Knowledge and Conditionals

Essays on the Structure of Inquiry

Robert C. Stalnaker

OXFORD
UNIVERSITY PRESS

9

A Theory of Conditionals

1. Introduction

A conditional sentence expresses a proposition that is a function of two other propositions, yet not one that is a *truth* function of those propositions. I may know the truth values of "Willie Mays played in the American League" and "Willie Mays hit four hundred" without knowing whether or not Mays would have hit four hundred if he had played in the American League. This fact has tended to puzzle, displease, or delight philosophers, and many have felt that it is a fact that calls for some comment or explanation. It has given rise to a number of philosophical problems; I will discuss three of these.[1]

Willie Mays: *a baseball star of the 1950s and 60s, who played only in the National League. To "hit four hundred" is to bat .400 over a season, which he never did.*

some comment or explanation: *the puzzle: the clauses' truth values don't settle the conditional, yet it's true or false. So something more enters its truth condition: a selected world, the paper will argue, p. 155.*

My principal concern will be with what has been called the *logical problem of conditionals*, a problem that is frequently ignored or dismissed by writers on conditionals and counterfactuals. This is the task of describing the formal properties of the *conditional function:* a function, usually represented in English by the words "if… then," taking ordered pairs of propositions into propositions. I will explain informally and defend a solution, presented more rigorously elsewhere, to this problem.[2]

The second issue—the one that has dominated recent discussions of contrary-to-fact conditionals—is the *pragmatic problem of counterfactuals*. This problem derives from the belief, which I share with most philosophers writing about this topic, that the formal properties of the conditional function, together with all of the *facts,* may not be sufficient for determining the truth value of a counterfactual; that is, different truth valuations of conditional statements may be consistent with a single valuation of all non-conditional statements. The task set by the problem is to find and defend criteria for choosing among these different valuations. This problem is different from the first issue because these criteria are pragmatic, and not semantic. The distinction between semantic and pragmatic criteria, however, depends on the construction of a semantic theory. The semantic theory that I shall defend will thus help to clarify the second problem by charting the boundary between the semantic and pragmatic components of the concept. The question of this boundary line is precisely what Rescher, for example, avoids by couching his whole discussion in terms of conditions for belief, or justified belief, rather than truth conditions.[3] Conditions for justified belief are pragmatic for any concept.

together with all of the facts: *fix every plain fact and it can still be open whether the match would have lit had it been struck. Something besides the facts, criteria for what to hold fixed, must decide.*

pragmatic, and not semantic: *semantic criteria come from the meaning of "if" and the facts alone, pragmatic ones from the context and purpose of the utterance.*

[1] This chapter, first published fifty years ago, has been lightly edited for style and clarity. I have left the content unchanged, but added several footnotes in addition to this one.

[2] Stalnaker & Thomason 1970. In this paper, the formal system, C2, is proved sound and semantically complete with respect to the interpretation sketched in the present paper. That is, it is shown that a formula is a consequence of a class of formulas if and only if it is derivable from the class in the formal system, C2.

[3] Rescher 1964.

The third issue is an epistemological problem that has bothered empiricist philosophers. It is based on the fact that many counterfactuals seem to be synthetic and contingent statements about unrealized possibilities. But contingent statements must be capable of confirmation by empirical evidence, and the investigator can gather evidence only in the actual world. How are conditionals that are both empirical and contrary-to-fact possible at all? How do we learn about possible worlds, and where are the facts (or counterfacts) that make counterfactuals true? Such questions have led philosophers to try to analyze the conditional in non-conditional terms[4]—to show that conditionals merely appear to be about unrealized possibilities. My approach, however, will be to accept the appearance as reality, and to argue that one can sometimes have evidence about non-actual situations.

both empirical and contrary-to-fact: *a triad that looks inconsistent: contingent claims need evidence, evidence comes only from the actual world, counterfactuals are about non-actual ones. Section 6 shows how all three can stand.*

In sections 2 and 3 of this chapter, I will present and defend a theory of conditionals that has two parts: a formal system with a primitive conditional connective, and a semantic apparatus that provides general truth conditions for statements involving that connective. In sections 4, 5, and 6, I will discuss in a general way the relation of the theory to the three problems outlined above.

2. The Interpretation

Eventually, I want to defend a hypothesis about the truth conditions for statements having conditional form, but I will begin by asking a more practical question: How does one evaluate a conditional statement? How does someone decide whether or not he or she believes it to be true? An answer to this question will not be a set of truth conditions, but it will serve as a heuristic aid in the search for such a set.

a heuristic aid: *the method of section 2: how we decide to accept a conditional isn't its truth condition, but the right truth condition should explain why that way of deciding works, p. 154.*

To make the question more concrete, consider the following situation: you are faced with a true-false political opinion survey.[5] The statement is: "If the Chinese enter the Vietnam conflict, the United States will use nuclear weapons." How do you deliberate in choosing your response? What considerations of a logical sort are relevant? I shall first discuss two familiar answers to this question, and then defend a third answer that avoids some of the weaknesses of the first two.

