad \text{prob}_S(Gx/Fx) = \mu_F([G]_F \cap [F]_F).$$

This measure-theoretic characterization of $\text{prob}_S$ looks at least superficially like Carnap's 'logical' interpretation of probability (although Carnap was working with definite probabilities rather than indefinite probabilities). However, there is a very important difference here. This is that the class $\mathfrak{F}$ of measure functions is not *a priori* definable. This can be seen as follows. In general, there will be two distinct elements that contribute to the value of $\text{prob}_S(Gx/Fx)$. On the one hand, there are strong subjunctive generalizations ('deterministic laws'), and they contribute by determining what physically possible ways there are of being an $F$, and hence determine membership in $[G]_F$ and $[F]_F$. Different deterministic laws will leave the class $\mathfrak{F}$ of measures unchanged, and will affect probabilities only by altering what sets are being measured. But on the other hand, there may also be some probabilistic laws. These cannot alter membership in $[G]_F$ or $[F]_F$ because they do not render anything physically impossible – only more or less improbable. Consequently, probabilistic laws must alter probabilities by altering the measure functions themselves. They result in

the same physical possibility receiving different weights. For example, let us suppose there are no deterministic laws, so that everything that is logically possible is physically possible. Then different probabilistic laws will all operate on the same sets of possible ways of being an $F$ or a $G$, ($[G]_F = [G]$ and $[F]_F = [F]$), and hence can only result in different values for $\text{prob}_S(Gx/Fx)$ by assigning different measures to $[G]$ and $[F]$. This means that the contents of the set $\mathfrak{F}$ are contingent, depending upon what probabilistic laws there are, and hence cannot be defined *a priori.*

On the other hand, if we suppose that all laws are deterministic, that there are no probabilistic laws, then $\mathfrak{F}$ becomes fixed, and differences in probability can only be generated by differences in membership of $[G]_F$ and $[F]_F$. In this case, it is not totally implausible to suppose that $\mathfrak{F}$ really is a priori definable, although I haven't the faintest idea how one might set about trying to define it. Let us call this set of measure functions the set $\mathfrak{F}_L$ of *logical measure functions.* These measure functions are intimately connected with Carnap's 'logical conception of probability'. Many of the difficulties to which Carnap's approach led can be traced to his supposition that we could define our probabilities in terms of a single logical measure rather than employing a whole class of measure functions. For example, if $X$ is finite, then it seems reasonable that the logical measure $\mu_x$ should simply count the members of $X$. However, if $X$ is infinite, then we need a different measure function, and there is no obvious connection between this measure function and the measure functions for smaller sets. We need a whole hierarchy of logical measure functions, and they seem to be largely independent of one another. The problem of how to define the logical measure functions seems extremely difficult.

Unless we somehow know that there are no probabilistic laws, we cannot employ the logical measure functions directly in computing probabilities. Probabilistic laws force us to adopt a new set of measure functions – the 'empirical' measures. To this extent, the traditional 'logical conception of probability' is bankrupt. Nevertheless, when we turn to the discussion of probabilistic laws, we will find that the logical measure functions do play a central role even when our measure functions are not those in $\mathfrak{F}_L$.

Our measure-theoretic characterization of $\text{prob}_S$ immediately im-

plies that it satisfies all of the following conditions, which will henceforth be called 'the measure-theoretic axioms':

(2.10) $0 \leq \text{prob}_S(Gx/Fx) \leq 1$.

(2.11) $(x)(Fx \rightarrow Gx) \supset \text{prob}_S(Gx/Fx) = 1$.

(2.12) $(x)(Fx \leftrightarrow Gx) \supset \text{prob}_S(Hx/Fx) = \text{prob}_S(Hx/Gx)$.

(2.13) $\Diamond(\exists x)Fx \supset \text{prob}_S(\sim Gx/Fx) = 1 - \text{prob}_S(Gx/Fx)$.

(2.14) $(x)(Fx \rightarrow Gx) \supset \text{prob}_S(Fx/Hx) \leq \text{prob}_S(Gx/Hz)$.

It is generally supposed that conditional probability satisfies an additional condition, which can be formulated by any of the following axioms:

(2.15) If $(x)(Fx \rightarrow Gx)$ and $(x)(Gx \rightarrow Hx)$, and $\text{prob}_S(Gx/Hx) \neq 0$, then $\text{prob}_S(Fx/Gx) = \dfrac{\text{prob}_S(Fx/Hx)}{\text{prob}_S(Gx/Hx)}$.

(2.16) If $(x)(Gx \rightarrow \sim Hx)$, then
$\text{prob}_S(Fx/(Gx \vee Hx)) = \text{prob}_S(Fx/Gx) \cdot \text{prob}_S(Gx/(Gx \vee Hx)) + \text{prob}_S(Fx/Hx) \cdot \text{prob}_S(Hx/(Gx \vee Hx))$.

(2.17) $\text{prob}_S((Fx \ \& \ Gx)/Hx) = \text{prob}_S(Fx/Hx) \cdot \text{prob}_S(Gx/(Fx \ \& \ Hx))$.

These 'product axioms' are interderivable given the measure-theoretic axioms, and together they would give us all of the normal principles of the probability calculus. In particular, they imply the axioms of Popper (1955) and (1959). Unfortunately, the product axioms are all false. For example, consider 2.15. By our definitions, given the antecedent of 2.15, we have:

$$\text{prob}_S(Fx/Gx) = \frac{\mu_{[G]_G}([F]_G)}{\mu_{[G]_G}([G]_G)}$$

$$\frac{\text{prob}_S(Fx/Hx)}{\text{prob}_S(Gx/Hx)} = \frac{\mu_{[H]_H}([F]_H)}{\mu_{[H]_H}([G]_H)}$$

These are only the same if $[F]_G = [F]_H$ and $[G]_G = [G]_H$, i.e., if the same worlds might be physically possible if $\ulcorner(\exists x)Gx\urcorner$ were physically

possible and if $\ulcorner(\exists x)Hx\urcorner$ were physically possible. This is the same as requiring that $\Omega_G = \Omega_H$. This will be true if $\ulcorner(\exists x)Gx\urcorner$ and $\ulcorner(\exists x)Hx\urcorner$ are already physically possible, or more generally if

$$(\underset{p}{\Diamond}(\exists x)Gx > \underset{p}{\Diamond}(\exists x)Hx) \;\&\; (\underset{p}{\Diamond}(\exists x)Hx > \underset{p}{\Diamond}(\exists x)Gx).$$

with this qualification, 2.15 becomes true:

(2.15*) If $(x)(Fx \rightarrow Gx)$ and $(x)(Gx \rightarrow Hx)$ and $(\underset{p}{\Diamond}(\exists x)Gx > \underset{p}{\Diamond}(\exists x)Hx)$ and $(\underset{p}{\Diamond}(\exists x)Hx > \underset{p}{\Diamond}(\exists x)Gx)$ and $\text{prob}_S(Gx/Hx) \neq 0$, then

$$\text{prob}_S(Fx/Gx) = \frac{\text{prob}_S(Fx/Hx)}{\text{prob}_S(Gx/Hx)}.$$

Similar qualifications lead to true versions of 2.16 and 2.17. In particular, the product axioms are all true if we suppose that $\ulcorner(\exists x)Fx\urcorner$, $\ulcorner(\exists x)Gx\urcorner$, and $\ulcorner(\exists x)Hx\urcorner$ are all physically possible.

It has been remarked that strong subjunctive generalizations can be regarded as a kind of limiting case of strong indefinite probabilities. Part of the meaning of this claim is provided by principle 2.1, according to which $(Fx \Rrightarrow Gx) \rightarrow \text{prob}_S(Gx/Fx) = 1$. However, as we remarked earlier, this entailment does not go the other way. The precise sense in which generalizations are limiting cases of probabilities is provided by the following theorem:

(2.18) $(Fx \Rrightarrow Gx) \equiv \{\underset{p}{\Diamond}(\exists x)Fx > (H)[\underset{p}{\Diamond}(\exists x)(Fx \;\&\; Hx) \supset \text{prob}_S(Gx/(Fx \;\&\; Hx)) = 1]\}$[2]

*Proof:* From left to right follows immediately from theorem 5.20 of Chapter VI according to which if $\ulcorner(Fx \Rrightarrow Gx)\urcorner$ is true then

$$\underset{p}{\Diamond}(\exists x)Fx > [\underset{p}{\Diamond}(\exists x)(Fx \;\&\; Hx) \supset [(Fx \;\&\; Hx) \Rrightarrow Gx]]$$

together with principle 2.1. Conversely, suppose the right side holds. Suppose

(2.19) $(H)[\underset{p}{\Diamond}(\exists x)(Fx \;\&\; Hx) \supset \text{prob}_S(Gx/((Fx \;\&\; Hx)) = 1].$

Suppose $\ulcorner \underset{p}{\Diamond}(\exists x)(Fx \ \& \sim Gx)\urcorner$. $\text{prob}_S(Gx/(Fx \ \& \sim Gx)) = 0 \neq 1$, which contradicts 2.19. Thus 2.19 entails $\ulcorner \underset{p}{\Box}(x)(Fx \supset Gx)\urcorner$, and hence the right side of 2.18 entails $\ulcorner[\underset{p}{\Diamond}(\exists x)Fx > \underset{p}{\Box}(x)(Fx \supset Gx)]\urcorner$. By theorem 5.21 of Chapter VI, this is equivalent to $\ulcorner(Fx \Rrightarrow Gx)\urcorner$.

The above remarks have been intended to elucidate the intuitive meaning of strong indefinite probability, and to elicit the logical properties of $\text{prob}_S$, but they cannot be regarded as constituting an analysis. We gave a measure-theoretic characterization of $\text{prob}_S$, but that is not an analysis because the measure $\mu_F$ is not constructed *a priori*. In fact, the only obvious way to define $\mu_F$ is in terms of $\text{prob}_S$ itself.

If we cannot analyze $\text{prob}_S$ measure-theoretically, how are we to analyze this concept? This is not a problem which I am now prepared to solve, although I can make some remarks which may, hopefully, point in the direction of a solution. The traditional attempts to define indefinite probability in terms of the limit of relative frequencies fail because of the subjunctive nature of non-material indefinite probabilities. Strong indefinite probabilities are intimately connected with strong subjunctive generalizations, the latter being a limiting case of the former. When it came time to analyze subjunctive generalizations, it was almost obviously hopeless to try to give a reductive analyses which would state their truth conditions in terms of other simpler concepts. Instead we sought an analysis in terms of their justification conditions. I argued at great length in Pollock (1974) that those intransigent concepts which resist truth-condition analysis can instead be analyzed by giving an account of how we operate with them, or more precisely, by giving an account of how we become justified in holding beliefs involving those concepts. In the case of subjunctive generalizations this came down to giving an account, first, of how basic subjunctive generalizations are confirmed directly by induction, and second, of how other subjunctive generalizations are derived from the basic ones. I believe that the same sort of account will work for strong indefinite probability. The justification conditions for this concept are of basically the same structure as those for its limiting case, the strong subjunctive generalization. There are two kinds of strong indefinite

probability statements. On the one hand there are those *basic* indefinite probability statements that can be confirmed directly by induction, and on the other hand there are non-basic indefinite probability statements that are derived from the basic ones.

Let us begin by looking at the basic indefinite probability statements. Herein lies the initial plausibility of the frequentist position. It is, I think, inescapable that subjunctive indefinite probabilities cannot be defined in terms of relative frequencies, but relative frequencies must be connected with probabilities somehow. And I think it is rather obvious what the connection is: observed relative frequencies provide the inductive grounds upon which we base our estimates of probabilities. As in the case of subjunctive generalizations, basic indefinite probability statements must involve pairs of *projectible* predicates. And, as is always the case with inductive grounds, the confirmation is defeasible. However, the defeaters involved in the confirmation of indefinite probability statements seem to be considerably more complicated than those involved in the confirmation of subjunctive generalizations. I will not attempt to spell them out here.

Turning next to the question of how we obtain non-projectible indefinite probability statements from the projectible ones, we come to an extraordinarily difficult problem. Sometimes we can simply use the laws of the probability calculus and calculate the probabilities for non-projectible predicates directly, but this is only a very small part of the story. It will become apparent in section three that concealed in this question are many of the most important and difficult problems of probability theory.

Although I am unable to give an account of the kind required, I do want to claim that when such an account is given, we will then have given an analysis of the meaning of strong indefinite probability statements. This will be no less an analysis than a truth-condition analysis would be. However, I am confident that a truth-condition analysis is impossible.

### 2.4. *Weak Indefinite Probability Statements*

Between strong and weak indefinite probability statements, the weak ones are commonly of more interest to us. We are generally more

interested in how likely an $F$ is to be a $G$ *given the way the world actually is* than we are in how likely an $F$ is to be a $G$ in all physically possible worlds. We can give a measure-theoretic characterization of $\text{prob}_W(Gx/Fx)$ that is completely analogous to our measure-theoretic characterization of $\text{prob}_S(Gx/Fx)$. Just as $\text{prob}_S(Gx/Fx)$ is supposed to be the proportion of physically possible ways of being an $F$ that are also ways of being a $G$, so $\text{prob}_W(Gx/Fx)$ is supposed to be the proportion of actually possible ways of being an $F$ that are also ways of being a $G$. Thus we can define:

(2.20) $\Omega^*_F = \{A;\ (D)[D \text{ is a predicate} \supset (D \in A \vee \ulcorner \sim D \urcorner \in A)]\ \&\ \Diamond(\exists x)(D)(D \bullet A \supset Dx)\ \&\ \sim[\underset{a}{\Diamond}(\exists x)Fx > \sim\underset{a}{\Diamond}(\exists x)(D)(D \in A \supset Dx)]\}$.

(2.21) $[G]^*_F = \{A;\ A \in \Omega^*_F\ \&\ G \in A\}$

Then there is a set $\mathfrak{F}^*$ of measures such that for each $X$, if $X \neq \varnothing$ and $X$ is in the domain of some member of $\mathfrak{F}^*$, then there is a unique $\mu$ in $\mathfrak{F}^*$ such that $X$ is in the domain of $\mu$ and $\mu(X) \neq 0$. Let $\mu^*_X$ be this unique $\mu$. In Section 3.2 below, I will argue that in general, $\mu^*_X = \mu_X$, because these measures reflect the same probabilistic laws. It will be maintained that there are not two kinds of probabilistic laws, one involved in strong probabilities and the other in weak probabilities. On this assumption, we have:

(2.22) $\mu^*_F(X) = \dfrac{\mu_{[F]^*_F}(X)}{\mu_{[F]^*_F}([F]^*_F)}$.

(2.23) $\text{prob}_W(Gx/Fx) = \mu^*_F([G]^*_F \cap [F]^*_F)$.

As before, this gives us most of the normal principles of the probability calculus for $\text{prob}_W$. The measure-theoretic axioms are true, and restrictions analogous to those involved in $\text{prob}_S$ give us qualified versions of the product axioms.

