cause some fundamental rethinking (at least in the "analytic" tradition, particularly in the science-centered portion) that would lead to some change of direction in philosophy from what I take to be a path filled with unfortunate misuses of very fine analytic powers.

HAO WANG

Rockefeller University

A COUNTEREXAMPLE TO MODUS PONENS*

THE rule of *modus ponens*, which tells us that from an indicative conditional ⌜If ϕ then ψ⌝,[1] together with its antecedent ϕ, one can infer ψ, is one of the fundamental principles of logic.[2] Yet, as the following examples show, it is not strictly valid; there are occasions on which one has good grounds for believing the premises of an application of modus ponens but yet one is not justified in accepting the conclusion. Later on, we shall see how these examples can be modified to give counterexamples to Stalnaker's semantics for the conditional:

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:
>
> If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
>
> A Republican will win the election.
>
> Yet they did not have reason to believe
>
> If it's not Reagan who wins, it will be Anderson.

> I see what looks like a large fish writhing in a fisherman's net a ways off. I believe
>
> If that creature is a fish, then if it has lungs, it's a lungfish.
>
> That, after all, is what one means by "lungfish." Yet, even though I believe the antecedent of this conditional, I do not conclude

*I would like to thank Ernest Adams for his great help in preparing this paper. He read the paper carefully and made a number of thoughtful and valuable suggestions.

[1] The corners, '⌜' and '⌝', are quasi-quotation marks. See Willard Van Orman Quine, *Mathematical Logic* (New York: Norton, 1940; Harper & Row, 1951), pp. 33-37.

[2] Here I speak of inferring the sentence ψ from the sentences ⌜If ϕ then ψ⌝ and ϕ, and at other places I shall speak of inferring the proposition ψ from the propositions ⌜If ϕ then ψ⌝ and ϕ. It would be more precise, but also more tedious, to say that we infer the proposition expressed by the sentence ψ from the propositions expressed by the sentences ⌜If ϕ then ψ⌝ and ϕ.

0022-362X/85/8209/0462$01.00

If that creature has lungs, it's a lungfish.

Lungfishes are rare, oddly shaped, and, to my knowledge, appear only in fresh water. It is more likely that, even though it does not look like one, the animal in the net is a porpoise.

Having learned that gold and silver were both once mined in his region, Uncle Otto has dug a mine in his backyard. Unfortunately, it is virtually certain that he will find neither gold nor silver, and it is entirely certain that he will find nothing else of value. There is ample reason to believe

If Uncle Otto doesn't find gold, then if he strikes it rich, it will be by finding silver.

Uncle Otto won't find gold.

Since, however, his chances of finding gold, though slim, are no slimmer than his chances of finding silver, there is no reason to suppose that

If Uncle Otto strikes it rich, it will be by finding silver.

These examples show that modus ponens is not an entirely reliable rule of inference. Sometimes the conclusion of an application of modus ponens is something we do not believe and should not believe, even though the premises are propositions we believe very properly.[3]

Modus ponens is sometimes thought of not as a rule of inference but as a law of semantics, to wit, whenever $\ulcorner$If ϕ then $\psi\urcorner$ and ϕ are both true, ψ is true as well. It is not at all obvious what we are to make of this law, since it is not evident what the truth conditions for the English conditional are or even whether it has truth conditions. Still it seems unlikely that, even if we learned the truth conditions for the English conditional, the semantic version of modus ponens would be vindicated. Let us imagine, on the contrary, that some time in the future linguists will determine the truth conditions for the English conditional and prove that modus ponens is truth-preserving. Assuming that basic zoology will not have changed, a future linguist who sees what looks like a large fish writhing in a fisherman's net a ways off will believe, as I believed,

If that animal is a fish, then if it has lungs it's a lungfish.
That animal is a fish.

Suppose he also believes this:

It is true that, if that animal is a fish, then if it has lungs it's a lungfish.
It is true that that animal is a fish.

[3] There are, of course, familiar cases in which we see that an application of modus ponens leads us from premises we reasonably believe to a conclusion we find utterly incredible, and we respond by repudiating the premises rather than accepting the conclusion. The present examples are not like this, since we do not renounce the premises.

