SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 21-35
Robots and if … then

ROBOTS AND *IF … THEN*

Ronald A. Cordero

<cordero@uwosh.edu>

0. Prefatory Note

For the sake of simplicity in this paper I am going to use expressions of cognitive process and propositional attitude, such as «reason,» «know,» «suspect,» «remember,» and «think,» without elaboration in referring to robots. In doing so, I do not mean to be taking a position on the question of whether or not robots can actually reason, think, remember, suspect, or know in the same way humans can. I simply wish to avoid the complexity of saying, for example, that a robot «believes$_R$,» where «believes$_R$» is defined as «is in the robotic state which corresponds to the state of mind of a human who believes.» I do think the question of whether or not entities like robots or non-human animals can have propositional attitudes and engage in cognitive processes is an interesting one, but I do not think that it has to be answered before issues regarding the use of «if…then» by robots can be resolved, and I do not propose to try to answer it here.

1. Introduction

We are going to have robots that reason. The scientific, commercial , and military motives for building them and letting them work with some degree of autonomy are going to prove irresistible. But this means that if we want to avoid awkward or even disastrous consequences, we must program these robots to make only valid, humanly intelligible inferences. And this may not be a simple thing to do where «if…then» is concerned. Conditional («if…then») statements are extremely important in human communication, but their analysis has long been a source of disagreement among logicians. In the present paper, I argue that unless we wish to court disaster, we cannot let our robots use the most widely accepted rules of inference concerning conditional statements and I propose alternative rules that I believe will stave off disaster.

There is at present a long list of commonly accepted «rules of inference» for what is commonly called «statement logic.» These «rules» are in essence analytically true statements of entailment relations between statements that contain expressions such as «not,» «or,» «and,» «if…then,» and «if and only if.» Because these analytically true statements are statements of fundamental truths of logic, I shall refer to them as axioms. Some statement-logic axioms express one-way entailment relations, saying that statements of one sort entail statements of another sort. Others express two-way entailment relations, saying that statements of a first sort entail and are entailed by — and thus are equivalent to — statements of a second sort. If, as is usually done, pairs of similar statement-logic axioms are counted as a single axiom, the total number of commonly accepted axioms is

eighteen. If each axiom were to be counted separately, the total number would be twenty-four. Fewer than half of these axioms involve conditional statements. Using a single-headed double-barred arrow (⇒) to indicate one-way entailment and a double-headed double-barred arrow (⇔) to indicate two-way entailment, those which do can be represented as follows:

Modus Ponens	$((p \supset q) \wedge p) \Rightarrow q$
Modus Tollens	$((p \supset q) \wedge \sim q) \Rightarrow \sim p$
Hypothetical Syllogism	$((p \supset q) \wedge (q \supset r)) \Rightarrow (p \supset r)$
Constructive Dilemma	$(((p \supset q) \wedge (r \supset s)) \wedge (p \vee r)) \Rightarrow (q \vee s)$
Transposition	$(p \supset q) \Leftrightarrow (\sim q \supset \sim p)$
Material Implication	$(p \supset q) \Leftrightarrow (\sim p \vee q)$
Material Equivalence	$(p \equiv q) \Leftrightarrow ((p \supset q) \wedge (q \supset p))$
	$(p \equiv q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$[1]
Exportation	$((p \wedge q) \supset r) \Leftrightarrow (p \supset (q \supset r))$

If our robots are not to make serious mistakes in reasoning with conditional statements, some of these axioms will have to be given to them in altered forms. In the following sections I shall illustrate the sort of problems that will arise if robots are allowed to use all these axioms without alterations — and shall suggest alternative axioms that can preclude the problems. In a sense, this will involve putting certain axioms «off limits» to robots, establishing what might be called a list of «forbidden inferences.»

2. Material Implication

Among the classic statement-logic axioms, one real potential trouble maker is the one commonly known as «Material Implication.» In its traditional form it encapsulates an analysis of conditional statements that has been widely used by logicians since early in the twentieth century — and which in fact goes back to Philo of Megara in the fourth century BC.[2] According to this analysis, when we say «If a, then b,» we are basically saying that either a is false or b is true. A conditional statement on this analysis, that is, is equivalent to the disjunction of the consequent with the negation of the antecedent. This core meaning has been held to be common to the different types of «if…then» statements,[3] and it is this meaning that the horseshoe (⊃) has been used to symbolize. In rendering «If a, then b,»

---

[1]This second Equivalence axiom effectively involves conditionalization, since in the light of the first Equivalence axiom it equates to $((p \supset q) \wedge (q \supset p)) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$.

[2]See David H. Sanford, *If P, then Q: Conditionals and the Foundations of Reasoning* (London: Routledge, 1989) 14-26.

[3]See, for example, Howard Kahane and Paul Tidman, *Logic and Philosophy: A Modern Introduction* (Belmont: Wadsworth, 1995) 25.

as «a ⊃ b» , we are taking the statement to mean ~a ∨ b.

Is there any reason why robots should not be allowed to use the Material Implication axiom in its traditional form? The answer is definitely «Yes»: if they are allowed to do so, their reasoning simply cannot be trusted. And there is not just one way in which the axiom can cause problems: it can generate unacceptable inferences in a variety of ways.