There is also another way of characterizing $\text{prob}_W$ which is in many ways more interesting. Weak indefinite probability statements and weak subjunctive generalizations are very much alike, as indeed they should be given that the latter is really a kind of limiting case of the former. In each case they are made true by some physically contingent facts about the world. We found that weak subjunctive generalizations

can be defined in terms of strong subjunctive generalizations:

$$(2.24)\quad (Fx \Rightarrow Gx) \equiv (\exists P)(\exists A)\{(x)(x \in A \equiv x = x)\ \&\ (Fx\ \&\ P. \Rightarrow Gx)\ \&\ PE(\exists x)(Fx\ \&\ x \notin A)\}.$$

Can we, in some analogous way, define weak indefinite probability statements in terms of strong indefinite probability statements? It seems we can. Let $P$ be that physically contingent proposition the truth of which is responsible for the weak probability being what it is. Then, analogous to the condition in 2.24 that $\ulcorner[(Fx\ \&\ P) \Rightarrow Gx]\urcorner$ be true, we have the condition that $\text{prob}_W(Gx/Fx) = \text{prob}_S(Gx/(Fx\ \&\ P))$. We also need a restriction on $P$ analogous to that contained in 2.24. To see what this restriction should be, consider the case of a coin whose physical characteristics are such as to make it a fair coin. Then we want to say that the probability of a flip of this coin at this time yielding a head is 1/2. Let $P$ be the statement describing the physical characteristics which make the coin a fair coin. Clearly it is required of $P$ that it would be true even if the coin were flipped, i.e., $\ulcorner PE(\exists x)Fx\urcorner$ is true. However, the condition $\ulcorner PE(\exists x)Fx\urcorner$ is clearly not sufficient by itself. If the coin is in fact being flipped at this time, then the condition $\ulcorner PE(\exists x)Fx\urcorner$ is automatically true because both $P$ and $\ulcorner(\exists x)Fx\urcorner$ are true. We must require of $P$ that it would still be true even if a different flip of the coin occurred. But this just means that $\ulcorner PE(\exists x)(Fx\ \&\ x \notin A)\urcorner$ should be true, where $A$ is the set of actually existing things. Thus our restriction becomes the same as in 2.24. This suggests that we should have the following:

$$(2.25)\quad \text{prob}_W(Gx/Fx) = r \equiv (\exists P)(\exists A)[\text{prob}_S(Gx/(Fx\ \&\ P)) = r\ \&\ (x)(x \in A \equiv x = x)\ \&\ PE(\exists x)(Fx\ \&\ x \notin A)].$$

Principle 2.25 is precisely analogous to 2.24. However, because we are now dealing with probabilities rather than strict generalizations, a further difficulty remains for 2.25 that did not occur in connection with 2.24. This is that the function $\text{prob}_W$ is not well defined because there could be two propositions $P$ and $P^*$ satisfying the constraints of 2.25 but such that $\text{prob}_S(Gx/(Fx\ \&\ P)) \neq \text{prob}_S(Gx/(Fx\ \&\ P^*))$. In such a case we would say that at least one of the two propositions $P$ and $P^*$ has not taken into account all of the relevant information. Intuitively, we want to do something like taking the conjunction of all of the

propositions $P$ satisfying the constraints of 2.25, and then look at the strong probability of an $F$ being a $G$ given that that conjunction is true. However, taking such an infinite conjunction is only a heuristic approximation to the right idea. One might question whether such infinite conjunctions really make sense. It seems that what we really want to require of $P$ is that it be strong enough to take into account all of the relevant information, in the sense that if there is another proposition $P^*$ also satisfying the constraints of 2.25 but yielding a different probability, then $P \rightarrow P^*$ and hence takes account of its information. This condition is still a bit too strong, because given a $P^*$ of this sort and an arbitrary irrelevant proposition $Q$ such that $\ulcorner QE(\exists x)(Fx \ \& \ x \not\in A)\urcorner$ is true, we would have $\text{prob}_S(Gx/(Fx \ \& \ P^*)) = \text{prob}_S(Gx/(Fx \ \& \ P^* \ \& \ Q))$, and hence the above condition would require that $P \rightarrow Q$. This would lead to the result that $P$ would have to entail every irrelevant proposition $Q$ satisfying the constraint that $\ulcorner QE(\exists x)(Fx \ \& \ x \not\in A)\urcorner$ is true. This is much too strong a requirement. We can avoid requiring $P$ to entail such irrelevant propositions by only requiring $P$ to entail $P^*$ if there is no $R$ such that $P^*$ entails $R$ but $R$ does not entail $P^*$, and $\ulcorner RE(\exists x)(Fx \ \& \ x \not\in A)\urcorner$ is true and $\text{prob}_S(Gx/(Fx \ \& \ R)) = \text{prob}_S(Gx/(Fx \ \& \ P^*))$. If there is such an $R$, then $P^*$ is an irrelevant strengthening of it. So our analysis becomes:

(2.26) $\text{prob}_W(Gx/Fx) = r \equiv (\exists P)(\exists A)[(x)(x \in A \equiv x = x)$
$\& \ \text{prob}_S(Gx/(Fx \ \& \ P)) = r \ \& \ (P^*)\{[P^*E(\exists x)(Fx \ \& \ x \not\in A)$
$\& \ \text{prob}_S(Gx/(Fx \ \& \ P^*)) \neq \text{prob}_S(Gx/(Fx \ \& \ P))$
$\& \ PE(\exists x)(Fx \ \& \ x \not\in A) \ \& \sim(\exists R)((P^* \rightarrow R)$
$\& \ (R \not\rightarrow P^*) \ \& \ \text{prob}_S(Gx/(Fx \ \& \ R))$
$= \text{prob}_S(Gx/Fx \ \& \ P^*)))\} \supset (P \rightarrow P^*)]$.

But there remains one final difficulty. There seems to be no logical guarantee that there is such a maximally strong $P$. It seems logically conceivable for there to be an infinite progression of stronger and stronger $P$'s without limit each giving a different probability.[3] In such a case, it would be unreasonable to identify $\text{prob}_W(Gx/Fx)$ with the probability given by any of the $P$'s. In such an eventuality, there would be no $P$ by virtue of which there would be a weak indefinite probability different from the strong indefinite probability, and so we should

have $\text{prob}_W(Gx/Fx) = \text{prob}_S(Gx/Fx)$. Thus our final analysis becomes:

(2.27) $\text{prob}_W(Gx/Fx) = r$ iff either

(i) $(\exists P)(\exists A)[(x)(x \bullet A \equiv x = x)$
$\& \text{prob}_S(Gx/(Fx \& P)) = r \& PE(\exists x)(Fx \& x \in A)$
$\& (P^*)\{[P^* E(\exists x)(Fx \& x \notin A)$
$\& \text{prob}_S(Gx/(Fx \& P^*)) \neq \text{prob}_S(Gx/(Fx \& P))$
$\& \sim(\exists R)((P^* \rightarrow R)$
$\& (R \not\rightarrow P^*) \text{prob}_S(Gx/(Fx \& R))$
$= \text{prob}_S(Gx/(Fx \& P^*))\} \supset (P \rightarrow P^*)]$; or

(ii) $\sim(\exists P)(\exists A)[(x)(x \in A \equiv x = x) \& PE(\exists x)(Fx \& x \notin A)$
$\& (P^*)\{[P^* E(\exists x)(Fx \& x \notin A)$
$\& \text{prob}_S(Gx/(Fx \& P^*)) \neq \text{prob}_S(Gx/(Fx \& P))$
$\& \sim(\exists R)((P^* \rightarrow R) \& (R \not\rightarrow P^*)$
$\& \text{prob}_S(Gx/(Fx \& R)) = \text{prob}_S(Gx/(Fx \& P^*))\}$
$\supset (P \rightarrow P^*)] \& \text{prob}_S(Gx/Fx) = r.$

## 3. The redefinition of **M**

One of the major reasons for discussing subjunctive indefinite probabilities at such length is that it seems they should be relevant to the definition of the relation **M** which is involved in the analysis of subjunctive conditionals. We have entertained the hypothesis that there are fundamental laws of nature which are essentially probabilistic. For example, we may have a law which tells us that if an electron is in state $S_1$ at time $t$, then the probability is 1/3 that it is in state $S_2$ at time $t + \Delta t$. Given such a probabilistic law, it would seem to follow that if an electron is in state $S_1$ at time $t$, then it might be in state $S_2$ at time $t + \Delta t$. Non-zero probabilities arising from fundamental probabilistic laws create 'might be's'. This is a genuinely new source of 'might be's'. Without probabilistic laws, the only way a 'might be' can arise is when $P$ counter-implicates a conjunction of true propositions without counter-implicating either conjunct.

Let us see how we can make all of this precise. We want to say, roughly, that if $Q$ is false but $\text{prob}(Q/P) \neq 0$, and this probability arises from fundamental probabilistic laws, then $QMP$, and hence there must be a world $\beta$ such that $\beta \mathbf{M} P$ and $Q$ is true in $\beta$. The first thing to notice here is that the probability $\text{prob}(Q/P)$ to which we are appealing is not an indefinite probability. It is the probability of a proposition $Q$ given a proposition $P$. This is a definite probability, and is something new which we have not yet defined. We must begin by constructing a definition of this probability function.

### 3.1. *Strong and Weak Definite Probability*

The definite probabilities in which we are interested are probabilities of propositions and arise directly out of the indefinite probabilities we have already discussed. These definite probabilities are based solely and exclusively on the indefinite probabilities. We obtain two kinds of definite probabilities depending upon whether they are based on strong indefinite probabilities or weak indefinite probabilities. We will call these 'strong-' and 'weak definite probability' respectively. We will symbolize them using the same symbols we used for the indefinite probabilities: $\ulcorner \text{prob}_S(Q/P) \urcorner$ and $\ulcorner \text{prob}_W(Q/P) \urcorner$. However, there is no danger of confusing the definite probabilities with the indefinite probabilities, because the arguments for the former must be propositions whereas the arguments for the latter must be predicates.

We want $\ulcorner \text{prob}_S(Q/P) \urcorner$ to be a definite probability based solely on what strong indefinite probabilities there are. How is this to be done? In a purely formal way, I think that this is quite simple. First define:

(3.1) $T(P)$ = the set of all possible worlds in which the proposition $P$ is true.

Then as a first approximation, we can define quite simply:

(3.2) $\text{prob}_S(Q/P) = \text{prob}_S(x \in T(Q)/x \in T(P))$.

In other words, the strong probability of $Q$ given $P$ is simply the strong probability of a world in which $P$ is true being a world in which $Q$ is true. Thus we reduce the definite probabilities to the indefinite probabilities in a very simple way.

It may seem that we are somehow cheating in this definition. I think we are cheating, but the cheating does not occur at this point but earlier on when we eschewed the attempt to give an account of how non-projectible indefinite probabilities are computed on the basis of projectible ones. Just what is concealed in that problem now becomes apparent. Hidden there is the whole problem of how one computes definite probabilities on the basis of inductively confirmed indefinite probabilities. The probability $\text{prob}_S(x \bullet T(Q)/x \in T(P))$ is itself a perfectly reasonable indefinite probability. However, it is clearly not an indefinite probability whose value is subject to direct inductive confirmation. Thus its value must be indirectly determined by those basic strong indefinite probabilities which can be confirmed inductively. I will not try to say how this is done, because I do not know how it is done. This involves all the traditional problems of randomness, etc. Without a solution to this problem, we cannot claim to have given an analysis of probability. However, that is not my present purpose. My objective in this chapter is not the analysis of probability per se, but rather the exhibition of the intimate connections between probability and subjunctive conditionals, and we can accomplish the latter without a complete analysis of probability.

Returning now to 3.2, we can see that it does not provide quite the definition we want. Intuitively, we want $\text{prob}_S(Q/P)$ to be the proportion of physically possible worlds making $P$ true which also make $Q$ true, but our definition fails to give us this result. This can be seen by looking at our measure-theoretic analysis of $\text{prob}_S$. $[\ulcorner x \bullet T(Q) \urcorner]_{\ulcorner x \in T(P) \urcorner}$ consists of certain maximal sets of predicates including $\ulcorner x \in T(Q) \urcorner$ which are such that it is logically possible for something to satisfy them all simultaneously. The only things that can satisfy $\ulcorner x \in T(Q) \urcorner$ are possible worlds, so these are maximal consistent set of predicates of possible worlds. A predicate of possible worlds picks out a set of possible worlds. It is frequently maintained that any set of possible worlds determines a proposition. This might be denied on the grounds that propositions are 'intensional', and hence are only picked out by 'definable' sets of possible worlds. However, it is at least clear that any set of possible worlds picked out by a predicate does determine a proposition. Conversely, any proposition $R$ determines a predicate of possible worlds, viz., $\ulcorner x \in T(R) \urcorner$. Consequently, the members of $[\ulcorner x \in$

$T(Q)^{\urcorner}]_{x \in T(P)^{\urcorner}}$ correspond one-one to maximal consistent sets of propositions, which in turn correspond to possible worlds. Thus we can regard $[^{\ulcorner}x \in T(Q)^{\urcorner}]_{x \in T(P)^{\urcorner}}$ as picking out a set of possible worlds in which $Q$ is true. By definition 2.25, a member $A$ of this set of possible worlds must satisfy the additional constraint that

$$\sim[\underset{p}{\Diamond}(\exists x)x \in T(P) > \sim\underset{p}{\Diamond}(\exists x)(R)(^{\ulcorner}x \in T(R)^{\urcorner} \in A \supset x \in T(R)].$$

We want this constraint to say that if $P$ were physically possible, then $A$ would be one of those worlds that might be physically possible. At first it looks like it does say this, but it doesn't. The difficulty is that $^{\ulcorner}(\exists x)x \bullet T(P)^{\urcorner}$ simply says that there is a possible world in which $P$ is true, which is to say $^{\ulcorner}\Diamond P^{\urcorner}$. Similarly, $^{\ulcorner}(\exists x)(R)(^{\ulcorner}x \in T(R)^{\urcorner} \in A \supset x \in T(R))^{\urcorner}$ say that there is a world making all of the propositions corresponding to $A$ true at the same time, which is just to say that $A$ determines a possible world. But we already know this. Consequently, our constraint is vacuous. But this means that $[^{\ulcorner}x \in T(Q)^{\urcorner}]_{x \in T(P)^{\urcorner}}$ picks out all logically possible worlds in which $Q$ is true, and not just the physically possible ones. Consequently, $\text{prob}_S(^{\ulcorner}x \bullet T(Q)^{\urcorner}/^{\ulcorner}x \in T(P)^{\urcorner})$ is the proportion of all *logically* possible worlds making $P$ true which also make $Q$ true, rather than being the proportion of all *physically* possible worlds making $P$ true which also make $Q$ true.

Formally, our difficulty arises from the fact that $^{\ulcorner}\underset{p}{\Diamond}(\exists x)x \in T(Q)^{\urcorner}$ is equivalent to $^{\ulcorner}\underset{p}{\Diamond}\Diamond Q^{\urcorner}$, and hence to $^{\ulcorner}\Diamond Q^{\urcorner}$, whereas we would like it to be equivalent to $^{\ulcorner}\underset{p}{\Diamond}Q^{\urcorner}$. The source of our difficulty lies in a strange equivocation of which philosophers are frequently guilty in connection with possible worlds. First consider possible men. Corresponding to each man is his 'diagram' – the set of all the predicates he satisfies. Talk about possible men is talk about possible ways men could be, which is to talk about all those diagrams which logically could be the diagrams of men. No one would confuse a man with his diagram. Non-actual men do not exist, but their diagrams do. Hence there is a clear distinction between saying that it is physically possible for there to be a man of a certain description and saying that it is physically possible for there to be a man-diagram containing that description.