Then he will be able to prove, using the well-established principle of future semantics that modus ponens is truth-preserving:

It is true that, if that animal has lungs, it is a lungfish.

He will not, however, believe

If that animal has lungs, it is a lungfish.

any more than I did. Thus our future linguist will be either in the awkward position of believing the premises of the argument without believing that those premises are true, or else in the equally awkward position of not believing the conclusion of the argument even though he does believe that that conclusion is true.[4] Thus the only way that we can hold on to the doctrine that modus ponens is truth-preserving will be to accept an unexpected disparity between believing a proposition and believing that that proposition is true.

In an attempt to supply truth conditions where nature provides none, philosophers have settled upon material implication: Count $\ulcorner$If ϕ then $\psi\urcorner$ as true if either ϕ is false or ψ is true. Sometimes this is intended as a proposal for linguistic reform, a suggestion that, at least in our scientific discourse, we ought to use the "If-then" construction in a new way, treating it as the material conditional rather than the ordinary conditional. Our examples do not raise any difficulties for this proposal, since if we reinterpret them this way, our examples become arguments with true premises and true conclusions. Sometimes, however, material implication is proposed as an account of how we presently use the "If-then" construction. This is surely wrong. If we have seen the polls showing Reagan far ahead of Carter, who is far ahead of Anderson, we will not for a moment suppose that

If Reagan doesn't win, Anderson will.

is true, even though we will resign ourselves to the truth of

Reagan will win.

Our counterexamples to modus ponens have a characteristic logical form. Each has as a premise a conditional whose consequent is itself a conditional. In general, we assert, accept, or believe a conditional of the form $\ulcorner$If ϕ, then if ψ then $\phi\urcorner$ whenever we are willing

[4] The first horn of this dilemma would not be uncomfortable to someone like Adams [*The Logic of Conditionals* (Boston: Reidel, 1975)] who doubts that conditionals are either true or false. By hypothesis, this is not the situation of our future linguist.

to assert, accept, or believe the conditional ⌜If ϕ and ψ, then θ⌝. It appears, from looking at examples, that the law of exportation.

⌜If ϕ and ψ, then θ⌝ entails ⌜If ϕ, then if ψ then θ⌝.

is a feature of English usage.[5] If so, then our counterexamples to modus ponens are not isolated curiosities but rather symptoms of a basic difficulty. It is natural to suppose that the English indicative conditional is intermediate in strength between strict implication and material implication. That is to say, whenever ψ is a logical consequence of ϕ, ⌜If ϕ then ψ⌝ will be true, and whenever ⌜If ϕ then ψ⌝ is true, either ϕ will be false or ψ true (and so modus ponens is truth-preserving). It now appears that we also want to require that the law of exportation be valid. But there is no connective other than the material conditional that meets all these requirements.

Theorem. Suppose that we have a logical consequence relation $\vdash$ on a language whose connectives comprise the ordinary Boolean connectives '$\vee$', '&', '$\sim$', '$\supset$', and '$\equiv$', as well as an additional conditional '$\Rightarrow$', satisfying the following conditions:

(Cons) $\vdash$, a relation between sets of sentences and sentences, is a consequence relation:
If $\phi \in \Gamma$, then $\Gamma \vdash \phi$.
If $\Gamma \vdash \phi$ and $\Gamma \subseteq \Delta$, then $\Delta \vdash \phi$.
If $\Delta \vdash \psi$ for each $\psi \in \Gamma$ and $\Gamma \vdash \phi$, then $\Delta \vdash \phi$.

(Exp) The law of exportation for "$\Rightarrow$":
$\{\ulcorner \phi \,\&\, \psi \Rightarrow \theta \urcorner\} \vdash \ulcorner \phi \Rightarrow (\psi \Rightarrow \theta)\urcorner$.

(MP) Modus ponens for both conditionals "$\Rightarrow$" and "$\supset$":
$\{\ulcorner \phi \Rightarrow \psi \urcorner, \phi\} \vdash \psi$
$\{\ulcorner \phi \supset \psi \urcorner, \phi\} \vdash \psi$

(StrImp) Strict implication is as strong or stronger than either conditional: If $\{\phi\} \vdash \psi$, then $\Phi \vdash \ulcorner \phi \Rightarrow \psi \urcorner$ and $\Phi \vdash \ulcorner \phi \supset \psi \urcorner$ (where Φ is the empty set).