[4] Cf. Chisholm 1946. The problem is sometimes posed (as it is here) as the task of analyzing the *subjunctive* conditional into an indicative statement, but I think it is a mistake to base very much on the distinction of mood. As far as I can tell, the mood tends to indicate something about the attitude of the speaker, but in no way affects the propositional content of the statement.

the distinction of mood: *the grammatical difference between the subjunctive "if he had played, he would have hit" and the indicative "if he played, he hit".*

[5] These motivating remarks are indebted to Saul Kripke, who used the device of a true-false opinion survey, in the discussion following a colloquium talk, to give a counterexample to a defense of the material conditional analysis of the indicative conditional. The talk (given in 1965) was by James Thomson, and Thomson acknowledged in discussion that the counterexample was decisive, which is probably why his paper was not published in his lifetime. (The paper was published years later: Thomson 1990.) Thomson argued, using Gricean considerations, that conditionals are true when both antecedent and consequent are true even when the two propositions are unrelated, and I thought these arguments were entirely persuasive. But I also agreed that Kripke's counterexample, which concerned a conditional with a false antecedent, was decisive against the material conditional analysis, which of course cannot work for counterfactuals in any case. So, when I turned my attention to the problem, I was looking for an analysis that was stronger than the truth-functional analysis, but that could still deliver the result that conditionals could be true even when the antecedent and consequent were unrelated. [Note added in 2018]

Gricean considerations: *Grice's rules of conversation. A conditional with unrelated clauses is odd to assert, but on this view the oddity is in the saying, not in its truth.*

The first answer is based on the simplest account of the conditional: the truth-functional analysis. According to this account, you should reason as follows in responding to the true-false quiz: you ask yourself, first, will the Chinese enter the conflict? Second, will the United States use nuclear weapons? If the answer to the first question is no, *or* if the answer to the second is yes, then you should place your *X* in the "true" box. But this account is unacceptable since the following piece of reasoning is an obvious *non sequitur:* "I firmly believe that the Chinese will stay out of the conflict; *therefore*, I believe that the statement is true." The falsity of the antecedent is never sufficient reason to affirm a conditional, even an indicative conditional.

The falsity of the antecedent: *truth-functionally "if A, B" is "not-A or B", so every counterfactual, its A false, would be true. Note 5 takes that as settled. Kripke's survey, used here, shows it fails even for indicatives.*

A second answer is suggested by the shortcomings of the truth-functional account. The material implication analysis fails, critics have said, because it leaves out the idea of *connection* that is implicit in an if-then statement. According to this line of thought, a conditional is to be understood as a statement that affirms that some sort of logical or causal connection holds between the antecedent and the consequent. In responding to the true-false quiz, then, you should look, not at the truth-values of the two clauses, but at the relation between the propositions expressed by them. If the "connection" holds, you check the "true" box. If not, you answer "false."

If the second hypothesis were accepted, then we would face the task of clarifying the idea of "connection," but there are counter-examples even with this notion left as obscure as it is. Consider the following case: you firmly believe that the use of nuclear weapons by the United States in this war is inevitable because of the arrogance of power, the bellicosity of our president, growing pressure from congressional hawks, or other *domestic* causes. You have no opinion about future Chinese actions, but you do not think they will make much difference one way or another to nuclear escalation. Clearly, you believe the opinion survey statement to be true even though you believe the antecedent and consequent to be logically and causally independent of each other. It seems that the presence of a "connection" is not a necessary condition for the truth of an if-then statement.

logically and causally independent: *the case isolates B: your grounds for it have nothing to do with A, yet you'd tick "true". That is the result note 5 says the theory was built to keep, truth for unrelated clauses.*

The third answer I shall consider is based on a suggestion made some time ago by F. P. Ramsey.[6] Consider first the case where you have no opinion about the statement, "The Chinese will enter the Vietnam war." According to your suggestion, your deliberation about the survey statement should consist of a simple thought experiment: add the antecedent (hypothetically) to your stock of knowledge (or beliefs), and then consider whether or not the consequent is true. Your belief about the conditional should be the same as your hypothetical belief, under this condition, about the consequent.

add the antecedent (hypothetically): *Ramsey's test. Ask not whether A is true or causes B, but whether you'd believe B having supposed A. Belief in "if A, B" is belief in B under that supposition.*

What happens to the idea of connection on this hypothesis? It is sometimes relevant to the evaluation of a conditional, and sometimes not. If you believe that a causal or logical connection exists, then you will add the consequent to your stock of beliefs along with the antecedent, since the rational man accepts the consequences of his beliefs. On the other hand, if you already believe the consequent (and if you also believe it to be causally independent of the antecedent), then it will remain a part of

a causal or logical connection exists: *two rules of rational belief: accept what your beliefs entail, change none without reason. The second returns for worlds, p. 156: the selected world differs only where A requires.*

[6] Ramsey 1929. The suggestion is made on p. 248. Chisholm quotes the suggestion and discusses the limitations of the 'connection' thesis that it brings out (Chisholm 1946), but he develops it somewhat differently.

your stock of beliefs when you add the antecedent, since the rational man does not change his beliefs without reason. In either case, you will affirm the conditional. Thus, this answer accounts for the relevance of "connection" when it is relevant, without making it a necessary condition of the truth of a conditional.