The latter is equivalent to saying that it is physically possible for it to be logically possible for there to be a man of a certain description (i.e., it is logically possible for there to be a man of that description) because (unlike men) man-diagrams exist just so long as they are logically consistent.

Philosophers tend not to make the analogous distinction between worlds and world-diagrams. Instead, they generally identify a world with its world diagram. This is the customary procedure, and it has been our procedure throughout this book. But now that customary procedure has landed us in difficulty. We must distinguish between saying ⌜It is physically possible for there to be a world in which $Q$ is true⌝ and ⌜It is physically possible for there to be a world-diagram containing $Q$⌝. As we have seen, the latter is equivalent to ⌜It is logically possible that $Q$⌝, and given our identification of worlds with their diagrams, it is this rather than the former which is captured by $\ulcorner \underset{p}{\Diamond}(\exists x) x \in T(Q) \urcorner$.

It is now clear how to solve our problem. We must replace the predicate $\ulcorner x \bullet T(Q) \urcorner$ by another predicate $\ulcorner Q(x) \urcorner$ such that $\ulcorner (\exists x) Q(x) \urcorner$ says that $Q$ is true (in the actual world) rather than that $Q$ is true in some world or other. We can accomplish this without changing our formal definition of 'possible world' if we introduce the predicate ⌜$x$ is actual⌝ to mean ⌜$x$ is a (the) actual world⌝. Then we can define:

$$(3.3) \qquad \text{prob}_S(Q/P) = \text{prob}_S(x \in T(Q)/(x \in T(P) \ \&\ x \text{ is actual})).$$

This has the desired result that $\text{prob}_S(Q/P)$ is the proportion of physically possible worlds making $P$ true which make $Q$ true.[4] Analogously, we can define:

$$(3.4) \qquad \text{prob}_W(Q/P) = \text{prob}_W(x \in T(Q)/(x \in T(P) \ \&\ x \text{ is actual})).$$

This gives us the proportion of actually possible worlds making $P$ true which make $Q$ true.

## 3.2. *Probabilistic Laws*

I have talked glibly about probabilistic laws as being probabilistic analogues of deterministic laws (subjunctive generalizations). But what

does this mean? We might naturally suppose that any strong indefinite probability statement expresses a probabilistic law, but it takes little reflection to see that such a view is mistaken. As we have seen, there are two contributing factors involved in generating strong probabilities – deterministic laws, which alter the contents of the sets $[G]_F$ by determining what worlds are physically possible, and probabilistic laws which dictate different weightings to the physically possible worlds. Even if there were no probabilistic laws (so our measures would be $\mathfrak{F}_L$, the logical measure functions), there would still be strong probabilities. Thus strong probability statements do not automatically express probabilistic laws.

To say that there is a probabilistic law operating in a certain case is just to say that our measure functions are not those in $\mathfrak{F}_L$. To make this precise, let us define a kind of hybrid probability function which uses the logical measures on the physically possible worlds. For each $X$, if $X \neq \varnothing$ and $X$ is in the domain of some member of $\mathfrak{F}_L$, let $\mu_{L,X}$ be the unique $\mu$ in $\mathfrak{F}_L$ such that $X$ is in the domain of $\mu$ and $\mu(X) \neq \varnothing$. Then we define:

$$(3.5) \qquad \mathrm{prob}_{L,S}(Gx/Fx) = \frac{\mu_{L,[F]_F}([G]_F \cap [F]_F)}{\mu_{L,[F]_F}([F])_F)}.$$

Analogously,

$$(3.6) \qquad \mathrm{prob}_{L,W}(Gx/Fx) = \frac{\mu_{L,[F]_F{}^*}([G]_F^* \cap [F]_F^*)}{\mu_{L,[F]_F{}^*}[F]_F^*)}.$$

Then to say that there is a probabilistic law operating is just to say that $\mathrm{prob}_S(Gx/Fx) \neq \mathrm{prob}_{L,S}(Gx/Fx)$. In such a case, let us say that $\ulcorner Fx \urcorner$ is *strongly relevant* to $\ulcorner Gx \urcorner$. Analogously, we can define weak relevance, and positive and negative relevance:

(3.7) $\ulcorner Fx \urcorner$ is strongly relevant to $\ulcorner Gx \urcorner$ iff
$\mathrm{prob}_S(Gx/Fx) \neq \mathrm{prob}_{L,S}(Gx/Fx)$.

(3.8) $\ulcorner Fx \urcorner$ is weakly relevant to $\ulcorner Gx \urcorner$ iff
$\mathrm{prob}_W(Gx/Fx) \neq \mathrm{prob}_{L,W}(Gx/Fx)$.

(3.9) $\ulcorner Fx \urcorner$ is strongly positively relevant to $\ulcorner Gx \urcorner$ iff
$\mathrm{prob}_{L,S}(Gx/Fx) < \mathrm{prob}_S(Gx/Fx)$.

(3.10) $\ulcorner Fx \urcorner$ is weakly positively relevant to $\ulcorner Gx \urcorner$ iff
$\text{prob}_{L,W}(Gx/Fx) < \text{prob}_{W}(Gx/Fx)$.

(3.11) $\ulcorner Fx \urcorner$ is strongly negatively relevant to $\ulcorner Gx \urcorner$ iff
$\text{prob}_{S}(Gx/Fx) < \text{prob}_{L,S}(Gx/Fx)$.

(3.12) $\ulcorner Fx \urcorner$ is weakly negatively relevant to $\ulcorner Gx \urcorner$ iff
$\text{prob}_{W}(Gx/Fx) < \text{prob}_{L,W}(Gx/Fx)$.

We can extend these definitions to strong and weak definite probabilities in the obvious way. Then to say that there is a probabilistic law operating in $\text{prob}_S(Gx/Fx)$ is just to say that $\ulcorner Fx \urcorner$ is strongly relevant to $\ulcorner Gx \urcorner$. Similarly, a probabilistic law is operating in $\text{prob}_W(Gx/Fx)$ iff $\ulcorner Fx \urcorner$ is weakly relevant to $\ulcorner Gx \urcorner$.

We know what it is for there to be a probabilistic law operating in a particular case, but what is the nature of the beast itself? What is a probabilistic law? What a probabilistic law does is shift our measures from the logical measures $\mathfrak{F}_L$ to a new set of empirical measures $\mathfrak{F}$, so it seems reasonable to simply identify the probabilistic laws with the set of empirical measures. That is, a probabilistic law is any set of probability measures other than $\mathfrak{F}_L$. This is not entirely in accord with our ordinary way of thinking of probabilistic laws according to which, e.g., the quantum mechanical equations expressing probability distributions for various quantities represent probabilistic laws. However, this can be reconciled with the view that a probabilistic law is a set of empirical probability measures. In order for the quantum mechanical equations to represent probabilistic laws, the distributions they dictate must differ from the 'chance' distributions generated by the logical measure functions. Insofar as they do differ, in order for the quantum mechanical equations to be true the set of empirical measures must differ from $\mathfrak{F}_L$. The quantum mechanical laws do not completely determine the set $\mathfrak{F}$ of empirical measures, but they do partially locate $\mathfrak{F}$ as being a member of that class of sets of measures which would yield the probability distributions expressed by the quantum mechanical laws. It seems reasonable to call such a partial characterization of $\mathfrak{F}$ a 'probabilistic law'. A complete specification of $\mathfrak{F}$ is just a maximally strong probabilistic law. The weaker probabilistic laws are equivalent

to the indefinite probability statements which generate them, so we can regard the latter as probabilistic laws too if we wish.

We have two kinds of subjunctive probability – strong and weak. Are there correspondingly two kinds of probabilistic laws? The answer to this seems to be, 'No'. A probabilistic law simply selects a set of empirical measure functions. There is nothing about such a set of measure functions to relate it to one kind of probability rather than the other. The difference between strong and weak probability would seem to be, not in the measure functions employed, but in the sets measured. Strong probability results from measuring sets of physically possible combinations. Weak probability only diverges from strong probability by virtue of measuring smaller sets, i.e., by measuring sets of actually possible combinations.

### 3.3. *The Analysis of* **M**

In our original analysis of **M**, which overlooked probabilities, the only way $\ulcorner QMP \urcorner$ could be true, where $Q$ is false in the actual world, is for there to be some false proposition $R$ such that $P \Rightarrow (Q \vee R)$, but $P \not\Rightarrow R$. But if the only true subjunctive generalizations were entailments, other putative subjunctive generalizations being replaced by subjunctive probability statements, this would not in fact diminish the set of 'might be' statements. On the contrary, whenever $P$ is positively relevant to $Q$, we would conclude that $QMP$. Thus it seems that there is a second way that 'might be' statements can arise, viz., from the positive relevance of $P$ to $Q$.

However, we have two kinds of subjunctive probabilities – strong and weak. Which kind is involved in generating 'might be's'? It is not difficult to see that it is the weak probabilities that are involved, the reason being that they take into account more about what is actually the case in the world. For example, consider Chisholm's bottle of rat poison again, but suppose now that all laws governing the effect of poison on the human system are probabilistic rather than deterministic. It is clear that the $\text{prob}_{L,S}(Sx/Dx)$ of the survival $Sx$ of a person who drank $Dx$ from the bottle is non-zero – there are numerous physically possible circumstances under which a person could do so and survive, e.g., by washing the bottle out first and then filling it with clean water. Let us suppose further that due to some strange psychological and

genetic features of human beings, people who drink from bottles are apt to live longer than people who do not drink from bottles. Then $\text{prob}_S(Sx/Dx) > \text{prob}_{L,S}(Sx/Dx)$, i.e., drinking from the bottle is (slightly) strongly positively relevant to surviving. But we can suppose that because the bottle is filled with very dilute poison, this is just enough to overcome that strong positive relevance that drinking from the bottle has to survival, with the result that drinking from the bottle is not weakly relevant to surviving. Under these circumstances, if we consider a person who died of old age, we would not say that he might have lived longer had he drunk from the bottle. If it weren't that the bottle contains the poison, the strong probability and the weak probability would coincide, there being nothing by virtue of which the weak probability would differ from the strong probability, and hence drinking from the bottle would be weakly positively relevant to survival. In that case we would say that the person might have lived longer had he drunk from the bottle; but *because* the bottle does contain the poison, it is not true that the person might have lived longer if he had drunk from it. Thus the same circumstances (the bottle's containing poison) which make the weak probability diverge from the strong probability, and hence prevent drinking from the bottle from being weakly positively relevant to survival, also determine that $\sim(QMP)$. This indicates that it is weak subjunctive probability and weak positive relevance that are involved in the analysis of **M** and allow additional 'might be' statements to be true.

How, precisely, is weak positive relevance involved in the analysis of **M**? We will start from analysis 2.15 of Chapter VI and see how it must be modified. The first thing we must do is modify the definition of a possible world so that it contains information on weak positive relevance. Rather than taking a possible world to be a quadruple $\langle S, \eta, N, W\rangle$ as in Chapter VI we will now take it to be a quintuple $\langle S, \eta, N, W, P\rangle$ where $P$ is the set of all ordered pairs $\langle \varphi, \psi\rangle$ such that $\varphi$ is weakly positively relevant to $\psi$.

Let us begin with the simplest case – that in which there are no deterministic laws, i.e., $N_\alpha = W_\alpha = \varnothing$, and $\varphi$ is indicative. In this case, definition 2.15 requires that if $\beta \mathbf{M}_\alpha\ \varphi$ then:

for $t \in Rl$, $S_\alpha(t) \triangle S_\beta(t)$ is a minimal $\varphi$-change to $T_\alpha$ at $t$.

In this case, noting that we have no contingent subjunctive generalizations with which to contend, it should be that case that if $\langle \varphi, \psi \rangle \in P_\alpha$, then $\psi M \varphi$, and hence there is some $\beta$ such that $\beta \mathbf{M}_\alpha \varphi$ and $\psi$ is true in $\beta$. We must liberalize the definition of $\mathbf{M}$ to allow such a $\beta$. This can be done very simply by disjoining the following to the above requirement:

(3.13) There is a $\psi$ such that $\alpha \notin T(\psi)$ and $\langle \varphi, \psi \rangle \in P_\alpha$ and for $t \in Rl$, $S_\alpha(t) \Delta S_\beta(t)$ is a minimal $\ulcorner \varphi \ \& \ \psi \urcorner$-change to $T_\alpha$ at $t$.

Next, what happens when $\alpha$ contains true contingent subjunctive generalizations? Let us still suppose that $\varphi$ is indicative, and suppose that $\varphi$ is not counter-legal, i.e., $\varphi$ is consistent with $W_\alpha$. What happens now is that having $\langle \varphi, \psi \rangle \in P_\alpha$ no longer guarantees that $\psi M \varphi$. This is because we can now talk about the historical antecedents of a proposition. For example, suppose that getting plenty of vitamin C in your diet is positively relevant to living a long life, and consider Jones who dies in an auto accident. Although his getting plenty of vitamin C is still positively relevant to his living a long life, we would not take this as implying that had he gotten plenty of vitamin C, he might have lived longer than he did. This is because although getting plenty of vitamin C is positively relevant to his not dying when he did, i.e., it is negatively relevant to his dying when he did, it is not negatively relevant to his dying in an auto accident. In general, given a true proposition $Q$ which has historical antecedents going back indefinitely far in time, $P$'s being negatively relevant to $Q$ only ensures $(\sim Q)MP$ if $P$ is also negatively relevant to the historical antecedents of $Q$. For example, supposing that getting plenty of vitamin C is positively relevant to living longer because it is negatively relevant to dying of certain diseases including pneumonia, given Smith who died of pneumonia we would say that he might have lived longer if he had gotten plenty of vitamin C; but given Jones who died in an auto accident, we would not say that Jones might have lived longer had he gotten plenty of vitamin C.

In order for considerations of relevance to bring it about that $(\sim Q)MP$, it is not required that $P$ be negatively relevant to *all* of the historical antecedents of $Q$. For example, suppose that getting plenty of vitamin C is negatively relevant to dying of pneumonia, but just

slightly positively relevant to dying of heart failure. Suppose Smith dies from pneumonia, with the actual causal mechanism consisting of the pneumonia bringing about heart failure. We would still agree that he might not have died when he did had he gotten plenty of vitamin C, even though his getting plenty of vitamin C is not negatively relevant to his suffering heart failure. The reason we would agree that he might not have died had he gotten plenty of vitamin C is that if we trace the historically antecedent circumstances implicating his dying back to a certain point in time (namely, the time at which he contracted pneumonia), we find that from that time back his getting plenty of vitamin C is negatively relevant to his being in those circumstances.

It seems then that the basic condition under which considerations of relevance can bring it about that $(\sim Q)MP$ is the following:

> $P$ is false and $Q$ is true, $P$ is weakly negatively relevant to $Q$, and there is a time $t$ such that $P$ is weakly negatively relevant to all historical antecedents of $Q$ predating $t$.