(Taut) Ordinary Boolean connectives behave normally: If ϕ is a tautology,[6] then $\Phi \vdash \phi$.[7]

Then the two conditionals "$\Rightarrow$" and "$\supset$" are logically indistinguishable. More precisely, if ϕ and ϕ' are alike except that '$\Rightarrow$' and '$\supset$' have been exchanged at some places, then $\{\phi\} \vdash \phi'$ and $\{\phi'\} \vdash \phi$.

[5] It would appear that the law of importation, the converse of the law of exportation, is also valid.

[6] To see whether ϕ is a tautology, apply the following test: First replace every subformula of ϕ of the form ⌜$\psi \Rightarrow \theta$⌝ that is not itself contained in such a subformula by a new sentential letter. Then apply the usual truth-table test.

[7] We get an equivalent set of conditions by replacing (Exp) and (StrImp) for '$\Rightarrow$' by the principle

(Cond) If $\Gamma \cup \{\phi\} \vdash \psi$, then $\Gamma \vdash \ulcorner \phi \Rightarrow \psi \urcorner$.

This rule reflects the way we customarily prove conditionals: Add ϕ hypothetically to our body of theory. If we can prove ψ in the augmented theory, count ⌜If ϕ then ψ⌝ as proved.

The idea of the proof, which proceeds by induction on the complexity of ϕ, is contained in the proof that $\{\ulcorner \psi \supset \theta \urcorner\} \vdash \ulcorner \psi \Rightarrow \theta \urcorner$:[8]

(i) $\Phi \vdash \ulcorner ((\psi \supset \theta) \,\&\, \psi) \supset \theta \urcorner$ by (Taut).
(ii) $\{\ulcorner ((\psi \supset \theta) \,\&\, \psi) \supset \theta \urcorner; \ulcorner (\psi \supset \theta) \,\&\, \psi \urcorner\} \vdash \theta$ by (MP) for '$\supset$'
(iii) $\{\ulcorner (\psi \supset \theta) \,\&\, \psi \urcorner\} \vdash \theta$ from (i) and (ii) by (Cons)
(iv) $\Phi \vdash \ulcorner ((\psi \supset \theta) \,\&\, \psi \Rightarrow \theta \urcorner$ from (iii) by (StrImp) for '$\Rightarrow$'
(v) $\{\ulcorner ((\psi \supset \theta) \,\&\, \psi) \Rightarrow \theta \urcorner\} \vdash \ulcorner (\psi \supset \theta) \Rightarrow (\psi \Rightarrow \theta) \urcorner$ by (Exp)
(vi) $\{\ulcorner (\psi \supset \theta) \Rightarrow (\psi \Rightarrow \theta) \urcorner; \ulcorner \psi \supset \theta \urcorner\} \vdash \ulcorner \psi \Rightarrow \theta \urcorner$ by (MP) for '$\Rightarrow$'
(vii) $\{\ulcorner \psi \supset \theta \urcorner\} \vdash \ulcorner \psi \Rightarrow \theta \urcorner$ from (iv), (v), and (vi) by (Cons)

The theorem points to a tension between modus ponens and the law of exportation. According to the classical account, which does not recognize any conditional other than the material, both are valid; but we will not expect them both to come out valid on any nonclassical account.

We have explicit examples to show that the indicative conditional does not satisfy modus ponens. It is not so easy to test whether the rule is valid for the subjunctive conditional, since we seldom use the subjunctive conditional in situations in which we are confident that the antecedent is true. On the other hand, it is easy to find natural instances of the law of exportation that employ the subjunctive mood; for example,

> If Juan hadn't married Xochitl and Sylvia hadn't run off to India, Juan and Sylvia would have become lovers.

entails

> If Juan hadn't married Xochitl, then if Sylvia hadn't run off to India, Juan and Sylvia would have become lovers.