Ramsey's suggestion covers only the situation in which you have no opinion about the truth-value of the antecedent. Can it be generalized? We can of course extend it without problem to the case where you believe or know the antecedent to be true; in this case, no changes need be made in your stock of beliefs. If you already believe that the Chinese will enter the Vietnam conflict, then your belief about the conditional will be just the same as your belief about the statement that the United States will use the bomb.

What about the case in which you know or believe the antecedent to be false? In this situation, you cannot simply add it to your stock of beliefs without introducing a contradiction. You must make adjustments by deleting or changing those beliefs which conflict with the antecedent. Here, the familiar difficulties begin, of course, because there will be more than one way to make the required adjustments.[7] These difficulties point to the pragmatic problem of counterfactuals, but if we set them aside for a moment, we will see a rough but general answer to the question we are asking. This is how to evaluate a conditional.

more than one way: *which beliefs to give up when the antecedent contradicts them. Goodman's problem of relevant conditions, from "Counterfactual conditionals".*

First, add the antecedent: *skirting the adjustments is harmless, since they return as the choice of a selection function, p. 157, a pragmatic matter. Tying truth to someone's beliefs isn't, and avoiding it leads to worlds.*

First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true. It is not particularly important that our answer is approximate—that it skirts the problem of adjustments—since we are using it only as a way of finding truth conditions. It is crucial, however, that the answer should not be restricted to some particular context of belief if it is to be helpful in finding a definition of the conditional function. If the conditional is to be understood as a function of the propositions expressed by its component clauses, then its truth-value should not in general be dependent on the attitudes that anyone has toward those propositions.

Now that we have found an answer to the question "How do we decide whether or not we believe a conditional statement?", the problem is to make the transition from belief conditions to truth conditions; that is, to find a set of truth conditions for statements having conditional form which explains why we use the method we do use to evaluate them. The concept of a *possible world* is just what we need to make this transition, since a possible world is the ontological analogue of a stock of hypothetical beliefs. The following set of truth conditions, using this notion, is a first approximation to the account that I shall propose:

the ontological analogue: *on the side of what there is, not of what anyone believes. A world is like a stock of hypothetical beliefs, but complete and nobody's.*

> Consider a possible world in which *A* is true, and which otherwise differs minimally from the actual world. *"If A, then B" is true (false) just in case B is true (false) in that possible world.*

differs minimally: *one world is picked, the nearest, not a set of equally near worlds. This uniqueness assumption is where Lewis will part company.*

[7] Rescher 1964, 11–16, contains a very clear statement and discussion of this problem, which he calls the problem of the ambiguity of belief-contravening hypotheses. He argues that the resolution of this ambiguity depends on pragmatic considerations. Cf. also Goodman's problem of relevant conditions, Goodman 1983, 17–24.

An analysis in terms of possible worlds also has the advantage of providing a ready-made apparatus on which to build a formal semantic theory. In making this account of the conditional precise, we use the semantic models for modal logics developed by Saul Kripke.[8] Following Kripke, we first define a *model structure.* Let *M* be an ordered triple ⟨*K,R,*λ*).* *K* is to be understood intuitively as the set of all possible worlds; *R* is the relation of relative possibility that defines the structure. If α and β are possible worlds (members of *K*), then αRβ reads, "β is possible with respect to α." This means that, where α is the actual world, β is a possible world. *R* is a reflexive relation; that is, every world is possible with respect to itself. If your modal intuitions so incline you, you may add that *R* must be transitive, or transitive and symmetrical.[9] The only element that is not a part of the standard modal semantics is λ, a member of *K* that is to be understood as the *absurd world*—the world in which contradictions and all their consequences are true. It is an isolated element under *R;* that is, no other world is possible with respect to it, and it is not possible with respect to any other world. The purpose of A is to allow for an interpretation of "If *A,* then *B*" in the case where *A* is impossible; for this situation one needs an impossible world.[10]

the absurd world: *a device, not a way things could be. Stalnaker's 2018 note 10 says it wasn't meant seriously: a conditional with no world eligible for selection can simply be true.*

In addition to a model structure, our semantic apparatus includes a *selection function, f,* which takes a proposition and a possible world as arguments and a possible world as its value. The s-function selects, for each antecedent *A,* a particular possible world in which *A* is true. The assertion that the conditional makes, then, is that the consequent is true in the world selected. A conditional is true in the actual world when its consequent is true in the selected world.

a particular possible world: *one world, not all the A-worlds, as strict implication has it, nor just some, which gives only a "might". That single world is why, for a possible A, "if A, B" or "if A, not-B" holds, p. 158.*

Now we can state the semantic rule for the conditional more formally (using the corner, >, as the conditional connective):

A > *B* is true in *a* if *B is* true in *f*(*A*,*a*);
A > *B is* false in *a* if *B is* false in *f(A,a).*

The interpretation shows conditional logic to be an extension of modal logic. Modal logic provides a way of talking about what is true in the actual world, in all possible worlds, or in at least one, unspecified world. The addition of the selection function to the semantics and the conditional connective to the object language of modal logic provides a way of talking also about what is true in *particular* non-actual possible situations. This is what counterfactuals are: statements about particular counterfactual worlds.