However, once we have generated some 'might be' statements in this way, we automatically get some more. For example, suppose $Q$ is a conjunction $\ulcorner Q_1 \ \& \ Q_2 \urcorner$. Then we must have either $(\sim Q_1)MP$ or $(\sim Q_2)MP$. There is an obvious constraint here. If $P$ is negatively relevant to $Q$ *because* it is negatively relevant to $Q_1$ and not relevant to $Q_2$, then we should have $(\sim Q_1)MP$, but we should not have $(\sim Q_2)MP$. More generally, if $\sim Q \Rightarrow (\sim R \vee \sim S)$, we have to have either $(\sim R)MP$ or $(\sim S)MP$. If $P$ is negatively relevant to $R$ but not to $S$, then we are constrained from having $(\sim S)MP$. When there is a choice between two propositions $\sim R$ and $\sim S$, and $P$ is negatively relevant to one of them, then that is the one we should choose as being one that might be true if $P$ were true. We can capture all of this by ruling:

(3.14) $\beta \mathbf{M}_\alpha \varphi$ if $\varphi$ is false in $\alpha$, and there is a $\psi$ true in $\alpha$ such that $\ulcorner(\varphi \ \& \ \sim\psi)\urcorner$ is true in $\beta$, and

(1) $\varphi$ is weakly negatively relevant to $\psi$, and there is a $t \in Rl$ such that for all $t^* < t$ and finite $X \subseteq S_\alpha(t^*)$, if $\Pi X$ is the conjunction of $X$ and $\Pi X \Rightarrow \psi$ then $\varphi$ is weakly negatively relevant to $X$; and

(2) for all $t \in Rl$.

(a) $(S_\alpha(t) \triangle S_\beta(t))$ is a minimal $(\forall N_\beta \cup \forall W_\beta \cup \{\ulcorner \varphi \ \& \sim \psi \urcorner\})$-change to $T_\alpha$ at $t$; and

(b) there is no indicative $\theta$ and $\gamma \in [[I]]_\alpha$ such that $(S_\alpha(t) \triangle S_\gamma(t))$ is a minimal $(\forall N_\gamma \cup \forall W_\gamma \cup \{\ulcorner \varphi \ \& \sim \psi \urcorner\})$-change to $T_\alpha$ at $t$, and $N_\gamma = N_\beta$ and $W_\gamma = W_\beta$, and $\theta \in (T_\beta - T_\alpha)$ but $\theta \notin T_\gamma$ and $\varphi$ is weakly negatively relevant to $\theta$.

Finally, consider what happens when we remove the restriction that $\varphi$ be indicative and not counter-legal. Removing this restriction does not seem to make any appreciable change to the analysis other than those changes already embodied in definition 2.15 of Chapter VI. The new definition becomes the same as the old 2.15 except that clause (vi) is replaced by the disjunction of the old clause (vi) and the new condition 3.14 above; and finally a new seventh clause must be added:

(VII) $P_\alpha = P_\beta$.

In defense of this final clause, notice that there is no way we can formulate within our language any sentence incompatible with the statements of positive relevance reflected in the set $P_\alpha$. To force changes in $P_\alpha$ we would at the very least have to be able to formulate probability statements in our language.

I believe that this new definition of **M** takes full account of the possibility of there being probabilistic laws. The definition of truth given in Chapter VI and based upon the definition of **M** can still be used without change. This completes the first of the two basic tasks of this chapter, which was to see what changes must be made to the definition of **M** to accommodate probabilistic laws.

## 4. Simple subjunctive probability

The second major task of this chapter is to examine those probability statements like ⌜If it were true that $P$, then it would probably be true that $Q$⌝ and ⌜If it were true that $P$, then it would almost certainly be true that $Q$⌝ which we are often inclined to make in lieu of asserting a

simple subjunctive. The suggestion was made above that in the absence of any true subjunctive generalizations (that is, if we had only probabilistic laws), these probability statements would be all that we would be warranted in asserting.

Statements like ⌜If it were true that $P$, then it would probably be true that $Q$⌝ *look* like simple subjunctive conditionals whose consequents are probability statements. But I think that this grammatical form is misleading. Taken at face value, it would give these probability statements the form $\ulcorner(P > \text{prob}(Q) \geq 1-\varepsilon)\urcorner$. The first difficulty with this is that it requires a notion of unconditional probability which proves very hard to come by. But supposing for the moment that we have such a probability concept, it would still not be reasonable to regard ⌜If it were true that $P$, then it would probably be true that $Q$⌝ as having the form $\ulcorner(P > \text{prob}(Q) \geq 1-\varepsilon)\urcorner$. The latter statement would require $\ulcorner\text{prob}(Q) \geq 1-\varepsilon\urcorner$ to be true in *every* $P$-world, but that is an unreasonably strong requirement. For example, suppose that $\ulcorner\text{prob}(Q) \geq 1-\varepsilon/2\urcorner$ is true in every $P$-world with the exception of one, $\beta$. Suppose further that if $P$ were true, it would be *much* less probable that $\beta$ would be the actual world than it would be that any other $P$-world would be the actual world. Surely, in this case, we would agree that if it were true that $P$ then it would probably be true that $Q$, although $\ulcorner(P > \text{prob}(Q) \geq 1-\varepsilon)\urcorner$ would be false.

These considerations indicate that ⌜If it were true that $P$ then it would probably be true that $Q$⌝ is not really a conditional. Given that it is not a conditional, it seems rather obvious that it must be a conditional probability statement. In this section I will define the kind of conditional probability which I believe to be involved here, and show how it is related to simple subjunctive conditionals.

### 4.1. *The Variety of Probabilities*

Philosophers sometimes make the mistake of supporting that there is just one reasonable concept of probability, and then criticize one another's pronouncements on probability from that point of view even though they are in fact talking about different concepts. We must avoid making that mistake.

The term 'probability' has been used to talk about a number of

different concepts. Three such concepts are 'degree of confirmation', 'degree of belief', and 'degree of rational belief'. It has never been obvious to me that degree of confirmation has the structure of a probability concept. Degree of belief (taken literally, and not idealized) almost certainly does not have such a structure, simply because people can hold irrational combinations of beliefs. However, degree of rational belief pretty clearly does have the structure of a probability concept, and it is a very interesting kind of probability to investigate.

I mention degree of rational belief primarily to contrast it with the kind of probability which will be investigated here. Let us symbolize degree of rational belief as '$\text{prob}_R$'. $\text{Prob}_R$ is primarily of use in deciding what actions to take in cases of partial ignorance. If $P$ represents everything we know, then $\text{prob}_R(Q/P)$ represents how likely it is that $Q$ is true given everything we know. As our state of knowledge changes, $\text{prob}_R$ changes (by changing $P$). This is all in strong contrast to the kind of subjunctive probability involved in ⌜If it were true that $P$, then it would probably be true that $Q$⌝. I will symbolize the latter kind of probability simply as 'prob'. $\ulcorner\text{prob}_R(Q/P) = r\urcorner$ is an indicative statement – it is about the probability that $Q$ *is* true given $P$. On the other hand, $\ulcorner\text{prob}(Q/P) = r\urcorner$ is a subjunctive statement – it is about the probability that $Q$ *would* be true *if*, contrary to fact, $P$ were true. $\text{Prob}(Q/P)$ is really only of interest in the case where $P$ is now false (this is also true of simple subjunctive conditionals). In determining the value of $\text{prob}(Q/P)$, we make use of everything that is true in the actual world, just as in evaluating the truth value of a subjunctive conditional we make use of everything that is true in the actual world. $\text{Prob}(Q/P)$ looks at all those worlds that might be actual if $P$ were true, i.e., at all $P$-worlds, and gives us the proportion of them that would make $Q$ true. This is a weighted proportion, the different $P$-worlds being weighted according to their relative likelihood of being actual if $P$ were true.

### 4.2. *Simple Subjunctive Probability*

The probability concept that will now be defined will be called 'simple subjunctive probability'. It is so-called because the simple subjunctive conditional is a limiting case of this probability concept. $\text{Prob}(Q/P)$ is

about what would be the case *if P were true*, thus it is a measure of the proportion of worlds making $Q$ true, not among all possible worlds making $P$ true, but among all worlds that might be actual if $P$ were true. We can define it in a way completely analogous to our definitions of strong and weak definite probabilities. First, define:

(4.1) $\mathbf{M}(P) = \{\beta; \beta \mathbf{M} P\}$

Then:

(4.2) $\text{prob}(Q/P) = \text{prob}_W(x \bullet T(Q)/(x \bullet \mathbf{M}(P)\ \&\ x \text{ is actual}))$.

We can also define 'prob' directly in terms of weak definite probabilities. First we define:

(4.3) $m_\alpha P \equiv (Q)[\alpha \bullet T(P > Q) \supset Q \text{ is true}]$.

(4.4) if $\alpha_0$ is the actual world, $mP \equiv m_{\alpha_0} P$.

$\ulcorner mP \urcorner$ is a proposition which is true in a world just in case that world is a $P$-world. Then we can define:

(4.5) $\text{prob}(Q/P) = \text{prob}_W(Q/mP)$.

Our simple subjunctive probability is a conditional probability. It is of course possible to define a non-conditional probability in the normal way:

(4.6) $\text{prob}(P) = \text{prob}(P/Q \vee \sim Q))$.

However, this turns out to be the same thing as the truth value of $P$. This results from the following two theorems:

(4.7) If $P$ is true, then $\text{prob}(Q/P) = \begin{cases} 1 \text{ if } Q \text{ is true} \\ 0 \text{ otherwise} \end{cases}$

*Proof:* if $P$ is true, then $\mathbf{M}(P) = \{\alpha_0\}$ (where $\alpha_0$ is the actual world). If $Q$ is true then $Q$ is true in all members of $\mathbf{M}(P)$, and so $\text{prob}(Q/P) = 1$. If $Q$ is false, then $Q$ is true in no member of $\mathbf{M}(P)$, so $\text{prob}(Q/P) = 0$.

(4.8) $\text{prob}(P) = \begin{cases} 1 \text{ if } P \text{ is true} \\ 0 \text{ if } P \text{ is false} \end{cases}$

Philosophers have often supposed that the connection between subjunctive conditionals and conditional probability should be that

$\text{prob}(P > Q) = \text{prob}(Q/P)$. Theorem 4.8 shows that this is not possible, at least for our notion of simple subjunctive probability. We will always have either $\text{prob}(P > Q) = 1$ or $\text{prob}(P > Q) = 0$, depending upon whether $\ulcorner P > Q \urcorner$ is true or false, but it will not generally be the case that either $\text{prob}(Q/P) = 1$ or $\text{prob}(Q/P) = 0$. What then is the connection between simple subjunctive probabilities and simple subjunctive conditionals? Just as in the case of subjunctive generalizations, the connection should be that the subjunctive conditional is a limiting case of subjunctive probability. Part of what this means is contained in the following trivial theorem:

(4.9) $\quad (P > Q) \supset \text{prob}(Q/P) = 1.$

If the situation here were completely analogous to the situation with subjunctive generalizations, then the full content of saying that the conditional is a limiting case of the probability statement would be contained in the following principle:

(4.10) $\quad (P > Q) \equiv (R)\,\text{prob}(Q/(P \;\&\; R)) = 1.$

Unfortunately, 4.10 is false. From right to left fails because we can trivially prove (substituting $\ulcorner \sim Q \urcorner$ for $R$):

(4.11) $\quad (R)\,\text{prob}(Q/(P \;\&\; R)) = 1 \supset (P \rightarrow Q).$

The other half of 4.10 fails for the same reason the principle

$$(P > Q) \supset [(P \;\&\; R) > Q]$$

fails. The source of both of these difficulties is that when we change subjunctive hypotheses, we change the set of possible worlds under consideration (from $\mathbf{M}(P)$ to $\mathbf{M}(P \;\&\; R)$), and there may be no simple connection between the two sets of possible worlds. In particular, we do not generally have that $\mathbf{M}(P \;\&\; R) = \mathbf{M}(P) \cap \mathbf{M}(R)$. For this same reason, it turns out that simple subjunctive probability does not satisfy all of the normal axioms for conditional probabilities. The measure-theoretic principles automatically hold:

(4.12) $\quad 0 \leq \text{prob}(Q/P) \leq 1.$

(4.13) $\quad (P \rightarrow Q) \supset \text{prob}(Q/P) = 1$

(4.14) $\quad (P \leftrightarrow Q) \supset \text{prob}(R/P) = \text{prob}(R/Q).$

(4.15) $\Diamond R \supset \text{prob}(\sim Q/P) = 1 - \text{prob}(Q/P)$.

(4.16) $(P \rightarrow Q) \supset \text{prob}(P/R) \leq \text{prob}(Q/R)$.

However, the following 'product axioms' (which are interderivable given 4.12–4.16) all fail:

(4.17) If $P \rightarrow Q$ and $Q \rightarrow R$ and $\text{prob}(Q/R) \neq 0$,

$$\text{then } \text{prob}(P/Q) = \frac{\text{prob}(P/R)}{\text{prob}(Q/R)}.$$

(4.18) If $Q \rightarrow \sim R$, then

$$\text{prob}(P/Q \vee R) = \text{prob}(P/Q) \cdot \text{prob}(Q/Q \vee R) + \text{prob}(P/R) \cdot \text{prob}(R/Q \vee R).$$

(4.19) $\text{prob}((P \ \& \ Q)/R) = \text{prob}(P/R) \cdot \text{prob}(Q/(P \ \& \ R))$.

As 4.17–4.19 are interderivable, it suffices to give a counter-example to just one of them. 4.19 is particularly interesting, because it can be viewed as a probabilistic analogue of principle 7.1 of Chapter III:

$$(R > Q) \ \& \ PMR. \supset [(R \ \& \ P) > Q].$$

This is the principle which, when added to *SS* yielded Lewis' *C*1, and which expressed the semantical principle that the ordering of possible worlds which generates subjunctive conditionals is connected. We can generate counter-examples to 4.19 in precisely the same way we generated counter-examples to 7.1. All we need is a case in which $\text{prob}(P/R) \neq 0$ (so *PMR*); and *P* and *Q* are independent so that $\text{prob}((P \ \& \ Q)/R) = \text{prob}(P/R) \cdot \text{prob}(Q/R)$; and $(R > Q)$ so that $\text{prob}(Q/R) = 1$, but $\text{prob}(Q/(P \ \& \ R)) \neq 1$. Given such a case, $\text{prob}(P/R) \cdot \text{prob}(Q/(P \ \& \ R)) < \text{prob}(P/R) = \text{prob}((P \ \& \ Q)/R)$. Such a case can be constructed using the same example which was a counter-example to 7.1. Let *S*, *T*, and *U* be the three unrelated false statements 'My car is painted black', 'My garbage can blew over', and 'My maple tree died'. Then let $P = \ulcorner (U \vee T) \urcorner$, $R = \ulcorner (S \vee T) \urcorner$, and $Q = \ulcorner \sim U \urcorner$.

Because of the failure of 4.17–4.19, we cannot express the sense in which the simple subjunctive conditional is a limiting case of simple subjunctive probability statements by asserting principle 4.10. We

must resort to more complicated techniques, which will be the topic of the next section.