For one thing, application of Material Implication in conjunction with the axiom called «Addition» (a ⇒ (a ∨ b)) can generate the infamous «paradoxes of material implication» — in which one finds that any true statement is materially implied by absolutely any other statement, and that any false statement materially implies any other statement whatsoever! Suppose, for example, that a robot, ℜ, knows that Ms. Gonzales is in Paris. (The statement P is true.) With Material Implication and Addition, ℜ could reason as follows…

1	She is in Paris.	P	Known
2	Either she's not in Germany or she's in Paris.	~G ∨ P	1;Addit., Commutat.
3	So if she's in Germany, she's in Paris.	G ⊃ P	2; Material Implication

If ℜ realizes that G is false — since to be in Paris is not to be in Germany — ℜ might phrase 3 counterfactually — in terms of what *would* be the case *if* G were true: «If she were in Germany, she would be in Paris.» But that would hardly alleviate the paradox. And of course, ℜ could similarly infer that she would be in Paris if she were in Spain, India, or on the moon! Clearly something isn't right here.

When statements known to be false are involved, the results of using Material Implication together with Addition can be just as bizarre. A robot can easily reach the conclusion that contrary counterfactual conditional statements are both true. Suppose ℜ knows that a certain stock did not go up. (The statement U is false.) Material Implication and Addition would permit the following peculiar line of reasoning…

1	The stock did not go up.	~U	Known
2	Either it didn't go up or we made a profit.	~U ∨ P	1; Addition
3	So if it had gone up, we would have made a profit.	U ⊃ P	2;Material Implication
4	Either it didn't go up or we didn't make a profit.	~U ∨ ~P	1; Addition
5	So if it had gone up, we would not have made a profit.	U ⊃ ~P	4;Material Implication

Presumably, we would not want to have ℜ thinking both 3 and 5! We could ill afford to have our robots endorsing contrary assertions. Nor would we want them to reach a conclusion like either line 3 or line 5 on the mere «evidence» that the stock did not go up!

However troublesome — and worth avoiding — they may be, such paradoxical results are by no means the extent of the trouble that the use of Material Implication in its usual form can occasion.[4] Especially serious problems have been noted to arise when the axiom is applied to statements involving negated conditionals. A beautiful example was presented by Charles L. Stevenson in 1970:

[4]For a brief discussion of problem cases, see Robert E. Rodes, Jr., and Howard Pospesel, *Premises and Conclusions: Symbolic Logic for Legal Analysis* (Upper Saddle River, NJ: Prentice Hall, 1997) 261-66.

> This is false: if God exists then the prayers of *evil* men will be answered. So we may conclude that God exists, and (as a bonus) we may conclude that the prayers of evil men will not be answered.[5]

Using the original version of the Material Implication axiom, we must regard the conditional $G \supset P$ as logically equivalent to $\sim G \vee P$ and it's negation as equivalent to $G \wedge \sim P$. It seems that we have either a new proof of the existence of God or a serious problem with Material Implication!

Evidently, reasoning such as this could make robots highly unreliable. Suppose we instruct ℜ, whom we have programmed to use Material Implication in its original form, to monitor various transmissions for remarks on certain topics and to verify intercepts on those topics by turning to a reliable source (RS) — perhaps a human operator. Imagine then that one day ℜ encounters the assertion «If Jones was involved, Smith was involved.» Following instructions, ℜ seeks confirmation from RS:

> ℜ: «Is it true that Smith was involved if Jones was?»
>
> RS: «No.»

The problem now is that ℜ, applying the standard axioms, can reason as follows:

1	It's not true that Smith was involved if Jones was.	$\sim(J \supset S)$	From RS
2	It's not the case that either Jones wasn't involved or Smith was.	$\sim(\sim J \vee S)$	1; Material Implication
3	Jones was involved but Smith wasn't.	$J \wedge \sim S$	2;DeMorgan,Double Negat.
4	Jones was involved.	J	3; Simplification
5	Smith wasn't involved.	$\sim S$	3; Simplification

This leaves ℜ with two beliefs confirmed by a reliable source: (1) that Jones was involved and (2) that Smith was not. If later queried as to who was involved, ℜ will reply accordingly: Jones was, but Smith was not. And any decisions subsequently made by ℜ will be based on these conclusions. For instance, if instructed to distribute certain sensitive information exclusively to individuals not involved, ℜ will send that information to Smith but not to Jones.

But there is no reason to believe that RS in replying as it did meant to say that one of the two individuals was involved and the other was not. RS's reply may have reflected no more than the belief that Jones *could* have been involved without Smith being involved as well. As a consequence of applying Material Implication, ℜ has acquired beliefs that could well be false.

Will this problem with negated conditionals arise only when a robot is told that some conditional statement is false? Evidently not. There are various other ways in which a reasoning robot could encounter negated conditionals. This could happen, for example, if a robot thinking for itself applied the axiom Modus Tolens to a conditional statement with a conditional antecedent. Suppose ℜ knows that if the object breaks when Jane drops it, it isn't made of plastic. A problem could arise if ℜ learned that the object is in fact plastic and proceeded to reason as follows…

---

[5] «If-iculties,» *Philosophy of Science* 37 (1970): 28.