This is what counterfactuals are: *the paper's thesis. Modal logic speaks of all worlds or some, the conditional of one in particular, so a counterfactual is as definite a claim as one about the actual world, only about another.*

But the world selected cannot be just any world. The s-function must meet four conditions. (I will use the following terminology for talking about the arguments and values of s-functions: where *f*(*A*,α) = β, *A* is the *antecedent*, α is the *base world*, and β is the *selected world.*)

[8] Kripke 1963.

[9] The different restrictions on the relation *R* provide interpretations for the different modal systems. The system we build on is von Wright's M. If we add the transitivity requirement, then the underlying modal logic of our system is Lewis's S4. and if we add both the transitivity and symmetry requirements, then the modal logic is *SS.* Cf. Kripke, op. cit.

[10] The absurd world was not meant to be taken seriously. The analysis could be stated without it, simply by stipulating that the conditional is true when there is no world eligible to be selected. [Note added in 2018].

(1) For all antecedents A and base worlds α, A must be true in $f(A,\alpha)$.
(2) For all antecedents A and base worlds α, $f(A,\alpha) = \lambda$ only if there is no world possible with respect to α in which A is true.[11]

The first condition requires that the antecedent be true in the selected world. This ensures that all statements like "if snow is white, then snow is white" are true. The second condition requires that the absurd world be selected only when the antecedent is impossible. Since everything is true in the absurd world, including contradictions, if the selection function were to choose it for the antecedent A, then "If A, then B and not B" would be true. But one cannot legitimately reach an impossible conclusion from a consistent assumption.

The informal truth conditions that were suggested above required that the world selected *differ minimally* from the actual world. This implies, first, that there are no differences between the actual world and the selected world except those that are required, implicitly or explicitly, by the antecedent. Further, it means that among the alternative ways of making the required changes, one must choose one that does the least violence to the correct description and explanation of the actual world. These are vague conditions that are largely dependent on pragmatic considerations for their application. They suggest, however, that the selection is based on an ordering of possible worlds with respect to their resemblance to the base world. If this is correct, then there are two further formal constraints that must be imposed on the s-function:

the least violence: *so for "if this match were struck" keep the oxygen, the dry head and the laws, and change only the hand. A world where it's struck but chemistry differs does more violence.*

(3) For all base worlds α and all antecedents A, if A is true in α then $f(A,\alpha) = \alpha$.
(4) For all base worlds a and all antecedents B and B', if B is true in $f(B',\alpha)$ and B' is true in $f(B,\alpha)$, then $f(B,\alpha) = f(B',\alpha.)$.

(4): *in words: if the world selected for B′ has B and the world selected for B has B′, they're one world. Otherwise each would be ranked nearer to α than the other.*

The third condition requires that the base world be selected if it is among the worlds in which the antecedent is true. Whatever the criteria for evaluating resemblance among possible worlds, there is obviously no other possible world as much like the base world as the base world itself. The fourth condition ensures that the ordering among possible worlds is consistent in the following sense: if any selection established β as prior to β' in the ordering (with respect to a particular base world α), then no other selection (relative to that α) may establish β' as prior to β.[12] Conditions (3) and (4) together ensure that the s-function establishes a total ordering of all

the base world itself: *the upshot of (3): if A and B are both true, so is "if A, B", however unrelated they are. That's the Ramsey test's case of a believed A, p. 154, and the result note 5 says he wanted.*

[11] Shouldn't it be "if and only if" rather than just "only if"? I had thought that the "if" part was a consequence of the other conditions, but closer examination reveals an anomaly in the semantics that went unnoticed (at least by me) until Matt Mandelkern pointed it out recently to me. The original idea was to start with a Kripke model, with its interpretation of the necessity and possibility operators, and then add some additional structure with the selection function. But then it was observed that the necessity and possibility operators could be defined in terms of the conditional, so there is no separate semantic rule for the standard modal operators. The accessibility relation becomes an idle wheel in the semantics. But we still want to show that the standard modal operators have their standard interpretation, and to do this we need to show that the Kripkean semantic rules for the standard operators could be derived from the semantic rules for the conditional connective. Condition (2), strengthened to "if and only if" and now thought of as a constraint on R rather than a constraint on the selection function, will allow us to show that the defined "□ϕ" is true in world x iff ϕ is true in all β such that αRβ. (Thanks to Matt Mandelkern for his observation, and discussion of it.) [ote added in 2018.]