Theorem 4.7, according to which

$$\text{If } P \text{ is true, then } \text{prob}(Q/P) = \begin{cases} 1 \text{ if } Q \text{ is true} \\ 0 \text{ if } Q \text{ is false} \end{cases}$$

expresses what is perhaps an unfortunate characteristic of simple subjunctive probability. By virtue of this theorem, these probabilities are really only useful in talking counterfactually about what would probably be the case if $P$ (which is now false) were true. But sometimes we employ a kind of probability statement which seems to circumvent this difficulty. For example, suppose a car goes out of control in heavy traffic and after careening about for a while collides with another car. We might observe that under the circumstances, *since* the car went out of control it was quite probable that there would be a collision. This notion of $Q$'s being probable since $P$ was true is an important notion that is worth investigating. It seems to be a probabilistic analogue of the necessitation conditional, which can be expressed (when it has a true antecedent) as ⌜It was true that $Q$ since it was true that $P$⌝. The more general form of a necessitation conditional is ⌜If it were true that $P$, then it would be true that $Q$ since it would be true that $P$⌝, and analogously the general form of our probability statement is ⌜If it were true that $P$, then it would be probable that $Q$ since it would be true that $P$⌝. Presumably, necessitation conditionals are the limiting case of these probability statements, and hence the analysis of necessitation conditionals can guide us in the analysis of the probability statements. The necessitation conditional is analyzable as:

$$(P \gg Q) \equiv \cdot(P > Q) \;\&\; [(\sim P \;\&\; \sim Q) > (P > Q)].$$

Correspondingly, I would suggest that ⌜If it were true that $P$, then it would be probable that $Q$ since it would be true that $P$⌝ is analyzable as:

> $Q$ would probably be true if $P$ were true, and if ⌜$(\sim P \;\&\; \sim Q)$⌝ were true, it would still be the case that $Q$ would probably be true if $P$ were true.

Presumably, to say that $Q$ would probably be true if $P$ were true is to

say that $\text{prob}(Q/P) \geq r$, for some particular $r$. Thus the above comes to:

$$\text{prob}(Q/P) \geq r \ \& \ [(\sim P \ \& \ \sim Q) > \text{prob}(Q/P) \geq r].^5$$

A reasonable reading of this would be ⌜$P$'s being true would dispose $Q$ to be true with at least a degree $r$⌝. Let us introduce a short symbol for this:

(4.20) $\quad (P \gg_r Q) \equiv \cdot \text{prob}(Q/P) \geq r \ \& \ [(\sim P \ \& \ \sim Q) > \text{prob}(Q/P) \geq r].$

We clearly have:

(4.21) $\quad (P \gg Q) \supset (P \gg_1 Q).$

It is worth noting in passing that conditionals like ⌜$(\sim P \ \& \ \sim Q) > \text{prob}(Q/P) \geq r$⌝ illustrate that we can have some true counterfactuals even if we have only probabilistic laws.

In 4.20 we introduced the notion of the degree to which $P$ disposes $Q$ to be true. It might be supposed that this is another kind of probability. It is easy enough to construct measures of this degree. An obvious definition would be:

$$\text{disp}(Q/P) = \text{l.u.b.}\{r; P \gg_r Q\}.$$

Actually, I think that a more interesting measure would be one taking a weighted average of the values $\text{prob}(Q/P)$ might have if ⌜$(\sim P \ \& \ \sim Q)$⌝ were true:

$$\text{disp}(Q/P) = \text{prob}(Q/P) \cdot \int_{x=0}^{x=1} x \cdot d \ \text{prob}(\text{prob}(Q/P) \leq x/(\sim P \ \& \ \sim Q)).$$

However, all of this is really rather beside the point, because it is easy to see that no matter how we define 'disp', it will not be a probability, i.e., it will not satisfy the measure-theoretic axioms of the probability calculus. This results from the fact, noted in Chapter III, that the following is not valid:

$$(P \gg Q) \supset [P \gg (Q \vee R)]$$

and hence we can have cases in which $\text{disp}(Q/P) = 1$, but $\text{disp}((Q \vee R)/P) < 1$.

4.3. *A Probability Algebra*

There is reason to believe that the relation ⌜If it were true that $P$, then it would be more probable that $Q$ than that $R$⌝ should have a fine structure not reflected in the values of prob($Q/P$) and prob($R/P$). Let us symbolize this relation as $\ulcorner Q <_P R \urcorner$. We can construct cases in which we seem to have $Q <_P R$, and yet prob($Q/P$) = prob($R/P$). The simplest such cases arise when prob($Q/P$) = prob($R/P$) = 0. For example, let $P$ be 'Joe is taller than six feet'. Let us suppose that if $P$ were true, then for any finite interval $\delta$ between 6′2″ and 6′4″, the probability that Joe's height lies in that interval is directly proportional to the length of the interval. Then it follows that if Joe were taller than six feet, then the probability of his height being any particular value between 6′2″ and 6′4″ is zero. If this were not the case, by adding a single point to an interval we could discontinuously increase the probability of Joe's height being in that interval. For each $x$ in the interval between 6′2″ and 6′4″, let $Q_x$ be the proposition that Joe's height is $x$. Then for each $x$, prob($Q_x/P$) = 0. Nevertheless, it seems clear that $(P \,\&\, \sim P) <_P Q_x$. Similarly, if $A$ is a finite set of points containing $x$, and $Q_A$ is the proposition that Joe's height lies in $A$, then $Q_x <_P Q_A$, although prob($Q_x/P$) = prob($Q_A/P$) = 0. Similarly, if $B$ is a denumerable set of points in the interval, and $A \subset B$, then it seems clear that $Q_A <_P Q_B$.

What the above examples illustrate is that among propositions of probability zero, the relation '$<_P$' has a fine structure not reflected in the probability values. But these examples yield additional examples concerning propositions having non-zero probability. For example, if $Q <_P R$, we surely want to say that $\sim R <_P \sim Q$. But if prob($Q/P$) = prob($R/P$) = 0, then prob($\sim Q/P$) = prob($\sim R/P$) = 1. So every example of fine structure on propositions of probability zero yields examples of fine structure on propositions of probability one.

It seems that we can go even further. We can construct intermediate cases with probability between zero and one. Suppose $Q$ is more probable than a contradiction, i.e., $(P \,\&\, \sim P) <_P Q$, but prob($Q/R$) = 0. Suppose prob($R/P$) ≠ 0, and $R \rightarrow \sim Q$. Then prob($(Q \vee R)/P$) = prob($Q/P$) + prob($R/P$) = prob($R/P$), but it seems that we should have $R <_P (Q \vee R)$. In general, it seems we should have:

(4.22) $\quad T(Q) \cap \mathbf{M}(P) \subset T(R) \cap \mathbf{M}(P) \supset Q <_P R.$

However, as we will see, there are some difficulties for this principle.

What are we to make of this relation '$\prec_P$' which exhibits a finer structure than our probability function? It turns out that it is definable in terms of our probabilities. It is not definable simply as

$$R \prec_P Q \text{ iff } \text{prob}(Q/P) < \text{prob}(R/P)$$

but it is definable in a more complicated way.

First consider the case of propositions of zero probability. We may have $\text{prob}(Q/P) = 0$ even though $Q$ might be true if $P$ were true, because $\mathbf{M}(P)$ is so much bigger than $\mathbf{M}(P) \cap T(Q)$ that any measure in $\Sigma$ sufficiently coarse to give $\mathbf{M}(P)$ a finite measure must give $\mathbf{M}(P) \cap T(Q)$ a zero measure. However, for the purpose of comparing $Q$ and $R$ (on the hypothesis $P$), we need not employ a measure which gives a finite value for $\mathbf{M}(P)$. It is not $\mathbf{M}(P)$ that we are interested in, but rather its two subsets $\mathbf{M}(P) \cap T(Q)$ and $\mathbf{M}(P) \cap T(R)$. We want to compare the sizes of these sets, and any measure in $\Sigma$ which gives them both finite values and makes at least one non-zero will do the job. This suggests the following definition:

(4.23) $Q \prec_P^{(1)} R$ iff $\text{prob}_W(Q/[mP \;\&\; (Q \vee R)]) < \text{prob}_W(R/[mP \;\&\; (Q \vee R)])$.

(4.24) $Q \approx_P^{(1)} P$ iff $\text{prob}_W(Q/[mP \;\&\; (Q \vee R)]) = \text{prob}_W(R/[mP \;\&\; (Q \vee R)])$.

What this definition does is move us from a finite measure for $\mathbf{M}(P)$ to a finite measure for the smaller set $\mathbf{M}(P) \cap T(Q \vee R)$, and this measure must give a non-zero measure to at least the larger of $\mathbf{M}(P) \cap T(Q)$ and $\mathbf{M}(P) \cap T(R)$.

Definition 4.23 works for the case of propositions of probability zero, but it does not work for the case of propositions of probability one. If $\text{prob}(Q/P) = \text{prob}(R/P) = 1$, then $Q \approx_P^{(1)} R$. However, we can rectify this situation easily enough as follows:

(4.25) $Q \prec_P^{(2)} R$ iff $Q \prec_P^{(1)} R$ or $\sim R \prec_P^{(1)} \sim Q$.

(4.26) $Q \approx_P^{(2)} R$ iff neither $Q \prec_P^{(2)} R$ nor $R \prec_P^{(2)} Q$.

The way in which this second ordering relation works can be seen

easily from the following theorems:

(4.27) $Q \prec_P^{(2)} R$ iff either:

(i) $\text{prob}(Q/P) = \text{prob}(R/P) = 0$ and $Q \prec_P^{(1)} R$; or

(ii) $\text{prob}(Q/P) = \text{prob}(R/P) = 1$ and $\sim R \prec_P^{(1)} \sim Q$; or

(iii) $\text{prob}(Q/P)$ and $\text{prob}(R/P)$ are neither both zero nor both one, and $\text{prob}(Q/P) < \text{prob}(R/P)$.

(4.28) $Q \approx_P^{(2)} R$ iff either:

(i) $\text{prob}(Q/P) = \text{prob}(R/P) = 0$ and $Q \approx_P^{(1)} R$; or

(ii) $\text{prob}(Q/P) = \text{prob}(R/P) = 1$, and $\sim Q \approx_P^{(1)} \sim R$; or

(iii) $\text{prob}(Q/P)$ and $\text{prob}(R/P)$ are neither both zero nor both one, and $\text{prob}(Q/P) = \text{prob}(R/P)$.

From this it follows easily that '$\approx_P^{(2)}$' is an equivalence relation, and '$\prec_P^{(2)}$' is a linear ordering relative to the equivalence relation.[6]

However, clause (iii) of 4.27 shows that '$\prec_P^{(2)}$' does not make finer discriminations than 'prob' in the intermediate case where $\text{prob}(Q/P)$ and $\text{prob}(R/P)$ are neither both zero nor both one. Hence '$\prec_P^{(2)}$' does not satisfy principle 4.22. We can construct a relation which does satisfy 4.22;

(4.29) $Q \prec_P^{(3)} R$ iff $(Q \ \& \sim R) \prec_P^{(1)} (R \ \& \sim Q)$.

The idea behind this definition is that rather than compare the entire sets $\mathbf{M}(P) \cap T(Q)$ and $\mathbf{M}(P) \cap T(R)$, we only compare those parts which they do not have in common: $\mathbf{M}(P) \cap T(Q \ \& \sim R) = (\mathbf{M}(P) \cap T(Q)) - (\mathbf{M}(P) \cap T(R))$; and $\mathbf{M}(P) \cap T(R \ \& \sim Q) = (\mathbf{M}(P) \cap T(R)) - ((\mathbf{M}(P) \cap T(Q))$. This new relation does seem to yield all of the fine structure we want:

(4.30) If $\text{prob}(Q/P) = \text{prob}(R/P) = 0$, then $Q \prec_P^{(3)} R$ iff $Q \prec_P^{(1)} R$.

(4.31) If $\text{prob}(Q/P) = \text{prob}(R/P) = 1$, then $Q \prec_P^{(3)} R$ iff $\sim R \prec_P^{(1)} \sim Q$.

(4.32) If $T(Q) \cap \mathbf{M}(P) \subset T(R) \cap \mathbf{M}(P)$ then $Q \prec_P^{(3)} R$.

So far, '$\prec_P^{(3)}$' looks like a fine relation – just the one we want. But there

is a difficulty. The difficulty concerns how we are to define equiprobability. As far as I can see, the only plausible definition is:

(4.33) $Q \approx_P^{(3)} R$ iff neither $Q \prec_P^{(3)} R$ nor $R \prec_P^{(3)} Q$.

Unfortunately, so defined, '$\approx_P^{(3)}$' is not an equivalence relation. This can be seen by employing the following theorem:

(4.34) If $P > \sim(Q \ \& \ R)$ is true (i.e., $\mathbf{M}(P) \cap T(Q)$ is disjoint from $\mathbf{M}(P) \cap T(R)$), then $Q \approx_P^{(3)} R$ iff $\text{prob}(Q/P) = \text{prob}(R/P)$.

Given this theorem, we can immediately see that it is possible to have $Q \prec_P^{(3)} R$, but $S \approx_P^{(3)} Q$ and $S \approx_P^{(3)} R$. For example, we simply choose $Q$, $R$, and $S$ so that (1) $S$ is disjoint from both $Q$ and $R$ (i.e., $\ulcorner P > \sim(S \ \& \ Q)\urcorner$ and $\ulcorner P > \sim(S \ \& \ R)\urcorner$ are true), and $\text{prob}(S/P) = \text{prob}(Q/P)$, and (2) $Q \rightarrow R$, $\mathbf{M}(P) \cap T(R \ \& \ \sim Q) \neq \varnothing$, but $\text{prob}((R \ \& \ \sim Q)/P = 0$. This result makes '$\prec_P^{(3)}$' a rather peculiar relation. It may be quite useful for some purposes, but it is not as useful for our present purposes as is '$\prec_P^{(2)}$', which is better behaved but agrees with '$\prec_P^{(3)}$' at the extremities of probability zero or probability one. Thus I shall define:

(4.35) $Q \prec_P R$ iff $Q \prec_P^{(2)} R$.

(4.36) $Q \approx_P R$ iff $Q \approx_P^{(2)} R$.

We have a well-behaved probability algebra which makes discriminations beyond those made by 'prob', and in terms of which we can explain in precisely what sense simple subjunctive conditionals are a limiting case of simple subjunctive probabilities. To be maximally probable is to be equiprobable with a tautology. We now have the simple theorem:

(4.37) $\ulcorner(P > Q)\urcorner$ is true iff $Q \approx_P (Q \vee \sim Q)$.

*Proof:* $\ulcorner(P > Q)\urcorner$ is true iff $\mathbf{M}(P) \subseteq T(Q)$, iff $\mathbf{M}(P) \cap T(\sim Q) \neq \varnothing = \mathbf{M}(P) \cap T(\sim(Q \vee \sim Q))$, iff $Q \approx_P (Q \vee \sim Q)$.

Thus for the conditional to be true is just for $Q$ to be maximally probable given $P$. Similarly, for $\ulcorner QMP\urcorner$ to be true is for $Q$ not to be minimally probable given $P$:

(4.38) $\ulcorner QMP\urcorner$ is true iff $Q \not\approx_P (Q \ \& \ \sim Q)$.