1	If it breaks if she drops it, it's not plastic.	$(D \supset B) \supset \sim P$	Known
2	But it is plastic.	P	Learned
3	So it's not the case that it will break if she drops it.	$\sim(D \supset B)$	1,2;DoubleNegation, MT
4	It's not true that either she won't drop it or it will break.	$\sim(\sim D \vee B)$	3; Material Implication
5	She will drop it and it won't break.	$D \wedge \sim B$	4;DeMorgan,DoubleNegat.
6	So she is going to drop the object.	D	5; Simplification

$\mathfrak{R}$ is thus left thinking that Jane is going to drop the object. By using the original Material Implication axiom, $\mathfrak{R}$ has arrived at a belief that does not follow from the premises and may well be false. No careful human reasoner would infer the last line from the first two.

Use of the original Material Implication axiom on negated conditionals could also lead robots to have false beliefs about the beliefs of others. Suppose $\mathfrak{R}$ monitors the following conversation on the assumption that neither of the parties is lying:

Smith: They won't attack unless they are threatened.

Jones: I don't agree.

In this case, $\mathfrak{R}$ will understand Smith to believe that $\sim T \supset \sim A$ and Jones to believe that $\sim(\sim T \supset \sim A)$, which by the original Material Implication axiom equates to $\sim TA$. $\mathfrak{R}$, that is, has Jones down as thinking that they will not be threatened but *will* attack. But Jones might not have that belief at all: he might only (and in fact would probably only) believe that the attack *could* occur without any threat being made.

Clearly, reasoning robots cannot be allowed to use the Material Implication axiom in its traditional form: the problems that can result are just too serious. But then the question arises as to what form of the axiom, if any, robots should be given. What I am going to suggest is that they be given precisely half of it. In its usual form, the axiom is a statement of two-way entailment: $(p \supset q) \Leftrightarrow (\sim p \vee q)$. And this is of course equivalent to the conjunction of two one-way entailment axioms: $(p \supset q) \Rightarrow (\sim p \vee q)$ and $(\sim p \vee q) \Rightarrow (p \supset q)$. Of these, I propose letting our robots have only the first, putting the second, as it were, on the list of «forbidden inferences.» I want to have them think, as does P. F. Strawson, that while conditional statements entail disjunctions, disjunctions do not entail conditional statements.[6] Obviously this involves taking the position (or at least giving our robots to understand) that a conditional statement means something significantly more than the disjunction of its consequent with the negation of its antecedent. To say «If p, then q,» is, on this position, to say more than just «Either p is false or q is true.» But, at least insofar as statement logic is concerned, our robots will get only the one-way entailment axiom $(p \supset q) \Rightarrow (\sim p \vee q)$. And to avoid confusion, in fact, I think it may be best not even to use the name «Material Implication» in referring to it. For want of a better name, I shall refer to it simply as «Deconditionalization.»

Furthermore, in the same interest of avoiding confusion, I think it may be a good idea not to continue the traditional use of the horseshoe symbol ($\supset$) in representing conditional statements. Use of this symbol goes back to Giuseppe Peano, who created it

[6] See Strawson, *Introduction to Logical Theory* (London: Methuen, 1952) 82-83.

in 1891 by turning the letter «c» around. He would use it, he said, to represent «*b* is a consequence of *a*,» or «if *a* then *b*,» and in doing so he may well have meant to capture the full meaning of «if…then.»[7] But since that time the symbol, slightly elongated in form and dubbed the «horseshoe,» has become so thoroughly associated with the material implication interpretation of «if…then» that to use it while placing half of the traditional Material Implication axiom «off limits» could be badly confusing. According, I shall use the pound sign (#), rendering «If p, then q,» as (p # q), with the understanding that conditional statements so represented retain the full meaning of natural language statements — not just a part of that meaning, however significant that part might be.

Two points should be noted here with respect to Deconditionalization. First, I am continuing to put the right-hand side of the axiom in terms of a disjunction, though it could just as well be put as the negation of a conjunction: $p \# q \Rightarrow \sim(p \wedge \sim q)$. And second, the axiom can be conveniently taken as having a second line:

Deconditionalization: $(p \# q) \Rightarrow (\sim p \vee q)$

$(p \wedge \sim q) \Rightarrow \sim(p \# q)$

The second line, which evidently results from transposing the first and applying De Morgan, may in certain cases allow robots to see the falsity of a conditional claim.

Now the question is whether giving our robots Deconditionalization instead of Material Implication will remedy the sort of problems already noted. I believe in fact that it effectively eliminates them. First of all, the paradoxes of material implication are prevented from arising. With regard to the first example above, for instance, ℜ can infer that Ms. Gonzales is either not in Germany or else is in Paris — but cannot get from that to the conclusion that she is (or would be) in Paris if she is (or were) in Germany. Without the «off-limits» axiom $(\sim p \vee q) \Rightarrow (p \supset q)$, $\sim G \vee P$ does not entail $G \supset P$. Nor could ℜ, in the second example, get from the knowledge that the stock did not go up to the conclusion that we would have made (or would *not* have made) a profit if it had gone up. In the list of axioms ℜ is allowed to use, there is no longer one for going from a disjunction to a conditional statement.