Multiplying such examples, we get good inductive evidence that the subjunctive conditional satisfies the law of exportation. If this evidence is correct, then no theory of the subjunctive conditional which denies the law of exportation will be entirely accurate. The most prominent logical theory of the subjunctive conditional is Robert Stalnaker's account,[9] according to which we test whether $\ulcorner \phi \Rightarrow \psi \urcorner$ is true in a possible world w by seeing whether ψ is true in the possible world most similar to w in which ϕ is true. Stalnaker's

[8] This conclusion already shows us that '$\Rightarrow$' is not genuinely stronger than the material conditional, as we would have hoped. Notice that to get it we need only this very weak form of (StrImp):

If ψ is a tautological consequence of ϕ, then $\Phi \vdash \ulcorner \phi \Rightarrow \psi \urcorner$.

[9] "A Theory of Conditionals," in Nicholas Rescher, ed., *Studies in Logical Theory. American Philosophical Quarterly* supplementary monograph series (Oxford: Blackwell, 1968), pp. 98–112.

system satisfies conditions (Cons), (MP), (StrImp), and (Taut), but it does not satisfy the law of exportation. Thus we are led to suspect that Stalnaker's analysis of the subjunctive conditional is inaccurate.

Concrete examples confirm our suspicions. We would ordinarily say (at least in contexts in which we are interested in the election results rather than, say, how else the primaries might have turned out),

> If Reagan hadn't won the election and a Republican had won, it would have been Anderson.

Appropriately, the Stalnaker semantics, under the natural comparitive similarity ordering among worlds, has this sentence come out true. As the law of exportation predicts, we also want to say,

> If Reagan hadn't won the election, then if a Republican had won, it would have been Anderson.

However, the possible world most similar to the actual world in which Reagan did not win the election will be a world in which Carter finished first and Reagan second, with Anderson again a distant third, and so a world in which "If a Republican had won it would have been Reagan" is true. Thus Stalnaker's theory wrongly predicts that, in the actual world,

> If Reagan hadn't won the election, then if a Republican had won, it would have been Reagan.

will be true. Thus, in this instance, the law of exportation is right and the Stalnaker semantics is wrong.

Another example: Let us imagine that, contrary to all our expectations, Uncle Otto finds a rich vein of gold, deeply buried in a distant corner of his property. We still believe this:

> If Uncle Otto hadn't found gold but he had struck it rich, it would have been by finding silver.

We also believe, as the law of exportation predicts,

> If Uncle Otto hadn't found gold, then if he had struck it rich, it would have been by finding silver.

What does the Stalnaker semantics say? The closest world to the actual world in which Uncle Otto does not find gold—call it *w*—will be a world in which the deposit of gold is located just on the other side of Otto's property line, or perhaps a world in which Otto does not dig quite deeply enough to reach the vein. The world closest to *w* in which Otto strikes it rich will be a world in which the gold is relocated back onto Otto's property and Otto digs deeply enough to find the gold. Thus the closest world to *w* in which Uncle Otto strikes it rich will be a world in which

> Uncle Otto finds gold.

is true. Therefore, in *w*,

> If Uncle Otto had struck it rich, it would have been by finding gold

is true, and so, according to Stalnaker's semantics,

> If Uncle Otto hadn't found gold, then if he had struck it rich, it would have been by finding gold.

is true in the actual world. Once again, the law of exportation scores a point against the Stalnaker semantics.

Our examples show us that an accurate logic for the English indicative conditional would have to restrict the rule of modus ponens somehow, and they suggest that the same would be true of an accurate logic of the subjunctive conditional. Nevertheless, all the familiar logics of the conditional countenance modus ponens without reservations. How do we account for this discrepancy? The simplest diagnosis is that we have committed an error of overly hasty generalization. We encounter a great many conditionals in daily life, and we have noticed that, when we accept a conditional and we accept its antecedent, we are prone to accept the consequent as well. We have supposed that this pattern held universally, with no exceptions. However, the examples we looked at were nearly always examples of simple conditionals, conditionals that did not themselves contain conditionals. Indeed there is every reason to suppose that, restricted to such conditionals, modus ponens is unexceptionable. But when we turn our attention to compound conditionals, new phenomena appear, and patterns that established themselves in the simple cases are disrupted.

The methodological moral to be drawn from this is that, when we formulate general laws of logic, we ought to exercise the same sort of caution we exercise when we make inductive generalizations in the empirical sciences. We must take care that the instances we look at in evaluating a proposed generalization are diverse as well as numerous.