[12] If f(A,α) = β, then β is established as prior to all worlds possible with respect to α in which A is true.

selected worlds with respect to each possible world, with the base world preceding all others in the order.

These conditions on the selection function are necessary in order that this account be recognizable as an explication of the conditional, but they are of course far from sufficient to determine the function uniquely. There may be further formal constraints that can plausibly be imposed on the selection principle, but we should not expect to find semantic conditions sufficient to guarantee that there will be a unique s-function for each valuation of non-conditional formulas on a model structure. The questions, "On what basis do we select a selection function from among the acceptable ones?" and "'What are the criteria for ordering possible worlds?" are reformulations of the pragmatic problem of counterfactuals, which is a problem in the application of conditional logic. The conditions that I have mentioned above are sufficient, however, to define the semantic notions of validity and consequence for conditional logic.

far from sufficient: *so C2's valid inferences hold however worlds are ranked, but whether the match would have lit depends on the ranking. Goodman's question, what to hold fixed, survives as the choice of f.*

3. The Formal System

The class of valid formulas of conditional logic, according to the definitions sketched in the preceding section, is coextensive with the class of theorems of a formal system, C2. The primitive connectives of C2 are the usual "⊃" and "~"{with "∨," "&," and "≡" defined as usual), as well as a conditional connective, ">" (called the corner). The necessity and possibility operators, and a bi-conditional operator can be defined in terms of the corner as follows:

The primitive connectives of C2: ⊃ *the material conditional,* ~ *not,* ∨ *or,* & *and,* ≡ *material equivalence,* > *the corner,* =*df* *"defined as",* □ *necessarily,* ◇ *possibly.*

$\Box A =_{df} (\sim A > A)$
$\Diamond A =_{df} \sim(A > \sim A)$
$(A \nsim B) =_{df} (A > B)\&(B>A)$

□A =df (~A > A): *A is necessary when even supposing not-A leads to A, which happens only when not-A is impossible. A is possible when supposing it doesn't lead to not-A.*

The rules of inference of C2 are *modus ponens* (if A and $A \supset B$ are theorems, then B is a theorem) and the rule of necessitation (If A is a theorem, then $\Box A$ is a theorem). There are seven axiom schemata:

(al) Any tautologous sentence is an axiom
(a2) $\Box(A \supset B) \supset (\Box A \supset \Box B)$
(a3) $\Box(A \supset B) \supset (A > B)$
(a4) $\Diamond A \supset ((A > B) \supset \sim(A >\sim B))$
(a5) $(A > (B \vee C)) \supset ((A > B) \vee (A > C))$
(a6) $(A > B) \supset (A \supset B)$
(a7) $((A \nsim B) \supset ((A > C) \supset (B > C))$

(a4), (a5), (a7): *the corner's own, in order: a possible A can't lead to both B and not-B. What leads to B or C leads to one of them. A and B leading to each other lead to the same things.*

The conditional connective, as characterized by this formal system, is intermediate between strict implication and the material conditional in the sense that $\Box(A \supset B)$ entails $(A > B)$ (by (a3)) and $(A > B)$ entails $(A \supset B)$ (by a6). It cannot, however, be analyzed as a modal operation performed on a material conditional (like Burks's causal implication, for example).[13] The corner lacks certain properties shared by the

strict implication: □*(A ⊃ B), true when no possible world has A true and B false. The corner sits between it and* ⊃*, yet unlike both it neither chains nor contraposes, p. 158.*

[13] Burks 1951. The causal implication connective characterized in this article has the same structure as strict implication. For an interesting philosophical defense of this modal interpretation of conditionals, see Mayo 1957.

two traditional implication concepts, and in fact these differences help to explain some peculiarities of counterfactuals. I shall point out three unusual features of the conditional connective.

nontransitive: *the Hoover case: the nearest world where he's born Russian has him a Communist but no traitor. The first premise is about another world, where he's an American Communist.*

(1) Unlike both material and strict implication, the conditional comer is a non-transitive connective. That is, from $A > B$ and $B > C$, one cannot infer $A > C$. While this may at first seem surprising, consider the following example: *Premises.* "If J. Edgar Hoover were today a Communist, then he would be a traitor." "If J. Edgar Hoover had been born a Russian, then he would today be a Communist." *Conclusion.* "If J. Edgar Hoover had been born a Russian, he would be a traitor." It seems reasonable to affirm these premises and deny the conclusion.

the following rule follows: *because $(A \,\&\, C) > A$ always holds, transitivity would turn $A > B$ into $(A \,\&\, C) > B$. The soaked match shows that rule false, so transitivity goes with it.*

If this example is not sufficiently compelling, note that the following rule follows from the transitivity rule: From $A > B$ to infer $(A\&C) > B$. But it is obvious that the former rule is invalid; we cannot always strengthen the antecedent of a true conditional and have it remain true. Consider "If this match were struck, it would light," and "If this match had been soaked in water overnight *and* it were struck, it would light."[14]

the denial of a conditional: *a possible A gets one ordinary world, which has B or lacks it, so denying "if A, B" is asserting "if A, not-B". An impossible A gets the absurd world, where both hold.*

(2) According to the formal system, the denial of a conditional is equivalent to a conditional with the same antecedent and opposite consequent (provided that the antecedent is not impossible). That is, $\Diamond A \supset ((A > B) \equiv \sim(A > \sim B))$. This explains the fact, noted by both Goodman and Chisholm in their early papers on counterfactuals, that the normal way to contradict a counterfactual is to contradict the consequent, keeping the same antecedent. To deny "If Kennedy were alive today, we wouldn't be in this Vietnam mess," we say, "If Kennedy were alive today, we would so be in this Vietnam mess."