But might not robots still encounter the paradoxes if they are allowed to use the logical technique known as conditional proof (CP) — assuming certain statements to be true and reaching conclusions about what follows if they are? In particular, could not ℜ reason as follows in the case of the first example?

1.	She's in Paris.	P	Known
2.	She's not in Germany or else she's in Paris.	$\sim G \vee P$	1; Addition, Commutation
3.	She's in Germany.	G	Assumed for CP.
4.	She's in Paris.	P	2,3; DoubNeg, Disj.Syll.
5.	So if she's in Germany, she's in Paris.	G # P	3-4; CP

In other words, could not ℜ reach the same paradoxical conclusion even without

[7] «Principii di logica matematica,» Rivista di matematica, 1 (1891): 1-10.

Material Implication by using conditional proof? The answer, I submit, is that ℜ could not. The proof above would not work because ℜ could not assume G on line 3. To do so would be, in effect, to add a premise inconsistent with the premise on line 1. If Ms. Gonzales is in Paris, she is not in Germany. And our robots cannot be allowed to assume things known to be false. ℜ would be instructed to reason only from consistent premises and so could not accept both «She is in Paris,» and «She is in Germany,» as premises. Since in this case ℜ knows the former to be true, ℜ could not assume the latter.

Yet isn't it ever possible to assume something false (suppose it to be true) for the sake of seeing what *would* follow? Could not ℜ reason about what would follow if Gonzales *were* in Germany, saying, as it were, «We know she's not in Germany, but what if she were?» That could indeed be done, but no paradox would arise — because in supposing for the sake of argument that she is in Germany, ℜ would have to abandon the original premise that she is in Paris. «Very well. We'll suppose she's in Germany and not in Paris.» And without P, the argument that generated the paradox would not go through.

Similarly, CP could not be used to revive the paradox from the second example either:

1.	The stock isn't going up.	~U	Known
2.	Either it isn't going up or we'll make a profit.	~U ∨ P	1; Addition
3.	It's going to go up.	U	Assumed for CP.
4.	We will make a profit.	P	2,3; DoubNeg, Disj.Syll.
5.	So if it goes (were to go) up, we'll make (would make) a profit.	U # P	3-4; CP

Here it is obvious that ℜ cannot make the assumption in 3: it directly contradicts the premise on line 1.

The point is really quite general: robots will have to be instructed not to accept any statement as a fact or an assumption unless it is consistent with all statements already accepted as facts or assumptions. Before committing any statement to memory as true, reasoning robots will have to check for the consistency of that statement with all other statements already held to be true. We cannot expect robots to reach acceptable conclusions if we permit them to tolerate inconsistency.

It may be worthwhile to note at this point that with Deconditionalization replacing Material Implication there seems to be no need to tell our robots anything at all about «possible worlds» in conjunction with conditional statements. The elaborate analyses introduced by David Lewis and Robert Stalnaker[8] will not have to be taken into consideration in order to have robots use «if…then» correctly — even in counterfactual cases.

And what now of the problems that can arise from the use of Material Implication in conjunction with negation? Our robots will not encounter them, I think, so long as they

[8] See Lewis, «Counterfactuals and Comparative Possibility,» and Stalnaker, «A Theory of Conditionals,» both in William L. Harper, Robert Stalnaker, and Glenn Pearce, eds., *Ifs* (Dordrecht: D. Reidel, 1981) 41-85.

stick to using Deconditionalization instead of Material Implication. The reason is simple: without the two-way entailment, robots will not be able to get from ~(p # q) to p∧~q. They will not find «This is false: if God exists then the prayers of *evil* men will be answered,» equivalent to «God exists, and the prayers of evil men will not be answered.» By the axioms permitted them, ~(G # P) is not equivalent to G∧~P. Nor is «It's not true that Smith was involved if Jones was,» ~(J # S), equivalent to «Jones was involved, but Smith wasn't,» J∧~S.

A question of course arises as to just what we should tell our robots the negation of a conditional statement is equivalent to — if not the conjunction of the antecedent with the negation of the consequent. The answer, I believe, is that we cannot tell them that the negation of a conditional is equivalent to anything that can be expressed in statement logic. «If a then b,» means something significantly more than «Either a is false or b is true,» so to deny that «If a then b,» is true is to do something more than assert that a is true and b is not. And the something more cannot be put in terms of simple statements and logical operators. In statement logic, it appears, our robots cannot have *any* axiom of the form «~(p # q) ⇔ …».

But will they at least have an axiom of the form «~(p # q) ⇒ …» so that they can infer *something* from the negation of a conditional statement? I am afraid the answer is that we cannot give them any such one-way entailment axiom in statement logic either. As I have said, in asserting «If a then b,» one is asserting more than «Either a is false or b is true.» The latter is, to be sure, part of what one means — as is explicitly acknowledged by Deconditionalization ((p # q) ⇒ (~p ∨ q)). But ~a ∨ b is only part of what is asserted when one asserts (a # b). So when (a # b) is denied, it is not necessarily the case that ~a ∨ b is being denied. By analogy, in asserting that Jones is a bachelor, one is asserting both that Jones is a man and that Jones is unmarried — so that, if I say «Jones is not a bachelor,» I am not necessarily saying that Jones is married. (I may be denying that Jones is a man.)