## 4.4. *Simple Subjunctive Probability and Simple Subjunctive Conditionals*

Our probability algebra demonstrates the intimate connection between simple subjunctive probability and simple subjunctive conditionals. The nature of this connection allows us to understand many features of subjunctive conditionals which previously seemed puzzling. For example, people often feel some misgivings about the analysis of $\ulcorner QMP \urcorner$ as $\ulcorner \sim(P > \sim Q) \urcorner$. The first misgiving is that $\ulcorner QMP \urcorner$ is expressed in English as a conditional: $\ulcorner$If it were true that $P$, then it might be true that $Q \urcorner$. Why, then, is it analyzed as the *negation* of a conditional? Principle 4.38 provides the answer by showing that $\ulcorner QMP \urcorner$ is a kind of conditional probability statement. It amounts to saying that if $P$ were true, then $Q$ would not be totally improbable.

The second misgiving about the analysis of $\ulcorner QMP \urcorner$ is that when we assert it to be the negation of $\ulcorner$If it were true that $P$, then it would be false that $Q \urcorner$ we feel a bit of strain, and are inclined to want to contrast it instead with $\ulcorner$If it were true that $P$, then it would *definitely* be false that $Q \urcorner$. The explanation for this hinges upon what I regard as an extremely important fact about subjunctive conditionals. This is that we frequently assert $\ulcorner (P > Q) \urcorner$ when it is not really true – when all that is literally true is something like $\ulcorner$If it were true that $P$, then it would almost certainly be true that $Q \urcorner$. It is this careless assertion of $\ulcorner (P > \sim Q) \urcorner$ we are resisting when we insert 'definitely' into the conditional.

The observation that we often assert a simple subjunctive when all that is really true is a probability statement is, I think, very important, partly because it is such a pervasive tendency. Consider an example. Suppose we have a cylinder of gas, one end of the cylinder being formed by a movable piston, and suppose there is a temperature mechanism which holds the temperature of the gas constant as we move the piston in and out. We would be very much inclined to say that if the pressure of the gas increased, the piston would have been depressed. But this subjunctive conditional is not really true. If the pressure were to increase, this *might* be because the temperature mechanism went awry. This is so much less probable an alternative that we are inclined to ignore it, but all that is strictly warranted here is

the statement that if the pressure were to increase then very probably, or almost certainly, the piston would have been depressed.

Lest someone think that perhaps all subjunctive conditionals are literally false, and all that is ever really true is a subjunctive probability statement, it is worth pointing out that the converse of the conditional in the above example is quite literally true. That is, if the piston had been depressed, then the pressure would have increased. This is because the piston's being depressed would undercut the pressure's being what it was, but would not undercut the proper functioning of the temperature mechanism.

The continuum of subjunctive probability statements with subjunctive conditionals as the upper bound makes it much easier to understand what is happening in many cases in which, if we are restricted to conditionals and not allowed to employ probability statements, we do not quite know what to say. For example, consider a person, Joe, who is 5′11″ tall. If Joe were over six feet tall, how tall might he be? Clearly, he might be 6′1″, and he might be 6′2″, and he might be 6′3″, and so on for a while. But it seems there must be an upper bound. Isn't it true that even if he were over six feet tall, he would not be one thousand miles tall? If so, then there must be some point $h$ in between such that he might be any height less than $h$, but he would definitely not be any height greater than $h$. The difficulty is that it is not at all clear what the value of $h$ might be. Unless there is some physical law which dictates that a man cannot have a height greater than a certain magnitude $h$ (and I know of no such law), then it is hard to see how any height $h$ could come to constitute a sharp dividing line between those heights Joe might be if he were over six feet tall and those height that he would not be if he were over six feet tall. Something here seems very mysterious.

I think that the solution to our difficulty lies in re-examining our initial supposition that if Joe were over six feet tall, he would not be one thousand miles tall. It is of course *extraordinarily* unlikely that he would be one thousand miles tall, so unlikely that for all practical purposes we can ignore the possibility altogether, but I doubt that there is anything which rules this out beyond all possibility. Because it is so extraordinarily unlikely that Joe would be one thousand miles tall, we also tend to ignore this possibility in our judgments regarding what

would be the case if he were over six feet tall, and so we assert that if Joe were over six feet tall, he would not be one thousand miles tall. But this conditional is literally false. There is just the slimmest possibility that he would be one thousand miles tall, and so all that is literally true is the subjunctive probability statement that if Joe were over six feet tall, he would almost certainly not be one thousand miles tall.

I suspect that a similar analysis can be given for many cases which might initially appear to be counter-examples to our analysis of subjunctive conditionals in terms of undercutting. For this reason an understanding of subjunctive probability is as important for understanding subjunctive conditionals as it is in its own right.

## 4.5. *Simple Indefinite Probabilities*

We have defined two kinds of indefinite probabilities corresponding to the two kinds of subjunctive generalizations. We can define a third kind of indefinite probability which corresponds to simple subjunctive definite probability. This will be called 'simple indefinite probability'. It will be of some importance in the next chapter where we undertake to analyze disposition statements.

We want prob($Gx/Fx$) to be the probability of an $F$ being a $G$, given the way the world actually is. It seems natural to suggest that this indefinite probability can be defined as being prob($(\exists x)Gx/(\exists x)Fx$). This is initially plausible, but that it will not work follows from Theorem 4.7 according to which if $\ulcorner(\exists x)Fx\urcorner$ is true, prob($(\exists x)Gx/(\exists x)Fx$) is either one or zero. However, I think that this general idea can be made to work.

Let us begin by asking what is the probability of an $F$ being a $G$ *given that there is just one* $F$? That is, taking $\ulcorner(\exists_1 x)Fx\urcorner$ to symbolize $\ulcorner$there is exactly one $F\urcorner$, we want to know the value of prob($Gx/(Fx\ \&\ (\exists_1 x)Fx)$). It seems to me that this should be:

(6.1) $\quad \text{prob}(Gx/(Fx\ \&\ (\exists_1 x)Fx)) = \text{prob}((\exists x)(Fx\ \&\ Gx)/(\exists_1 x)Fx).$

What then is the probability of an $F$ being a $G$ given that there are exactly two $F$'s? It seems that we should be able to calculate this like

an expectation value:

(6.2) $\text{prob}(Gx/(Fx \,\&\, (\exists_2 x)Fx)) = \text{prob}((\exists_2 x)(Fx \,\&\, Gx)/(\exists_2 x)Fx)$
$+\text{prob}(Gx/(Fx \,\&\, (\exists_1 x)(Fx \,\&\, Gx) \,\&\, (\exists_2 x)Fx))$
$\cdot\text{prob}((\exists_1 x)(Fx \,\&\, Gx)/(\exists_2 x)Fx).$

In general:

(6.3) $\text{prob}(Gx/(Fx \,\&\, (\exists_n x)Fx)$
$= \sum_{0<k\leq n} \text{prob}(Gx/(Fx \,\&\, (\exists_k x)(Fx \,\&\, Gx) \,\&\, (\exists_n x)Fx))$
$\cdot\text{prob}((\exists_k x)(Fx \,\&\, Gx)/(\exists_n x)Fx).$

$\text{prob}(Gx/(Fx \,\&\, (\exists_k x)(Fx \,\&\, Gx) \,\&\, (\exists_n x)Fx))$ is the probability of an $F$ being a $G$ given that there are exactly $n$ $F$'s and $k$ of them are $G$'s. This probability would seem to be $k/n$, so principle 6.3 can be simplified to read:

(6.4) $\text{prob}(Gx/(Fx \,\&\, (\exists_n x)Fx))$
$= \sum_{0<k\leq n} \frac{k}{n} \text{prob}((\exists_k x)(Fx \,\&\, Gx)/(\exists_n x)Fx).$

Symbolizing the statement that there are infinitely many $F$'s as $\ulcorner(\exists_\infty x)Fx\urcorner$, how should we define $\text{prob}(Gx/(Fx \,\&\, (\exists_\infty x)Fx))$? the only plausible way I can see to define this is as the limit of the probabilities for larger and larger finite sets of $F$'s:

(6.5) $\text{prob}(Gx/(Fx \,\&\, (\exists_\infty x)Fx)) = \lim_{n\to\infty} \text{prob}(Gx/(Fx \,\&\, (\exists_n x)Fx))$.

Letting $\ulcorner(\exists_{\leq n} x)Fx\urcorner$ symbolize the statement that there are no more than $n$ $F$'s, we can next define:

(6.6) $\text{prob}(Gx/(Fx \,\&\, (\exists_{\leq n} x)Fx))$
$= \sum_{0<k\leq n} \text{prob}(Gx/(Fx \,\&\, (\exists_k x)Fx))\cdot\text{prob}((\exists_k x)Fx/(\exists_{\leq n})Fx))$.

Then it seems reasonable to define $\text{prob}(Gx/(Fx \,\&\, \sim(\exists_\infty x)Fx))$, the probability of an $F$ being a $G$ given that there are finitely many $F$'s, to be the limit of $\text{prob}(Gx/(Fx \,\&\, (\exists_{\leq n} x)Fx))$ as $n$ becomes indefinitely

large, and so finally to define $\text{prob}(Gx/Fx)$ as:

(6.7) $\text{prob}(Gx/Fx) = \text{prob}(Gx/(Fx \,\&\, (\exists_\infty x)Fx))$
$\cdot \text{prob}((\exists_\infty x)Fx/(\exists x)Fx) + \text{prob}(\sim(\exists_\infty x)Fx/(\exists x)Fx)$
$\cdot \lim_{n\to\infty} \text{prob}(Gx/(Fx \,\&\, (\exists_{\leq n} x)Fx))$.

Principle 6.7 seems to constitute a reasonable definition of simple indefinite probability on the basis of simple subjunctive definite probability. However, there are some surprises forthcoming from this definition. By virtue of Theorem 4.7 we obtain:

(6.8) If $\ulcorner(\exists_n x)Fx\urcorner$ and $\ulcorner(\exists_k x)(Fx \,\&\, Gx)\urcorner$ are true, then $\text{prob}(Gx/Fx) = k/n$.

That is, if there are finitely many $F$'s, then $\text{prob}(Gx/Fx)$ is just the material probability $\text{prob}_M(Gx/Fx)$. For many purposes, this is an entirely reasonable result. We want to know the probability of an $F$ being a $G$ given the way the world actually is, so if there are some $F$'s in the world, this probability should be the actual proportion of $F$'s that are $G$'s.

However, for some purposes we would also like a different kind of indefinite probability which still reflects the way the world actually is. This would reflect not simply the actual proportion of $F$'s that are $G$'s, but the likelihood of a new $F$ being a $G$. For example, if we are flipping a coin, we would like to know not just what proportion of flips already made have resulted in heads, but how likely it is that a new flip would result in heads. This probability can be defined very easily using a now-familiar construction:

(6.3) $\text{prob}_+(Gx/Fx) = (\iota r)(\exists A)[(x)(x \in A \equiv x = x)$
$\&\ \text{prob}(Gx/(Fx \,\&\, x \notin A)) = r]$.

$\text{Prob}_+(Gx/Fx)$ is the simple indefinite probability of a new or different $F$ being a $G$. This will turn out to be a very important probability concept. In particular, it will be involved in the analysis of dispositions.

## NOTES

[1] Note that '$\mu_F = \mu_G$' is an equivalence relation. It will hold iff $\text{prob}_S(Gx/(Fx \vee Gx)) \neq 0$ and $\text{prob}_S(Fx/(Fx \vee Gx)) \neq 0$.

[2] It is worth remarking that the analogous theorem holds for weak indefinite probabilities, weak subjunctive generalizations, and actual possibility.

[3] If, as is often maintained, given any set of possible worlds there is a proposition true in just those worlds, then it follows that infinite conjunctions of propositions exist, and hence that there is always a maximal $P$.

[4] More precisely, this is the proportion of worlds making $Q$ true out of those worlds which might be physically possible if $P$ were physically possible and which make $P$ true.

[5] Up to this point I have been intentionally vague about whether $\ulcorner \text{prob}(Q/P) \urcorner$ is a metalinguistic relation between sentences or an object-language term-forming operation. It is clear that for our present purposes we must opt for the latter alternative.

[6] This means that the ordering of the equivalence classes imposed by the ordering '$\prec_P^{(2)}$' of their members is a linear ordering.

CHAPTER IX

# DISPOSITIONS

## 1. Introduction

The final kind of subjunctive statement to be considered in this book is that of a statement ascribing a disposition to an object. Traditional examples of such statements would be:

(1) That liquid is flammable.

(2) This vase is fragile.

(3) That chalk is friable.

(4) Joe is foolhardy.

(5) Mary is inquisitive.

The problem is how to analyze such statements.

Remarkably little has been written about dispositions in the last decade. I suspect that this is due largely to philosophers feeling that this problem has been solved, or at least successfully reduced to the problem of analyzing subjunctive conditionals. For example, it seems initially plausible to suppose that (1) is analyzable as:

(1*) If that liquid were heated, it would burn.

Thus the feeling is that the only problem remaining is that of analyzing subjunctive conditionals, and there is really no point in further discussion of dispositions *per se*.

We have presented an analysis of subjunctive conditionals, but unfortunately we cannot rest content that we have thereby solved the problem of dispositions. The traditional view according to which disposition statements are analyzable on the model of (1*) is completely and unalterably wrong. A first glimmering of difficulty for the traditional view was noted by Goodman (1955) who pointed out that

(1) does not entail (1*). There are numerous circumstances under which (1) would be true but (1*) false. For example, there might be no oxygen present. Thus, as Goodman observes, if we are to defend something like the traditional view, we must retreat to:

(1**) If conditions were propitious and that liquid were heated, then it would burn.

But of course, without an account of what it is for conditions to be propitious, this is not an analysis. As we will see, following this line of thought to its logical conclusion leads ultimately to a completely different kind of account of dispositions.

However, before continuing the discussion of the above difficulty, it behooves us to note an entirely different sort of difficulty. Our list of disposition statements contains statements about two entirely different sorts of dispositions. Ignoring the difficulties about propitious conditions, (1)–(3) seem to be analyzable roughly on the model of (1**). These disposition statements at least seem to entail conditionals to the effect that if something were the case then something else would be the case. But (4) and (5) are markedly different. They do not entail any such conditionals. (4) and (5) report *tendencies*. They tell us how Joe and Mary tend to behave under certain sorts of circumstances. Unlike (1)–(3), they do not tell us what definitely would happen under some circumstances, but only what would be likely to happen. We have a generally overlooked distinction here between what may be termed 'absolute dispositions' and 'probabilistic dispositions'. I do not know of a single discussion of the difference between these kinds of dispositions, and yet an account of the sort of (1**) is obviously inappropriate for probabilistic dispositions. This is a rather remarkable oversight because, historically, those dispositions which have most interested philosophers have tended to be mental or psychological dispositions, and these are perhaps without exception of the probabilistic variety.

It will turn out that there are definite similarities between the absolute and the probabilistic dispositions. These arise from the connections noted in the last chapter between subjunctive conditionals and subjunctive probabilities. Once we have seen how to analyze statements ascribing absolute dispositions, it will become much clearer how to analyze statements ascribing probabilistic dispositions.

## 2. Absolute dispositions

It is surprisingly difficult to find examples of absolute dispositions. Almost all of the dispositions which philosophers customarily discuss are of the probabilistic variety. However, the following, picked at random from the dictionary, all seem to be of the absolute variety: 'absorbant', 'addictive', 'adhesive', 'deflatable', 'fissionable', 'flammable', 'flexible', 'fluorescent', 'fragile', 'friable', 'magnetized', 'magnetizable', 'soluble'.