~B > ~A is false: *different antecedents pick different worlds. The nearest halt-world has no talks, but the nearest talks-world is further off: there the bombing stopped and much more changed besides.*

(3) The inference of contraposition, valid for both the truth-functional horseshoe and the strict implication hook, is invalid for the conditional corner. $A > B$ may be true while $\sim B > \sim A$ is false. For an example in support of this conclusion, we take another item from the political opinion survey: "If the U.S. halts the bombing, then North Vietnam will not agree to negotiate." A person would believe that this statement is true if he thought that the North Vietnamese were determined to press for a complete withdrawal of US troops. But he would surely deny the contrapositive, "If North Vietnam agrees to negotiate, then the US will not have halted the bombing." He would believe that a halt in the bombing, and much more, is required to bring the North Vietnamese to the negotiating table.[15]

How Goodman and Chisholm kept the connection theory despite (2) and (3): by paraphrasing semifactuals. The reply, on p. 159, is that this is ad hoc.

Examples of these anomalies have been noted by philosophers in the past. For instance, Goodman pointed out that two counterfactuals with the same antecedent and contradictory consequents are "normally meant" as direct negations of each other. He also remarked that we may sometimes assert a conditional and yet reject its contrapositive. He accounted for these facts by arguing that semifactuals

[14] Although the transitivity inference fails, a related inference is of course valid. From $A > B$, $B > C$. and A, one can infer C. Also, note that the bi-conditional connective is transitive. From $A \asymp B$ and $B \asymp C$, one can infer $A \asymp C$. Thus the bi-conditional relation is an equivalence relation since it is also symmetric and reflexive.

[15] Although contraposition fails, *modus tollens* is valid for the conditional: from $A > B$ and $\sim B$, one can infer $\sim A$.

(conditionals with false antecedents and true consequents) are for the most part not to be taken literally. "In practice," he wrote, "full counterfactuals affirm, while semifactuals deny, that a certain connection obtains between antecedent and consequent.... The practical import of a semifactual is thus different from its literal import.'[16] Chisholm also suggested paraphrasing semifactuals before analyzing them. "Even if you were to sleep all morning, you would be tired" is to be read, "It is false that if you were to sleep all morning, you would not be tired."[17]

A separate and non-conditional analysis for semifactuals is necessary to save the "connection" theory of counterfactuals in the face of the anomalies we have discussed, but it is a baldly ad hoc manoeuver. Any analysis can be saved by paraphrasing the counterexamples. The theory presented in section 2 avoids this difficulty by denying that the conditional can be said, in general, to assert a connection of any particular kind between antecedent and consequent. It is, of course, the structure of inductive relations and causal connections that makes counterfactuals and semifactuals true or false, but it does this by determining the relationships among possible worlds, which in turn determine the truth-values of conditionals. By treating the relation between connection and conditionals as an indirect relation in this way, the theory is able to give a unified account of conditionals that explains the variations in their behavior in different contexts.

but it does this: *laws and causes make conditionals true only by ranking worlds. So Chisholm's "even if you were to sleep all morning" needs no paraphrase: it's true if the nearest such world still has you tired.*

4. The Logical Problem: General Considerations

The traditional strategy for attacking a problem like the logical problem of conditionals was to find an *analysis*: to show that the unclear or objectionable phrase was dispensable, or replaceable by something clear and harmless. Analysis was viewed by some as an *unpacking*—a making manifest of what was latent in the concept; by others it was seen as the *replacement* of a vague idea by a precise one, adequate to the same purposes as the old expression, but free of its problems. The semantic theory of conditionals can also be viewed either as the construction of a concept to replace an unclear notion of ordinary language, or as an *explanation* of a commonly used concept. I see the theory in the latter way: no recommendation or stipulation is intended. This does not imply, however, that the theory is meant as a description of linguistic usage. What is being explained is not the rules governing the use of an English word, but the structure of a concept. Linguistic facts—what we would say in this or that context, and what sounds odd to the native speaker—are relevant as evidence, since one may presume that concepts are to some extent mirrored in language.