What then are robots to do when they encounter the negation of a conditional statement? Humans, after all, are not stopped cold by such an encounter! If we want our robots to reason like humans (who are reasoning correctly, that is), we cannot leave them stymied when conditionals are negated. We have to tell them how to proceed in such cases.

Consider first the sort of case in which the truth of a conditional is denied by someone with whom a robot is communicating. Here the robot can simply ask for an explanation, inquiring — as it were — why the other party (OP) believes the conditional in question to be false. In practice, this will involve asking whether the other party believes something that would entail ~(a # b). In most cases, if OP believes a conditional statement is false it will be because OP believes (1) that the contrary of the conditional is true or (2) that it is possible for the antecedent of the conditional to be true while the consequent is false. So in most cases our robot will be able to proceed by formulating questions about just these two possibilities.

If, however, we want to represent these questions symbolically, we will have to go beyond statement logic, since while the contrary of (a # b) can be represented in statement

logic as (a # ~b), a symbolic representation of the assertion that the antecedent can be true without the consequent being true requires a symbol for possibility. We need to represent the statement «It is possible that a∧~b». Moreover, we cannot use the ordinary possibility operator of modal logic (◊), since this symbol is usually used to represent *logical* possibility. If we wrote «◊(a∧~b),» we would be representing «It is logically possible that a∧~b.» But the possibility that our robot has to ask about when a conditional statement is denied is not limited to logical possibility. We need, accordingly, a symbol for possibility in general. For ease of understanding, I will use a black diamond (♦) for this purpose. The possibility it represents may be logical possibility — or may be possibility of some other sort:

1) They may not get here by five.
   ♦[~(They will get here by five.)]
   ♦(~G)
2) She may have read the report.
   ♦(She read the report.)
   ♦(R)

Similarly, if our robot has to speak of certainty in discussing conditionals, a black square (■) will be used to indicate certainty in general, as opposed to logical certainly in particular:

1) They will get here for sure by five.
   ■(They will get here by five.)
   ■(G)
2) She definitely read the report.
   ■(She read the report.)
   ■(R)

Naturally, we will also have to provide our robots with a general Possibility-Certainty axiom to the effect that ~♦(p) ⇔ ■(~p). Thus «It's not possible that they will get here by five,» ~♦(G) , will be interpreted by robots as equivalent to «It is certain that they will not get here by five,» ■(~G).

In asking about beliefs that would entail the negation of a conditional, our robots are going to be relying on certain other axioms that we ought to make explicit at this point and which, in fact, are not hard to state in statement logic augmented by general possibility and certainty operators. A first additional axiom, which can be called Conditional Contrariety, merely notes the fact that if a conditional statement is true, its contrary must be false:

Conditional Contrariety (p # q) ⇒ ~(p # ~q)

Loosely put, if the truth of a first statement (the antecedent) means that a second statement (the consequent) is true, it is not the case that the truth of the first statement means that the second is false. If it is the case that you will win if you enter the contest, it is not the case that you will not win if you enter. If it is true that she would go to Paris if she did not have to go to Chicago, it is false that she would not go to Paris if she did not have to go to Chicago. Our robots will have to be able to make inferences such as these.

A second additional axiom for the inference of the negation of a conditional records

the analytic truth that a conditional statement is false if it is possible for its antecedent and the negation of its consequent both to be true: $\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \ \# \ q)$. Interestingly, this axiom embodies an important part of the interpretation of «if…then» proposed by the Stoic Chrysippus, who headed the Stoic School in Athens after Zeno and Cleanthes in the third century BC. What Chrysippus apparently held was that a conditional proposition is true when (and only when) it is impossible for its antecedent to be true and its consequence false.[9] This means he would endorse both the axiom just stated and another that I am not willing to let our robots have: $\sim\blacklozenge(p \wedge \sim q) \Rightarrow (p \ \# \ q)$. (Giving them that would only, I fear, lead them into more paradoxes.) Still, out of historical deference, I think we can refer to the axiom I *am* proposing to give robots by his name:

Chrysippus $\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \ \# \ q)$

$(p \ \# \ q) \Rightarrow \sim\blacklozenge(p \wedge \sim q)$

The second line, of course, is merely the first line transposed.

It may be worth while noting that with Chrysippus, our robots will have an alternate «route» from $\sim p \wedge q$ to $\sim(p \ \# \ q)$ — in addition, that is, to that employing the second line of Deconditionalization. In all probability we will have to give our robots a modal axiom concerning the relation between what is and what is possible (in the general sense of possibility):

Actuality-Possibility $p \Rightarrow \blacklozenge p$

$\sim\blacklozenge p \Rightarrow \sim p$

(Here again, the second line of the axiom is merely the first line transposed.) But by applying this axiom, robots can reason from $\sim p \wedge q$ to $\sim(p \ \# \ q)$ without using Deconditionalization:

1	Smith was at the meeting and Jones wasn't.	$S \wedge \sim J$	Discovered
2	So it was possible for Smith to be at the meeting without Jones being there.	$\blacklozenge(S \sim J)$	1; Actuality-Possibility
3	So it's not the case that Jones was there if Smith was.	$\sim(S \ \# \ J)$	2; Chrysippus

Now we can return to the question of how to have robots respond when the other party, with whom they are communicating, negates a conditional statement. In the abstract, then, a robot could respond as follows:

OP: $\sim(A \ \# \ B)$

$\mathfrak{R}$: Do you think that $(A \ \# \sim B)$ or that $(A \wedge \sim B)$ or that $\blacklozenge(A \wedge \sim B)$?