It is perhaps surprising that, in constructing a logical theory, one comes upon the same pitfalls one encounters in the empirical sciences, since it is widely believed that logic is an a priori science. Upon reflection, however, we see that there is no cause for perplexity. If one believes that the correctness of a logically valid inference is recognized by an a priori intuition, what one believes is this:

> If $\mathcal{R}$ is a valid rule of inference, then whenever R is an instance of $\mathcal{R}$, one can see by an a priori intuition that R is a correct inference.

In order to conclude that the general laws of logic can be established purely by a priori reasoning, we would have to know something stronger, namely,

> If $\mathcal{R}$ is a valid rule of inference, then one can see by an a priori intuition that, whenever R is an instance of $\mathcal{R}$, R is a correct inference.

Our examples show that modus ponens is not strictly valid. They do nothing to dissuade us from our entrenched belief that modus ponens is valid for simple conditionals. They suggest that the law of exportation is valid for a wide range of cases, perhaps even valid universally. Beyond this, the examples give us no positive guidance toward constructing a correct logic of conditionals. It may be that some entirely new approach is needed, but it may also be that we can modify some existing theory to take the examples into account.

It is not hard to modify the Stalnaker semantics so that it has the right logical features. Instead of the simple notion of truth in a world, we develop a notion of truth in a world under a set of hypotheses. To be simply true in a world is to be true in that world under the empty set of hypotheses. If there is no world accessible from w in which all the members of Γ are true, then every sentence is true in w under the set of hypotheses Γ. Otherwise we have the following: An atomic sentence is true in w under the set of hypotheses Γ iff it is true in the possible world most similar to w in which all the members of Γ are true. A conjunction is true in a world under a given set of hypotheses iff each of its conjuncts is. A disjunction is true in a world under a set of hypotheses iff one or both disjuncts are. $\ulcorner \sim \phi \urcorner$ is true in w under the set of hypotheses Γ iff ϕ is not true in w under that set of hypotheses. Finally, $\ulcorner \phi \Rightarrow \psi \urcorner$ is true in w under the set of hypotheses Γ iff ψ is true in w under the set of hypotheses $\Gamma \cup \{\phi\}$. Thus to evaluate whether $\ulcorner \phi \Rightarrow (\psi \Rightarrow \theta) \urcorner$ is true under the set of hypotheses Γ, we add first ϕ and then ψ to our set of hypotheses, and we see whether θ is true under the augmented set of hypotheses $\Gamma \cup \{\phi, \psi\}$. This semantics gives a logic that is compact and decidable.

For each sentence constructed using this modified Stalnaker conditional, we can find a logically equivalent sentence that uses the original Stalnaker conditional. We use '$\Rightarrow$' to stand for the modified Stalnaker conditional and '>' to denote the connective Stalnaker originally described. We take the Boolean connectives to be '$\vee$', '&', '$\sim$', and a logically constant false sentence '$\bot$'. Define the operation * by:

$\phi^* = \phi$ if ϕ is an atomic sentence.
$'\bot'^* = '\bot'$
$\ulcorner(\phi \vee \psi)\urcorner^* = \ulcorner(\phi^* \vee \psi^*)\urcorner$
$\ulcorner(\phi \,\&\, \psi)\urcorner^* = \ulcorner(\phi^* \,\&\, \psi^*)\urcorner$
$\ulcorner{\sim}\phi\urcorner^* = \ulcorner{\sim}(\phi^*)\urcorner$
$\ulcorner(\phi \Rightarrow \psi)\urcorner^* = \ulcorner(\phi^* > \psi)\urcorner$ if ψ is an atomic sentence or '$\bot$'
$\ulcorner(\phi \Rightarrow (\psi \vee \theta))\urcorner^* = \ulcorner((\phi \Rightarrow \psi)^* \vee (\phi \Rightarrow \theta)^*)\urcorner$
$\ulcorner(\phi \Rightarrow (\psi \,\&\, \theta))\urcorner^* = \ulcorner((\phi \Rightarrow \psi)^* \,\&\, (\phi \Rightarrow \theta)^*)\urcorner$
$\ulcorner(\phi \Rightarrow {\sim}\psi)\urcorner^* = \ulcorner((\phi \Rightarrow \bot)^* \vee{\sim}((\phi \Rightarrow \psi)^*))\urcorner$
$\ulcorner(\phi \Rightarrow (\psi \Rightarrow \theta))\urcorner^* = \ulcorner((\phi \,\&\, \psi) \Rightarrow \theta)\urcorner^*$

ϕ and ϕ^* are logically equivalent.