A pause on method: the theory explains a concept we have, so what we'd say bears on it and each ad hoc plea costs. Hence one meaning counts in its favour, p. 160.

the replacement of a vague idea: *Carnap's explication (Logical Foundations of Probability, 1950) puts a precise concept in place of a vague one. The "explication" of p. 157 is the other kind, an explanation of the concept we have.*

The "facts," taken singly, need not be decisive. A recalcitrant counterexample may be judged a deviant use or a different sense of the word. We can claim that a paraphrase is necessary, or even that ordinary language is systematically mistaken about the concept we are explaining. There are, of course, different senses and times when "ordinary language" goes astray, but such ad hoc hypotheses and qualifications diminish both the plausibility and the explanatory force of a theory. While we are not irrevocably bound to the linguistic facts, there are no "don't cares"—contexts of use

[16] Goodman 1955, 15, 32. [17] Chisholm 1946, 492.

with which we are not concerned—since any context can be relevant as evidence for or against an analysis. A general interpretation which avoids dividing senses and accounts for the behavior of a concept in many contexts fits the familiar pattern of scientific explanation in which diverse, seemingly unlike surface phenomena are seen as deriving from some common source. For these reasons, I take it as a strong point in favor of the semantic theory that it treats the conditional as a univocal concept.

scientific explanation: *many surface phenomena traced to one source, as one law covers apples and planets. One meaning of "if" explaining all its uses beats a list of senses, so being univocal, one-sensed, is a strength.*

5. Pragmatic Ambiguity

I have argued that the conditional connective is semantically unambiguous. It is obvious, however, that the context of utterance, the purpose of the assertion, and the beliefs of the speaker or his community may make a difference to the interpretation of a counterfactual. How do we reconcile the ambiguity of conditional sentences with the univocality of the conditional concept? Let us look more closely at the notion of ambiguity.

A sentence is ambiguous if there is more than one proposition that it may properly be interpreted to express. Ambiguity may be syntactic (if the sentence has more than one grammatical structure), semantic (if one of the words has more than one meaning), or pragmatic (if the interpretation depends directly on the context of use). The first two kinds of ambiguity are perhaps more familiar, but the third kind is probably the most common in natural languages. Any sentence involving pronouns, tensed verbs, articles, or quantifiers is pragmatically ambiguous. For example, the proposition expressed by "L'etat, c'est moi" depends on who says it; "Do it now" may be good or bad advice depending on when it is said; "Cherchez la femme" is ambiguous since it contains a definite description; and the truth conditions for "All's well that ends well" depend on the domain of discourse. If the theory presented above is correct, then we may add conditional sentences to this list. The truth conditions for "If wishes were horses, then beggars would ride" depend on the specification of an s-function.[18]

"L'etat, c'est moi": *"I am the state", Louis XIV. "Cherchez la femme", "look for the woman", picks out a different woman each time it's said.*

The grounds for treating the ambiguity of conditional sentences as pragmatic rather than semantic are the same as the grounds for treating the ambiguity of quantified sentences as pragmatic: simplicity and systematic coherence. The truth-conditions for quantified statements vary with a change in the domain of discourse, but there is a single structure to these truth conditions that remains constant for every domain. The semantics for classical predicate logic brings out this common structure by giving the universal quantifier a single meaning and making the domain a parameter of the interpretation. In a similar fashion, the semantics for conditional logic brings out the common structure of the truth-conditions for conditional statements by giving the connective a single meaning and making the selection function a parameter of the interpretation.

a parameter of the interpretation: *"everyone came" has one meaning, yet is true or false only once context fixes who's counted. Likewise "if" has one meaning, and context fixes f, which worlds count as nearest.*

[18] I do not wish to pretend that the notions needed to define ambiguity and to make the distinction between pragmatic and semantic ambiguity (e.g., "proposition" and "meaning") are precise. They can be made precise only in the context of semantic and pragmatic theories. But even if it is unclear, in general, what pragmatic ambiguity is, it is clear, I hope, that my examples are cases of it.

Just as we can communicate effectively using quantified sentences without explicitly specifying a domain, so we can communicate effectively using conditional sentences without explicitly specifying an s-function. This suggests that there are further rules beyond those set down in the semantics, governing the use of conditional sentences. Such rules are the subject matter of a *pragmatics* of conditionals. Very little can be said, at this point, about pragmatic rules for the use of conditionals since the logic has not advanced beyond the propositional stage, but I will make a few speculative remarks about the kind of research that may provide a framework for treatment of the problem, and related pragmatic problems in the philosophy of science.

(1) If we had a quantified conditional logic, it is likely that $(\forall x)(Fx > Gx)$ would be a plausible candidate for the form of a law of nature. A law of nature says, not just that every actual F is a *G,* but further that for every possible *F,* if it were an *F,* it would be a *G.* If this is correct, then Hempel's confirmation paradox does not arise, since "All ravens are black" is not logically equivalent *to* "All non-black things are non-ravens." Also, the relation between counterfactuals and laws becomes clear: laws support counterfactuals because they entail them. "If this dove were a raven, it would be black" is simply an instantiation of "All ravens are black."[19]

Hempel's confirmation paradox: *a white shoe confirms "all non-black things are non-ravens" and so, if the two are equivalent, "all ravens are black" too.*

simply an instantiation: *Goodman explained counterfactuals by laws. Here a law is a universal conditional, so it entails its instances, and the hard work passes to the ordering of worlds.*