Here, $\mathfrak{R}$ is inquiring about beliefs which OP might have that would *entail* $\sim(A \ \# \ B)$. Now consider the following concrete example:

---

[9] Josiah B. Gould, *The Philosophy of Chrysippus*, (Albany: State University of New York Press, 1970) 80.

OP	It's not the case that if they drink that beverage they will get sick.	~(D # S)
ℜ	Do you mean that they can drink it without getting sick — or that if they drink it they won't get sick?	♦(D∧~S)? (D # ~S)?
OP	I mean it's possible for them to drink it and not get sick.	♦(D∧~S)

Notice that ℜ does not even have to ask whether OP believes that D∧~S. OP clearly does not have to believe that they are going to drink the beverage and not get sick in order to think it is false that D # S. And if as a matter of fact OP does happen to be convinced that the parties in question are going to drink and not get sick (D∧~S), OP can be assumed to be using the Actuality-Possibility axiom too, and thus to believe (or at least to *constructively* believe) that ♦(D∧~S). So the two alternatives suggested by ℜ are sufficient. The truth of either would entail (and thus explain the belief of OP) that ~(D # S). After OP's responds, ℜ will know what OP is thinking.

There will also be cases in which the party who denies the truth of a conditional does so because of the belief that its contrary is true:

ℜ:	If we take this road, will we get to the camp?	(R # C)?
OP	No.	~(R # C)
ℜ	Do you mean that we can take this road and not get to camp — or that we won't get to camp if we take it?	♦(R∧~C)? (R # ~C)?
OP	I mean we won't get to camp if we take this road.	R # ~C

Here again there is no point in ℜ inquiring whether OP thinks that R∧~C. OP may not know whether the road will be taken or not, but OP is definitely of the opinion that taking it would prevent getting to camp. By asking such questions when a conditional is negated by some other party, a robot can at least avoid mistaken conclusions as to what that party believes.

Is there nothing else that the party denying a conditional might mean? If fact, in rare cases, when inquiring why OP denies that (a#b), ℜ may find that OP thinks neither that a#~b nor that ♦(a∧~b). There is a third possibility, though it is not one that will be encountered very often. Moreover, as it happens, it is not something that can be couched in terms of statement logic — or in statement logic augmented by a modal operator. Suppose, for example, someone says, «If Jim has his lucky rabbit's foot with him when he plays, he'll win,» L # W, to which someone else replies, «That's not so,» ~(L # W). Suppose then that ℜ tries the usual approach, asking questions concerning beliefs that would entail the contradictory of the conditional:

OP	That's not so.	~(L # W)
ℜ:	Do you mean it's possible that he'll have his lucky rabbit's foot with him when he plays and still not win — or do you mean that if he has his lucky rabbit's foot with him when he plays, he won't win?	♦(L∧~W)? (L # ~W)?
OP	Neither: I mean that his having his lucky rabbit's foot with him when he plays would have no effect at all on whether or not he wins.	

OP may be of the opinion that Jim is going to win whether or not he carries the

rabbit's foot. (OP may know, for example, that the contest has been fixed, or that the opposition is just no match for Jim.) Or OP may have no idea of how the contest is going to turn out — but may be certain that the outcome is not going to be affected by the absence or presence of a rabbit's foot. In this case, OP denies that L # W , because OP believes that the truth of L has no bearing on the truth or falsity of W. This, however, is not something that can be expressed in statement logic with or without augmentation by a possibility operator. So in this case ℜ will have to attribute neither the belief that ♦(L∧~W) nor the belief that L # ~W to OP on the basis of the latter's assertion that ~(L # W). ℜ could, of course, proceed to ask, «Do you think the truth of L is unrelated to the truth of W?» and note the answer as indicative of OP's belief.

And what about cases in which a robot encounters the negation of a conditional while following a line of reasoning itself? What if ℜ infers a negated conditional as in the example about the object that might be dropped?

1.	If it breaks if she drops it, it's not plastic.	(D # B) # ~P	Known
2.	But it is plastic.	P	Discovered
3.	So it's not the case that it will break if she drops it.	~(D # B)	1,2; DoubleNegat, MT

Without Material Implication, ℜ is in no danger of going from line 3 to D∧~B. The improper inference from lines 1 and 2 to the conclusion that she is going to drop the object does not go through. But is there nothing that ℜ can infer from line 3? I think the answer has to be «Nothing in statement logic or statement logic with modal operators.» With axioms other than statement-logic axioms, ℜ could get something like «D would not mean that B» or «The facts that would make D true would not also make B true.» Ultimately, of course, our robots will have to have such non-statement-logic axioms.