Another approach we might use would be to continue to use a formal system in which modus ponens has unrestricted validity, and to take account of the invalidity of modus ponens in English by modifying our informal rules for translating English sentences into the formal language.[10] Thus we do not translate an English sentence of the form $\ulcorner$If ϕ, then if ψ then $\theta\urcorner$ in the natural way, as a formula of the form $\ulcorner(\phi \Rightarrow (\psi \Rightarrow \theta))\urcorner$; instead we translate it as $\ulcorner((\phi \,\&\, \psi) \Rightarrow \theta)\urcorner$. Thus the invalid English inference:

If ϕ, then if ψ then θ.
ϕ.
Therefore if ψ then θ.

is translated as the invalid formal inference:

$(\phi \,\&\, \psi) \Rightarrow \theta$.
ϕ.
Therefore $\psi \Rightarrow \theta$.

It is sometimes a bit arbitrary whether to account for a feature of English usage within our formal system or to account for it at the informal level of translation lore. For example, we just discussed a way of modifying the Stalnaker conditional so as to make the law of exportation generally valid. If we let Tr(ϕ) be the "natural" translation of an English sentence ϕ into a formal language whose connectives are the Boolean connectives and '$\Rightarrow$', we can equally well take the translation of ϕ to be Tr(ϕ) and use the modified

[10] Barry Loewer, "Counterfactuals with Disjunctive Antecedents," this JOURNAL, LXXIII, 16 (Sept. 16, 1976): 531-537, has proposed using this strategy for coping with a different difficulty with Stalnaker's analysis.

Stalnaker semantics or take the translation of ϕ to be $(\mathrm{Tr}(\phi))^*$ and use the original Stalnaker semantics.

The selective use of unnatural translations is a powerful technique for improving the fit between the logic of the natural language and the logic of a formal language. In fact, it is a little too powerful. One suspects that, if one is sly enough in giving translations, one can enable almost any logic to survive almost any counterexample. What is needed is a systematic account of how to give the translations. In the absence of such an account, the unnatural translations will seem like merely an ad hoc device for evading counterexamples.

There is no guarantee that any approach will work. It may be that it is not possible to give a satisfactory logic of conditionals. This is not to say that it is not possible to give a linguistic account of how we use conditionals, but only to say that such an account would not give rise to a tractable theory of logical consequence.

VANN McGEE

University of California at Berkeley

KRIPKE ON WITTGENSTEIN ON RULES*

THERE is no doubt that Ludwig Wittgenstein thought the topic of rule following to be important; nearly forty sections of the *Philosophical Investigations* are devoted to it, as are large swatches of the manuscripts published as *Remarks on the Foundations of Mathematics.*[1] Its relevance to Wittgenstein's philosophy of mathematics was emphasized early on by Michael Dummett;[2] but only recently has it received significant attention in the less specialized context of the *Investigations,* that is, with respect to questions of meaning and intentionality. This recent attention has, to a large extent, been engendered by Saul Kripke's exposition of Wittgenstein, first presented publicly at the 1976 Wittgenstein Colloquium in London, Ontario, and laid out more expansively in his recent book, *Wittgenstein on Rules and Private Language.*[3] Kripke reads Wittgenstein to be mounting a skeptical

*I am grateful to Burton Dreben, Paul Hoffman, Peter Hylton, Edward Minar, and, especially, Thomas Ricketts, for helpful comments and discussions.

[1] *Philosophical Investigations* (New York: Macmillan, 1953), hereafter cited as PI; *Remarks on the Foundations of Mathematics,* rev. ed. (Cambridge, Mass: MIT Press, 1978), hereafter cited as RFM.

[2] "Wittgenstein's Philosophy of Mathematics," *Philosophical Review,* LXIX, 3 (July 1959): 324-348.

[3] Cambridge, Mass: Harvard, 1982. Parenthetical page references to Kripke will be to this book.

0022-362X/85/8209/0471$01.80