According to the (modified) traditional view, '$x$ is $\varphi$-able' is analyzable as 'If conditions were propitious and $x$ were $\psi$-ed, then $x$ would $\varphi$', where $\psi$ is the appropriate antecedent for $\varphi$-ability. Ignoring for the moment the difficulties about what constitutes propitious conditions, we can see that this analysis is inadequate for another reason. Where does the antecedent $\psi$ come from? If this analysis is to work, $\psi$ must be involved in the concept of $\varphi$-ability. For certain dispositions, this seems plausible. For example, '$x$ is absorbant' appears to mean something like 'If conditions were propitious and $x$ were placed in contact with a liquid, $x$ would absorb the liquid'; and '$x$ is flammable' appears to mean something like 'If conditions were propitious and $x$ were heated sufficiently, $x$ would burn'. But if we turn to dispositions like 'deflatable', 'fissionable', and 'friable', this becomes much less plausible. What it takes to deflate something depends upon what it is that is being deflated. We *discover* how to make fissionable material fission. And many different circumstances will lead to friable material crumbling. It is not plausible to regard these antecedents as being built into the meaning of the associated dispositions. Consider friability again. We might have one kind of friable material which will crumble when hit with a hammer, but another kind of friable material which is unaffected by a sharp blow but which crumbles when subjected to sustained pressure. It is not plausible to regard either of these circumstances as being built into the concept of friability. We *discover* what it takes to make friable material crumble, and the world might have been other than it is so that different circumstances were involved.

The attempt to analyze 'deflatable', 'fissionable', and 'friable' in terms of conditionals is, I think, just wrong. These are best construed

as *capabilities*. To be deflatable is to be capable of being deflated; to be fissionable is to be capable of undergoing fission; to be friable is to be capable of being made to crumble. What about our other dispositions for which there are obvious antecedents? Consider 'absorbant'. The natural antecedent here is '$x$ is placed in contact with the liquid'. But this is only natural because we know something about the circumstances in which absorbant materials absorb liquids. It is still only a contingent fact that these are the appropriate circumstances. It might have been the case instead that absorbant materials only absorb liquids when there is a slight air space between them, the air somehow facilitating the movement of molecules. Turning to 'flammable', there is already something peculiar about the antecedent '$x$ is heated sufficiently'. What does 'sufficiently' mean here? And I should think once more that it is only a contingent fact that heating flammable objects makes them burn. There are other ways to ignite some flammable objects, e.g., pouring certain chemicals on them, and I see no reason in general why it couldn't have been the case that flammable objects are made to burn by cooling them rather than heating them. Heating an object just happens to be the most common and generally the simplest way to make it burn. But if there is *any* way to make an object burn, then it is flammable. To be flammable is to be *capable* of being made to burn. Similarly, to be absorbant is to be capable of absorbing liquids. There are no antecedents built into absolute dispositions. The extent to which we are able to supply such antecedents reflects on the one hand our knowledge of how these dispositions work, and on the other hand the uniformity of the way they work (e.g., there are too many different ways of being deflatable for us to be able to supply a single antecedent). The ability to supply the antecedents does not arise simply from an understanding of the concepts involved. In general, absolute dispositions are best regarded as *capabilities*. Philosophers have misappropriated the term 'disposition' in talking about absolute dispositions. The term 'disposition' is really only appropriate for probabilistic dispositions, which are truly 'tendencies'.

The observation that absolute dispositions are really capabilities explains the difficulty regarding propitious conditions. The reason we had to include the propitiousness of the conditions in the antecedent of our conditionals is that the antecedents supplied by our common

understanding of the functioning of an absolute disposition do not generally state a 'total cause' for the actuation of the capability. But this becomes irrelevant if the antecedent is not part of the meaning of the concept anyway.

Once it is realized that absolute dispositions are really capabilities, it is rather easy to see how they are to be analyzed. For example, to say that a certain liquid is flammable is to say that it is capable of burning, which seems to be to say that it is physically possible for it to burn. Similarly, to say that a piece of chalk is friable appears to be equivalent to saying that it is physically possible for it to crumble. In general,

(2.1) $x$ is $\varphi$-able iff $\underset{p}{\Diamond}(x \text{ is } \varphi)$.

I believe that 2.1 is basically correct, but more must be said about how it relates to particular disposition statements. Consider fragility. If we try to fit fragility into the format of 2.1, we find that it cannot be done. The nearest we can come is something like:

$x$ is fragile iff it is physically possible for $x$ to break.

But this is incorrect. This would at best be a definition of 'breakable' rather than 'fragile'. To be fragile is not just to be breakable, but to be 'quite breakable'. We distinguish between degrees of breakability, flammability, friability, etc., and many of our 'disposition words' refer to particular degrees of having capabilities. For example, a piece of 'non-flammable' plastic might be made to burn by heating it to a very high temperature in a pure oxygen environment. Thus the plastic is flammable (i.e., has the capability of burning), but just barely so. Because it is so hard to make the plastic burn, it has the capability to such a minimal degree that we say it is non-flammable. Our ordinary use of the word 'non-flammable' is such that non-flammable objects need not be *totally* non-flammable.

This indicates that 2.1 should not be taken as the scheme for defining disposition words, but rather as the scheme for defining what we might call 'basic capabilities'. Then most disposition words refer to varying degrees of these basic capabilities. For example, we might define the basic capability '$x$ is breakable' as 'It is physically possible for $x$ to break'. The degrees of this basic capability constitute a continuum stretching from total non-breakability through extreme

fragility, and different disposition words may pick out different regions of this continuum. In particular, 'breakable' refers to a rather large region including anything which is 'reasonably breakable'. Thus the term 'breakable' can be used both to refer to the basic capability and to refer to a region in the continuum of degrees of that basic capability. This seems to be true of most disposition words which can also be regarded as naming basic capabilities. They do double duty.

Principle 2.1 constitutes an analysis of basic capabilities, but it does not yet constitute an analysis of most disposition statements, because such statements generally refer to specific degrees of basic capabilities. We must explore this notion of the degree to which something has a capability. To say that $x$ is more flammable than $y$ is to say that it is 'easier' to get $x$ to burn than it is to get $y$ to burn. Similarly, to say that $x$ is more breakable than $y$ is to say that it is easier to break $x$ than it is to break $y$. In general, the degree to which an object has a capability is a measure of the ease with which that capability can be realized in that object. But 'ease' in what sense? We might naturally suppose that this has something to do with human abilities and how easy it is or hard it is for humans to cause the capability to be realized. But this cannot be right, because we can talk about a capability whose realization is simply beyond human ability. For example, an object which will ignite at a temperature of $10^{10^{10}}$ degrees centigrade is more flammable than one which will not ignite until it is heated to $10^{10^{10^{10}}}$ degrees centigrade, but it is beyond human ability to heat an object to either temperature. A natural second suggestion would be that talk about degrees of capabilities is to be analyzed in terms of the amount of energy that must be expended to realize the capability. But this cannot be right either. The concept of energy only plays the role it does in the realization of capabilities because of contingent physical laws, and as such it cannot be involved in the very concept of the degree to which something has a capability. Furthermore, the amount of energy that must be expended is not really a correct measure of the degree to which something has a capability anyway. Consider two substances, one of which can be ignited by heating it to any temperature above 100 degrees centigrade in the presence of any gas that would normally be considered 'breathable air'. Consider a second substance which can only be ignited by heating it to a temperature between 91.333 333 333

degrees centigrade and 91.333 333 334 degrees centigrade, holding it at that temperature for 37.375 seconds, and then abruptly changing the temperature to between 92.334 788 8 degrees centigrade and 92.334 788 9 degrees centigrade and holding it there for 10 seconds, all in the presence of pure oxygen. It takes less energy to ignite the second substance than to ignite the first substance, but surely we would consider the first substance more flammable than the second substance. Items made of the second substance would almost never burn because of the critical sequence of steps required to ignite them.

If an object is flammable, there will generally be infinitely many ways to make it burn, e.g., heating it to 100°, heating it to 101°, heating it to 102°, . . . , dipping it in sulphuric acid and then placing it in a pure oxygen environment, etc. To say that one object is more flammable than another is to compare how many ways there are to get them to burn. Or better, it is to employ a measure function which measures the physically possible ways to get the objects to burn. A physically possible way to get an object to burn can be thought of as picking out a set of physically possible worlds in which the object does burn. Thus to say that $x$ is more flammable than $y$ is to employ a measure comparing the set of all physically possible worlds in which $x$ burns and the set of all physically possible worlds in which $y$ burns. For example, if the only relevant difference between $x$ and $y$ is that $y$ has a higher kindling point than $x$, then any way of making $y$ burn is also a way of making $x$ burn, so the measure of the set of physically possible worlds in which $x$ burns must be greater than the measure of the set of physically possible worlds in which $y$ burns.

But what is the nature of this measure? It is surely inadequate simply to say that there is such a measure. Fortunately, the measure involved is a familiar one. All it does is measure the size of a set of physically possible worlds. Furthermore, if there are probabilistic laws which dictate that some possible worlds are less probable than others, the measure should reflect this. If all possible worlds in which $x$ would burn are highly improbable in light of some basic probabilistic laws, then it is correspondingly more difficult to get $x$ to burn, and so the measure of physically possible worlds in which $x$ does burn should be correspondingly smaller. Thus the measure involved here does precisely the same thing as the measures involved in strong definite

probabilities. To say that the measure of the set of physically possible worlds in which $x$ burns is smaller than the measure of the set of physically possible worlds in which $y$ burns is to compare the strong probability of $x$'s burning with the strong probability of $y$'s burning.

Strong probabilities are conditional probabilities, so we cannot talk about the probability of $x$'s burning *simpliciter*. A natural suggestion would be that as we are merely interested in comparing the set of worlds in which $x$ burns with the set of worlds in which $y$ burns, we should have that $x$ is more flammable than $y$ iff:

$$\text{prob}_S(y \text{ burns}/(x \text{ burns} \vee y \text{ burns})) < \text{prob}_S(x \text{ burns}/(x \text{ burns} \vee y \text{ burns})).$$

But this cannot be quite right. The difficulty is that if it were less likely for $x$ to exist than for $y$ to exist, this would correspondingly diminish the measure of the set of worlds in which $x$ burns, but this should not be relevant to the flammability of $x$. This suggests that the conditional probabilities we should be looking at are instead the probabilities of objects burning given that they exist. On this proposal:

(2.2) $x$ is more flammable than $y$ iff

$$\text{prob}_S(y \text{ burns}/y \text{ exists}) < \text{prob}_S(x \text{ burns}/x \text{ exists}).$$

Principle 2.2 is close to being correct, but it runs into a surprising difficulty. A piece of asbestos is much less flammable than a piece of sodium. So far, principle 2.2 seems to give the correct answer. But now consider a brick *made* of asbestos and a brick *made* of sodium. We want to say that the brick made of asbestos is less flammable than the brick made of sodium, but this result is not forthcoming from 2.2. The difficulty is that it is within the realm of physical possibility to transmute the elements in the asbestos into sodium atoms, thereby transforming the brick made of asbestos into a brick made of sodium. It is still the same brick, but it is no longer made of asbestos.[1] Analogously, it is physically possible for the atoms of sodium in the second brick to be transmuted into the elements that make up asbestos thereby transforming the second brick into a brick made of asbestos. This has the consequence that for every physically possible world in which $x$, the brick originally made of asbestos, exists, there is a physically possible

world in which $y$, the brick originally made of sodium exists and has the same attributes as $x$ which are relevant to their burning. Consequently, despite our judgment that $x$ is less flammable than $y$,

$$\text{prob}_S(x \text{ burns}/x \text{ exists}) = \text{prob}_S(y \text{ burns}/y \text{ exists}).$$

The difficulty arises here because our probability measure simply looks at the *size* of the set of all physically possible worlds in which $x$ burns, and does not take into account how likely it is for different members of the set to exist.

It might seem that we could avoid these difficulties by changing the antecedent of our probability from ⌜$x$ exists⌝ to something like ⌜$x$ has the nature it presently has⌝. The latter is obviously a problematic notion, but the intention would be to judge $x$'s flammability in terms of its present physical makeup. However, insofar as this would rule out our looking at those worlds in which transmutation of elements occurs, this is now too strong an antecedent. We do not want to rule out transmutation altogether; rather we want to take into account the fact that it is an unlikely occurrence. If instead transmutation were very easy to bring off, we might well find ourselves regarding bricks of asbestos as fuel and burning them in furnaces which first transmute their elements. In such a case, we would regard them as highly flammable. Thus we are not interested simply in the probability that $x$ burns given that it has the nature it does. Instead, we are interested in the probability of $x$ *coming to burn* given that it has the nature it does. This would allow for the possibility that the nature of $x$ changes before it begins to burn.

But what are we talking about when we talk about $x$ having a certain nature? The intention is, among other things, to talk about $x$'s physical composition. Thus, for example, $x$'s being composed of asbestos is relevant to the computation of the probability measuring $x$'s flammability. However, that $x$ is now submerged in water is not relevant to its flammability, although including this fact in our antecedent would certainly alter the probability of $x$'s coming to burn. Thus some facts about $x$ should be included in its 'nature', but not all facts. Can we make sense of this? In fact, I think we can. The facts about $x$ which are included in $x$'s nature are just those simple truths which are about $x$. That $x$ is composed of asbestos is a simple truth about $x$, but that $x$ is

immersed in water is not a simple truth. Rather, that $x$ is immersed in $y$ is a simple truth, and that $y$ is water is a simple truth, but the conjunction of these two simple truths is not a simple truth. Thus it appears that the probability which measures $x$'s degree of flammability is $\text{prob}_S(x$ will come to burn/$x$ has the set of simple attributes it in fact has). To make this precise, notice that this probability will be the same for any object at all having the same simple attributes as $x$. What is relevant to this probability is not the identity of the object $x$, but rather the set of simple attributes of $x$. Thus we can recast this as an indefinite probability. Let us define:

(2.3) $\Pi_{a,t}(x) \equiv x$ has all of the simple attributes possessed by $a$ at $t$.

Then we have:

(2.4) $a$ is more flammable than $b$ at time $t$ iff $\text{prob}_S(x$ will come to burn/$\Pi_{b,t}(x)) < \text{prob}_S(x$ will come to burn/$\Pi_{a,t}(x))$.

Notice that the reference to the time $t$ is essential here. We can make an object more flammable at one time than it was at another time by altering its nature, e.g., by transmuting its elements.

Principle 2.4 meets our previous difficulties, but now we encounter a new source of complexity in the notion of the degree to which an object has a capability. Consider deflatability. To say that one object is more deflatable than another would ordinarily be taken as meaning that the first object can be deflated more completely, not that it can be deflated more easily. Similarly, to say that one object is more fluorescent than another is to say that it fluoresces more, not that it fluoresces more easily.

What is happening here is that the notion of the degree to which an object has a capability is a two-dimensional notion. Consider flammability again. Suppose that when $x$ is heated to 100° centigrade it bursts into flame, but when $y$ is heated to 90° centigrade it just begins to smoulder and no matter how high the temperature of $y$ is raised it never does more than smoulder. Assuming that we count smouldering as the lower limit of burning, $x$ and $y$ are both flammable. There is a sense in which $y$ is more flammable than $x$ – it can be made to burn more easily. But there is also a clear sense in which $x$ is more flammable than $y$ – $y$ just barely burns whereas $x$ burns brilliantly.