It should also be noted that robots using the material implication interpretation of negated conditionals and not having access to the axiom I am calling Chrysippus would be unable to make certain valid inferences. Suppose, for example, that ℜ hears from a reliable source that either Jones is lying or area two was contaminated if area one was. ℜ interprets this as J ∨ (O # T). Then ℜ learns from another reliable source that in fact it is possible that area one was contaminated but area two was not: ♦(O∧~T).

1.	Either Jones is lying or else area two was contaminated if area one was.	J ∨ (O # T)	Known
2.	It may be that area one was contaminated but area two wasn't.	♦(O∧~T)	Learned

The problem here is that ℜ cannot infer from these premises that Jones is lying. In order to get J from premise 1 by Disjunctive Syllogism on the material implication interpretation of negated conditionals, ℜ would have to have O∧~T , not just ♦(O∧~T). ℜ would have to know, that is, that area one had actually been contaminated while area two had not. But the mere possibility that area one could have been contaminated without area two being contaminated would certainly be enough to eliminate the second disjunct in 1 and prove Jones a liar. The Chrysippus axiom would, of course, do the trick, by permitting an inference from line 2 to ~(O # T).

Or suppose that in a similar case ℜ knows that if both Smith and Jones were at the meeting, then Roberts was too. And then ℜ learns that Jones might have been at the meeting without Roberts being there as well. So ℜ starts to reason…

1.	If Smith and Jones were at the meeting, then Roberts was too.	(S∧J) # R	Given
2.	It's possible that Jones was at the meeting but Roberts wasn't.	♦(J∧~R)	Learned
3.	If Smith was at the meeting, then Roberts was too if Jones was.	S # (J # R)	1; Exportation

On the material implication interpretation of negated conditionals, ℜ would need J∧~R to get ~S from 3 by Modus Tollens. But ℜ has only ♦(J∧~R), and so could go no further — without Chrysippus. With Chrysippus, however, ℜ can reach the humanly obvious conclusion:

4.	It's not true that Roberts was at the meeting if Jones was.	~(J # R)	2; Chrysippus
5.	Smith wasn't at the meeting.	~S	3,4; Modus Tollens

3. Material Equivalence

As the reader may have suspected, replacement of the two-way-entailment axiom Material Implication by the one-way-entailment axiom Deconditionalization will necessitate analogous alteration of one of the pair of two-way-entailment axioms known as Material Equivalence:

a) $(p \equiv q) \Leftrightarrow ((p \supset q) \wedge (q \supset p))$

b) $(p \equiv q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$

First of all, to indicate that the full meaning of «if…then» is involved, the horseshoe will be replaced by the pound sign. But then what about the triple-barred equal sign? It represents «if and only if» but is intimately connected, through this axiom, to the material implication interpretation of biconditionalization. For clarity, I propose replacing it with a double pound sign, ##, understood to represent the full ordinary meaning of «if and only if»:

a) $(p \,\#\#\, q) \Leftrightarrow ((p \# q) \wedge (q \# p))$

b) $(p \,\#\#\, q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$

Aside from these changes, the first line of the axiom does not need to be altered. It simply notes that saying «p if and only if q,» is equivalent to saying «If p then q, and if q then p.» The second line of the axiom, however, says that the assertion that p if and only if q is *equivalent* to the assertion that either both statements are true or both are false. But this would permit ℜ to make inferences such as the following…

1.	Amanda is in Brasil.	A	Known
2.	Juan is in Argentina.	J	Known
3.	Amanda is in Brasil and Juan is in Argentina.	A∧J	1,2; Conjunction
4.	Either Amanda is in Brasil and Juan is in Argentina or else Amanda isn't in Brasil and Juan isn't in Argentina.	(A∧J) ∨ (~A∧~J)	3; Addition

5.	So Amanda is in Brasil if and only if Juan is in Argentina.	A ## J	4; Material Equivalence

Presumably, we do not want ℜ reaching a conclusion like 5 from premises like 1 and 2. Knowledge that two statements are both true should not lead to the conclusion that one of them is true if and only if the other one is! But in the light of what has already been said about Material Implication, the solution is obvious: we simply reduce the second line of Material Equivalence to a statement of one-way entailment:

a) $p \#\# q \Leftrightarrow (p \# q) \wedge (q \# p)$

b) $p \#\# q \Rightarrow p \wedge q \vee \sim p \wedge \sim q$

This will effectively preclude the move from line 4 to line 5 in the current example. Should we then continue to refer to the axioms (a) and (b) as Material Equivalence? Perhaps — because of the way in which we are rejecting the material implication interpretation of «if-then» and «if and only if» — we should not. For simplicity, I shall refer to this pair of axioms as «Biconditionalization.»