(2) Goodman argued that the pragmatic problem of counterfactuals is one of a cluster of closely related problems concerning induction and confirmation. He locates the source of these difficulties in the general problem of projectability, which can be stated roughly as follows: When can a predicate be validly projected from one set of cases to others? Or, when is a hypothesis confirmed by its positive instances? Some way of distinguishing between natural predicates and those that are artificially constructed is needed. If a theory of projection such as Goodman envisions were developed, it might find a natural place in a pragmatics of conditionals. Pragmatic criteria for measuring the inductive properties of predicates might provide pragmatic criteria for ordering possible worlds.[20]

natural predicates: *"green" extends from the emeralds examined to new ones, but "grue", green if examined before a set date and blue otherwise, doesn't. Whatever explains that might also rank worlds.*

(3) There are some striking structural parallels between conditional logic and conditional probability functions, which suggests the possibility of a connection between inductive logic and conditional logic. A probability assignment and an s-function are two quite different ways of describing the inductive relations among propositions; a theory that draws a connection between them might be illuminating for both.[21]

6. Conclusion: Empiricism and Possible Worlds

Writers of fiction and fantasy sometimes suggest that imaginary worlds have a life of their own beyond the control of their creators. Pirandello's six characters, for example,

Pirandello's six characters: *Six Characters in Search of an Author (1921): they turn up unwritten and demand their own story.*

[19] For a discussion of the relation of laws to counterfactuals, see Nagel 1961, 47–78. For a recent discussion of the paradoxes of confirmation by the person who discovered them, see Hempel 1966.

[20] Goodman 1955, especially ch. 4.

[21] Several philosophers have discussed the relation of conditional propositions to conditional probabilities. See Jeffrey 1964 and Adams 1966. I hope to present elsewhere my method of drawing the connection between the two notions, which differs from both of these.

rebelled against their author and took the story out of his hands. The skeptic may be inclined to suspect that this suggestion is itself fantasy. He believes that nothing goes into a fictional world, or a possible world, unless it is put there by decision or convention: it is a creature of invention and not discovery. Even the fabulist Tolkien admits that Faerie is a land "full of wonder, but not of information."[22]

a creature of invention: *the sceptic's point: what we make up holds no surprises, so there'd be nothing in a possible world to discover, and counterfactuals would record our decisions, not facts.*

For similar reasons, the empiricist may be uncomfortable about a theory that treats counterfactuals as literal statements about non-actual situations. Counterfactuals are often contingent, and contingent statements must be supported by evidence. But evidence can be gathered, by us at least, only in this universe. To satisfy the empiricist I must show how possible worlds, even if the product of convention, can be subjects of empirical investigation.

partially define a possible world: *the reply: to stipulate a world is to fix it by a description, not to know everything true in it. Say "like ours in China's population" and you've made it, yet much of its census you'll never know.*

There is no mystery to the fact that I can partially define a possible world in such a way that I am ignorant of some of the determinate truths in that world. One way I can do this is to attribute to it features of the actual world that are unknown to me. Thus, I can say, "I am thinking of a possible world in which the population of China is just the same, on each day, as it is in the actual world." *I* am making up this world—it is a pure product of my intentions—but there are already things true in it that I will never know.

Conditionals do implicitly: *a counterfactual fixes its world by ours, as ours except that A. So what you don't know of ours you don't know of it, and clues here, the boss's mood, bear on it, though not decisively.*

Conditionals do implicitly, and by convention, what is done explicitly by stipulation in this example. It is because counterfactuals are generally about possible worlds that are very much like the actual one, and defined in terms of it, that evidence is so often relevant to their truth. When I wonder, for example, what would have happened if I had asked my boss for a raise yesterday, I am wondering about a possible world that I have already roughly picked out. It has the same history, up to yesterday, as the actual world, the same boss with the same dispositions and habits. The main difference is that in that world, yesterday I asked the boss for a raise. Since I do not know everything about the boss's habits and dispositions in the actual world, there is a lot that I do not know about how he acts in the possible world that I have chosen, although I might find out by watching him respond to a similar request from another, or by asking his secretary about his mood yesterday. These bits of information about the actual world would not be decisive, of course, but they would be relevant, since they tell me more about the non-actual situation that I have selected.

subjunctive or otherwise: *one account for both moods. Whether it speaks of our world or another turns on whether A is in fact true, not on what the speaker believes, and shared conventions pick the rest.*

If I make a conditional statement, subjunctive or otherwise, and the antecedent turns out to be true, then whether I know it or not, I have said something about the actual world, namely that the consequent is true in it. If the antecedent is false, then I have said something about a particular counterfactual world, even if I believe the antecedent to be true. The conditional provides a set of conventions for selecting possible situations that have a specified relation to what actually happens. This makes it possible for statements about unrealized possibilities to tell us, not just about the speaker's intentions, but about the world.

[22] Tolkien 1966, 3.