For most capabilities we can talk about the extent to which an object realizes that capability on a particular occasion, e.g., the extent to which something burns, fluoresces, dissolves, crumbles, breaks, deflates; etc. How this notion is defined will depend upon what capability we are talking about. Let us suppose that we have a real valued function $E_{\varphi}(x)$ which measures the extent to which an object is $\varphi$-ing. Then in general, 'how $\varphi$-able an object is' cannot be given by any single measure. The degree of $\varphi$-ability of an object is best represented by the probability distribution

$$\text{prob}_S(\text{it will be the case that } E_{\varphi}(x) \geq r/\Pi_{a,t}(x)).$$

Given such a probability distribution we can generally make sense of statements comparing the extent to which two objects have a particular capability. For example, to say that one object $a$ is more fragile than another object $b$ generally means that it is easier to bring about a certain large degree of breaking in $a$ than in $b$, i.e., for some particular $r$,

$$\text{prob}_S(\text{it will be that case that } E_{\text{breaks}}(x) \geq r/\Pi_{b,t}(x)) < \text{prob}_S(\text{it will be the case that } E_{\text{breaks}}(x) \geq r/\Pi_{a,t}(x)).$$

On the other hand, to say that one object $a$ is more fluorescent than a second object $b$ is generally to say that a certain minimal 'amount of effort' will result in a greater degree of fluorescing in $a$ than in $b$, i.e., for some particular $p$,

$$\text{l.u.b. } \{r;\ \text{prob}_S(\text{it will be the case that } E_{\text{fluoresces}}(x) \geq r/\Pi_{b,t}(x)) \geq p\} < \text{l.u.b. } \{r;\ \text{prob}_S(\text{it will be the case that } E_{\text{fluoresces}}(x) \geq r/\Pi_{a,t}(x)) \geq p\}$$

In general, it seems that 'comparative capability statements' will be analyzable in one of these two ways, the correct analysis being determined by the context. For certain capabilities (e.g., fragility, fluorescence, deflatability) it will be more common for one of these analyses to be correct than for the other one to be correct, but in general either analysis could be correct and the context will determine which is appropriate.

It is generally claimed that absolute dispositions are dispositions to have certain 'manifest properties'. In light of our conclusions, this

should be rephrased by saying that capabilities are capabilities of having certain 'manifest properties'. This is probably all right if we do not bear down too heavily on 'manifest'. The properties involved in capabilities may be extremely complex, and may even be other capabilities. For example, the 'manifest property' involved in magnetizability is the property of being magnetic, but the property of being magnetic would itself seem to be a capability. It is also worth noting how 'non-manifest' the 'manifest properties' of fluorescing and dissolving are. We do not say that an object is fluorescing simply because it is glowing in the dark. E.g., a burning light bulb is glowing in the dark. Nor do we say that something is dissolving just because it is benignly disappearing in a liquid. For example, a submerged radioactive substance which is rapidly and spontaneously breaking up into elementary particles is not thereby dissolving. Fluorescing and dissolving are complex processes which only occur if certain microphysical processes occur. Thus one must not rely very heavily on this notion of a manifest property.

### 3. Probabilistic dispositions

Now let us turn to probabilistic dispositions. Except for their probabilistic aspect, they actually come closer to fitting the traditional analysis than do absolute dispositions. It is much easier to find examples of probabilistic dispositions than it is to find examples of capabilities. A random glance at the dictionary yielded the following list: 'abstemious', 'acquisitive', 'acrimonious', 'adamant', 'adaptable', 'affable', 'aggressive', 'courageous', 'cowardly', 'creative', 'credulous', 'critical', 'curious', 'inquisitive', 'deceitful', 'foolhardy', 'forgetful', 'forgiving'. Most of these terms are ambiguous between referring to enduring characteristics and (relatively) momentary states. For example, we can say either 'Joe is abstemious', or 'Joe is being abstemious'. The latter means simply that Joe is eating and drinking sparsely, and does not report a disposition. However, 'Joe is abstemious' reports a tendency in Joe to be abstemious. This tendency is what is being called 'a probabilistic disposition'. We make a similar distinction between 'Joe is acquisitive' and 'Joe is being acquisitive'; between 'Joe is acrimonious'

and 'Joe is being acrimonious'; between 'Joe is aggressive' and 'Joe is being aggressive'; etc. In each case, the former sentence reports a tendency to be in the state reported by the latter. It is the 'tendency' senses of these words which express probabilistic dispositions. It is worth noting that in some cases the momentary states reported by these words as used in their non-dispositional senses are themselves characterized dispositionally. For example, to say that Joe is pleasant is to say that Joe has a tendency to behave in a pleasant manner, and to say (non-dispositionally) that Joe is behaving in a pleasant manner on a particular occasion is to say that his behavior has the disposition to please people. Similar observations apply to 'abusive' and 'adamant'.

Statements of probabilistic dispositions report tendencies. In the common case of character traits these are tendencies to do certain things or behave in certain ways. For example, to say that a person is foolhardy is to say that he tends to act without sufficient attention to his personal safety. This is a probability statement. It amounts to saying that the probability of an action of this person being taken without sufficient attention to his personal safety is rather high. The probabilities in these statements are subjunctive probabilities, and hence are conditional probabilities. For example, for $S$ to be abstemious is for the probability of $S$'s eating and drinking sparsely on an eating and drinking occasion to be high; for $S$ to be acquisitive is for the probability of $S$'s trying to acquire a thing when it appeals to him to be high; for $S$ to be acrimonious is for the probability of $S$'s manner being bitter or harsh to be high; for $S$ to be courageous is for the probability of $S$'s actions to be taken without undue attention to personal safety to be high; for $S$ to be credulous is for the probability of $S$'s believing something he is told to be high; and so on. As conditional probability statements, these disposition statements have antecedents. Thus statements attributing probabilistic dispositions are much more like the traditional model of disposition statements than are statements attributing absolute dispositions or capabilities. The former have built-in antecedents whereas the latter do not.

We can talk about the degree to which a person (or object) has a probabilistic disposition, and just as in the case of capabilities, these degrees are two-dimensional. For example, in the case of foolhardiness we can talk both about how apt a person is to act in a foolhardy

manner, and about how foolhardy he is apt to act. In general, the extent to which a person is foolhardy is best measured by the distribution of the probabilities of his acting foolhardy to differing degrees. That is, if $\text{deg}_{\text{foolhardy}}(x)$ is the degree to which a particular act $x$ is foolhardy, then the measure of the extent to which a person $S$ has the disposition of foolhardiness is the distribution of the probabilities of an act $x$ of $S$ being such that $\text{deg}_{\text{foolhardy}}(x) \geq r$.

What kind of probability is involved here? It is a kind of subjunctive indefinite probability, but we have isolated several varieties of subjunctive indefinite probabilities. In judging whether a person is foolhardy, what we want to know is the probability of a new action of his being taken without sufficient attention to his personal safety. In computing this probability, we should take into account everything that is true about the person's actual situation. In other words, it is the simple subjunctive indefinite probability probe which is in question here. The degree to which a person is foolhardy is measured by the probability distribution:

$$\text{prob}_{+}(\text{deg}_{\text{foolhardy}}(x) \geq r / x \text{ is an action of } S).$$

It is worth pointing out that the use of subjunctive probabilities here is absolutely essential. It would be nonsensical to try to analyze probabilistic dispositions in terms of degree of rational belief or some other kind of indicative probability. What is at issue is not how likely (on the basis of incomplete knowledge) it is that something is now true of $S$'s actual actions, but rather how likely (on the basis of everything about $S$) it is that something would be true of non-actual actions of $S$. The tendency of philosophers to conflate different kinds of probability, no doubt fostered by their fear of subjunctive statements, has hopelessly muddled some earlier discussions of probabilistic tendencies.

This completes the list of subjunctive concepts to be discussed in this book. Analyses have been proposed for subjunctive generalizations, subjunctive conditionals, causal statements, statements of subjunctive probability, and disposition statements. If these analyses ultimately prove successful, this will free philosophers to use subjunctive notions in the analysis of other philosophically problematic concepts. In particular, I suspect that subjunctive concepts will prove of paramount importance in ethics. Judgments about moral responsibility and the

rightness and wrongness of actions involve in essential ways judgments about what would have resulted had persons acted in ways other than they did. I suspect that here and elsewhere, subjunctive concepts will prove to be indispensable tools for philosophical analysis.

## NOTES

[1] This implies, contrary to what many people have supposed, that the brick is not identical with the asbestos of which it is made. This is a conclusion I have defended at length in Pollock (1974), pp. 157–174. The brick is *composed* of the asbestos, but it is not identical with the asbestos.

# BIBLIOGRAPHY


Brody, Baruch, 1973, 'Why Settle for Anything Less than Good Old-Fashioned Aristotelian Essentialism?' *Nous* **7,** 351–365.

Carnap, Rudolph, 1950, *Logical Foundations of Probability*, University of Chicago Press.

Carnap, Rudolph, 1952, *The Continuum of Inductive Methods*, University of Chicago Press.

Carnap, Rudolph and Jeffrey, Richard C., 1971, *Studies in Inductive Logic and Probability*, vol. I, University of California Press.

Chisholm, Roderick, 1946, 'The Contrary-to-Fact Conditional', *Mind* **55,** 289–307.

Chisholm, Roderick, 1955, 'Law Statements and Counterfactual Inference', *Analysis* **15,** 97–105.

Chisholm, Roderick, 1967, 'Identity through Possible Worlds: Some Questions', *Nous* **1,** 1–8.

Davidson, Donald, 1967, 'Causal Relations', *Journal of Philosophy* **64,** 691–703.

Davidson, Donald, 1970, 'The Individuation of Events', in *Essays in Honor of Carl Hempel*, (ed. by Nicholas Rescher), Reidel, Dordrecht.

Ducasse, C. J., 1966, 'Critique of Hume's Conception of Causality', *Journal of Philosophy* **63,** 141–148.

Goodman, Nelson, 1955, *Fact, Fiction, and Forecast*, Harvard.

Jackson, Frank and Pargetter, Robert, 1973, 'Indefinite Probability Statements', *Synthese* **26,** 205–217.

Jeffrey, Richard and Carnap, Rudolph, 1971, *Studies in Inductive Logic and Probability*, vol. I, University of California Press.

Kaplan, David, 1964, *Foundations of Intensional Logic* (dissertation, UCLA 1964).

Kanger, Stig, 1957, 'The Morning Star Paradox', *Theoria* **23,** 1–11.

Kim, Jaegwon, 1970, 'Events and their Descriptions: Some Considerations', in *Essays in Honor of Carl Hempel*, (ed. by Nicholas Rescher), Reidel, Dordrecht.

Kim, Jaegwon, 1971, 'Causes and Events: Mackie on Causation', *Journal of Philosophy* **68,** 426–441.

Kim, Jaegwon, 1973, 'Causation, Nomic Subsumption, and the Concept of an Event', *Journal of Philosophy* **70,** 217–236.

Kim, Jaegwon, 1973a, 'Causes and Counterfactuals', *Journal of Philosophy* **70,** 570–572.

Kripke, Saul, 1971, 'Identity and Necessity', in *Identity and Individuation*, (ed. by Milton Munitz), New York University Press.

Kripke, Saul, 1971, 'Naming and Necessity', in *Semantics of Natural Language*, (ed. by Donald Davidson and Gilvert Harman), Reidel, Dordrecht.

Kyburg, Henry E., Jr., 1974, *The Logical Foundations of Statistical Inference*, Reidel, Dordrecht.

Lewis, David, 1968, 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* **65,** 113–126.

Lewis, David, 1972, 'Completeness and Decidability of Three Logics of Counterfactual Conditionals', *Theoria* **37,** 74–85.

Lewis, Dávid, 1973, *Counterfactuals*, Harvard.
Lewis, David, 1973a, 'Counterfactuals and Comparative Possibility', *Journal of Philosophical Logic* **2,** 418–446.
Lewis, David, 1973b, 'Causation', *Journal of Philosophy* **90,** 556–567.
Mackie, J. L., 1965, 'Causes and Conditions', *American Philosophical Quarterly* **2,** 245–264.
Pargetter, Robert and Jackson, Frank, 1973, 'Indefinite Probability Statements', *Synthese* **26,** 205–217.
Pollock, John L., 1967, 'Logical Validity in Modal Logic', *The Monist,* **51,** 128–135.
Pollock, John L., 1972, The Logic of Projectibility', *Philosophy of Science* **39,** 302–314.
Pollock, John L., 1973, 'Laying the Raven to Rest', *Journal of Philosophy* **70,** 747–754.
Pollock, John L., 1974, *Knowledge and Justification*, Princeton.
Pollock, John L., 1974a, 'Subjunctive Generalizations', *Synthese* **28,** 199–214.
Pollock, John L., 1975, 'Four Kinds of Conditionals', *American Philosophical Quarterly* **12,** 51–60.
Popper, Karl R., 1955, 'Two Autonomous Axiom Systems for the Calculus of Probabilities', *British Journal for the Philosophy of Science* **5,** 51–57.
Popper, Karl R., 1959, *The Logic of Scientific Discovery*, Basic Books.
Scriven, Michael, 1964, 'The Structure of Science', *Review of Metaphysics* **17,** 403–424.
Stalnaker, Robert C., 1968, 'A Theory of Conditionals', *American Philosophical Quarterly*, monograph series **2,** 98–112.
Stalnaker, Robert C., 1972, 'Pragmatics', *Semantics of Natural Language* (ed. by Donald Davidson and Gilbert Harman), Reidel, Dordrecht.
Stalnaker, Robert C. and Thomason, Richmond H., 1970, 'A Semantic Analysis of Conditional Logic'. *Theoria* **36,** 23–42.
Thomason, Richard H. and Stalnaker, Robert C., 1970, 'A Semantic Analysis of Conditional Logic', *Theoria* **36,** 23–42.
Vendler, Zeno, 1967, 'Causal Relations', *Journal of Philosophy* **64,** 704–713.
Vendler, Zeno, 1967a, *Linguistics in Philosophy*, Cornell, Ithaca.

# INDEX

# PHILOSOPHICAL STUDIES SERIES IN PHILOSOPHY



1. JAY F. ROSENBERG, *Linguistic Representation.* 1974, xii + 159 pp.
2. WILFRID SELLARS, *Essays in Philosophy and Its History.* 1974, xiii + 462 pp.
3. DICKINSON S. MILLER, *Philosophical Analysis and Human Welfare.* Selected Essays and Chapters from Six Decades. Edited with an Introduction by Loyd D. Easton. 1975, x + 333 pp.
4. KEITH LEHRER (ed.), *Analysis and Metaphysics.* Essays in Honor of R. M. Chisholm. 1975, x + 317 pp.
5. CARL GINET, *Knowledge, Perception, and Memory.* 1975, viii + 212 pp.
6. PETER H. HARE and EDWARD H. MADDEN, *Causing, Perceiving and Believing.* An Examination of the Philosophy of C. J. Ducasse. 1975, vii + 211 pp.
7. HECTOR-NERI CASTAÑEDA, *Thinking and Doing.* The Philosophical Foundations of Institutions. 1975, xviii + 366 pp.