4. Transposition

Troublesome as it may be, the Philonian (material-implication) interpretation of «if…then» is not the only source of potential difficulties for robots reasoning with conditional statements. Robots could also run into trouble when applying the traditional form of the axiom known as Transposition, $p \supset q \Leftrightarrow \sim q \supset \sim p$. In particular, problems could arise if a robot applied the axiom to conditionals having subcontraries as antecedent and consequent. When the antecedent and consequent of a conditional statement could both be true but could not both be false, Transposition can lead to trouble. For example, a robot knowledgeable about European geography could easily reason as follows…

1.	If he's not in Spain, he's in France.	~S # F	Learned from a reliable source
2.	If he's in France, he's not in Germany.	F # ~G	Known
3.	If he's not in Spain, he's not in Germany.	~S # ~G	1,2; Hypothetical Syllogism
4.	So if he's in Germany, he's in Spain.	G # S	3; Transposition

We cannot have our robots reasoning to impossible conclusions like that on line 4 — so something obviously has to be done. But what? Perhaps the best we can do is to block applications of Transposition like the one on line 4 by inserting a sort of «filter» into the axiom:

Transposition $((p \# q) \wedge \blacklozenge(\sim q \wedge \sim p)) \Rightarrow (\sim q \# \sim p)$

The effect of the conjunct with the general possibility operator will be to filter out cases in which antecedent and consequent are subcontraries. (For obvious reasons, conditionals in which antecedent and consequent are contraries will not be encountered as true premises.) Taking this approach will mean that Transposition cannot be written as a two-way entailment statement, but that should pose no particular problem. And the use of Transposition in this form would clearly block the step from line 3 to line 4 in the example just given.

Are there other problems with Transposition that could be prevented by giving robots

this axiom in the altered form? It does appear that the proposed «filter» — could also block certain absurd results that robots might encounter when using Transposition if they were told to interpret «q, even if p» as p # q. Suppose we tell ℜ that even if the hosts of some reception served fruit, they did not serve cherries. Without the filter, ℜ could proceed to reason as follows:

1.	Even if they served fruit, they did not serve cherries.	F # ~C	Acquired from a reliable source
2.	So if they served cherries, they did not serve fruit.	C # ~F	1; Transposition and Double Negation

But the conclusion on line 2 is the sort of nonsense we cannot permit. Using the restricted version of Transposition would, however, prevent the inference to line 2, since it is not possible that the hosts served cherries but did not serve fruit.

Moreover, precisely the same sort of problem could arise with a counterfactual «even if» conditional. Suppose ℜ knows that Jones is not in France and learns from a reliable source that even if he were in France, he would not be in Paris. It would be only too easy for ℜ to reason in the following manner:

1.	Jones is not in France	~F	Known
2.	Even if Jones were in France, he wouldn't be in Paris.	F # ~P	Learned from a reliable source
3.	If Jones were in Paris, he wouldn't be in France.	P # ~F	2; Transposit., and Double Negation

Here again, the absurd conclusion could be prevented by giving ℜ only the restricted version of Transposition. The incompatibility of P and ~F would keep Transposition from being applied to line 2.

But while the restricted version of Transposition can prevent the sort of problems indicated with «even if» statements, I do not think it is the best solution. I think in fact it would be preferable *not* to let our robots interpret «q, even if p» as p # q. The question is how we should have them interpret it. As a minimum, perhaps, we can tell them to take «q, even if p» to entail q. We can let them assume that someone who asserts an «even if» statement is asserting the truth of the part that is *not* preceded by «even if». (Because «even if» statements seem *not* to be ordinary conditional statements, it may be best not even to use the terms «antecedent» and «consequent» for their parts.) If a speaker says, «Even if they served fruit, they did not serve cherries,» we can have our robots think that the speaker has at least asserted «They did not serve cherries.» And if someone writes, «Even if Jones were in France, he wouldn't be in Paris,» we can have our robots take the writer to have endorsed the proposition that Jones is not in Paris.

But is there nothing more we can or should tell robots about the meaning of «q, even if p»? There *is* one thing we can add innocuously, although I am not sure that it will prove to be of great use. It is simply ~(p # ~q). We will then be telling our robots that «q even if p» at least entails q ∧ ~(p # ~q). However, as already noted, there is not much within statement logic that robots will be able to do with the negation of a conditional.

5. Conclusion

Here then are the «forbidden inferences» that we must declare off limits to our robots:

1) $(\sim p \vee q) \Rightarrow (p \supset q)$

2) $((p \wedge q) \vee (\sim p \wedge \sim q)) \Rightarrow (p \equiv q)$

And here are the alternate axioms that I have argued our reasoning robots will have to use:

Deconditionalization	$(p \# q) \Rightarrow (\sim p \vee q)$
	$(p \wedge \sim q) \Rightarrow \sim(p \# q)$
Conditional Contrariety	$(p \# q) \Rightarrow \sim(p \# \sim q)$
Chrysippus	$\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \# q)$
	$(p \# q) \Rightarrow \sim\blacklozenge(p \wedge \sim q)$
Biconditionalization	$(p \#\# q) \Rightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$
Transposition	$((p \# q) \wedge \blacklozenge(\sim q \wedge \sim p)) \Rightarrow (\sim q \# \sim p)$
Possibility-Certainty	$\sim\blacklozenge(p) \Leftrightarrow \blacksquare(\sim p)$
Actuality-Possibility	$p \Rightarrow \blacklozenge p$
	$\sim\blacklozenge p \Rightarrow \sim p$

Using these axioms, I submit, reasoning robots will be able to «get it right» when making inferences involving «if…then,» and we will consequently be able to trust their reasoning when conditional statements are involved.

Ronald A. Cordero
Department of Philosophy. The University of Wisconsin at Oshkosh
Oshkosh, Wisconsin, USA 54901
<cordero@uwosh.edu>