Logique & Analyse **139-140** (1992), 207-222

ON ESTABLISHING THE CONVERSE

Robert K. MEYER & Errol P. MARTIN

While it seems a little unfair to throw yet another brick at material implication (a.k.a. the "conditional"), and a little futile as well (bricks on the noggin having a way of just bouncing off its hard-headed defenders), we wish to report yet another ground to do so.(1) It is that material implication licenses a rather remarkable way to establish the converse of a given proposition. We shall see moreover that intuitionist implication is no better. And we shall report a valid core to the material anomaly, which relates a proposition to its converse even in quite minimal logics.

In a day now apparently passing, the school course in Euclidean geometry introduced everybody to axiomatic methods. While the course may have left something to be desired on the side of rigour, and while it may have survived as long as it did because of the prestige once enjoyed by classical studies, it at least had the merit of introducing one to applied logic in a preregimented form. These days, when boolean brainwashing begins in kindergarten, who will be able to distinguish the logic they use from that in which they have been instructed? There is a possibility of an Orwellian state in which people can no longer think clearly about their logic, because logical Newspeak has so tilted the use of common particles that the vocabulary in which we might have expressed what we really meant is gone. The widespread inclination to kick 'implies' upstairs to the metalanguage is an instance of what we have in mind. As a number of authors have shown (see, for example, Belnap [1975] and Curry [1963]), there is no force to the purported grammatical arguments ('Beware use-mention confusion' and the like) for the upstairs kick. Meanwhile, a very useful word, which English cannot spare, gets chopped from our formal grammar at the object level.

But Big Brother, though he has his partisans, has not triumphed yet. So, we recall that among mathematicians and others who are into such stuff, the demonstration of a central proposition is often accompanied by a proof of its converse. Moreover, at the level of implicational logic, the converse of a statement $A \to B$ is evidently and sensibly $B \to A$. This is so, too, at the

(1) See Routley, et al. [1982], Chapter 1, for a plenitude of examples, many of them due to Hunter.

level of logical practice.(2)

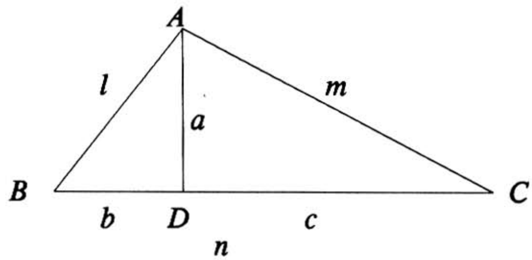
Let us then set up a little notation to deal specifically with this practice. We shall use A, B and C for arbitrary formulas, $\rightarrow$ for implication, and $\leftrightarrow$ for equivalence. We reserve I, J, K, etc., for formulas of the form $A \rightarrow B$. For the converse operation on implicational formulas, let us use ¢.(3) Thus ¢$(A \rightarrow B)$ is $B \rightarrow A$ by definition. Trying out our contextual notation, we can even report a few theorems. For example, ¢$I \rightarrow (I \rightarrow I)$ is not merely a classical tautology (by 'logical truth of consequent', as they say) but it is even a theorem of relevant logics like **R** in which one must use the premises to get the conclusion.

We shall have more to say on this theme. Meanwhile, let us leave off formal cataloguing, and return to intuitive exploration. We need an example. Having mentioned geometry, here is a typical one. A *right triangle*, naturally, is one whose largest angle measures 90°. A *Pythagorean triangle* has a side whose square is the sum of the squares of the other two sides. The *Pythagorean theorem* states that, if a triangle is a right triangle, then it is Pythagorean. Let us call this theorem just *PT*. Its converse, which we might as well call *TP*, states that if triangle *ABC* is Pythagorean, then it is a right triangle. Both because it will improve the reader's soul (cf. our opening remarks about the utility, or at least the humility, that attends trying to recall one's school geometry) and because we have announced that it is common and acceptable practice, let us use *PT* to derive *TP* as a corollary:

Let *ABC* be a Pythagorean triangle. We must show it a right triangle. We take for granted those geometrical facts that we can recall (which, unless they have recently been helping the kids with their homework, ought to be roughly those that readers recall). (It is definitely cheating to recall *TP*, or anything that implies it *too* obviously.) Here is a picture of *ABC* (a bit distorted, lest we draw the conclusion from the drawing.)

(2) While there is a use, in general relational algebra, for a converse operation on arbitrary inary relations, proving a converse usually just means reversing some implicational arrow.

(3) After all, inflation has rendered it an otherwise almost meaningless symbol.



We take *BC* to be the side whose square is the sum of the squares of *AB* and *AC*.

We may now sketch the proof of *TP* as follows. We use l, m, n for the lengths of *AB*, *AC*, *BC* respectively. So $l^2 + m^2 = n^2$. Now drop the perpendicular *AD* from *A* to the point labelled *D* on the line *BC*. We use a, b, c for the lengths of the segments *AD*, *BD*, *CD* respectively. (We note without proof that, since *BC* is the longest side, *D* must be between *B* and *C*. So $b + c = n$.) We next observe that, since *ADB* and *ADC* are both right triangles, we have by the *PT* that $a^2 + b^2 = l^2$ and $a^2 + c^2 = m^2$. So, collecting terms, $2a^2+b^2+c^2 = l^2 + m^2 = n^2$. But since $n = b + c$, we have also by the binomial theorem that $n^2 = 2bc + b^2 + c^2$. Subtracting equals from equals and then dividing by 2, this leaves $a^2 = bc$. Dividing again by ab, we have $a/b = c/a$. Trigonometrically speaking, this means that the tangent of the angle at *B* is the cotangent of the angle at *C*. It follows that the angles sum to 90°, whence the angle at *A* is 180° - 90° = 90°; i.e. *ABC* is a right triangle. Alternatively (we are, after all, supposed to be speaking geometrically), since the two angles at *D* are both right angles, and the adjoining sides of the triangles *BDA* and *ADC* are proportional, these two triangles are similar; but then it again follows quickly that the angles at *B* and *C* are complements, whence again (since the angles of *ABC* must sum to 180°), we have a right angle at *A*. Either way, we have established *TP* on its hypothesis, using *PT*.

That was fun. Moreover the logical form of the argument is unexceptionable. We have shown, given other salient facts, that if a certain proposition holds universally, so also must its converse. Moreover the proposition *PT* is known to hold. By the step that most people know as *modus ponens* the converse proposition *TP* certainly holds as well.

But wouldn't it be easier, and even more fun, if we didn't have to rely on Pythagoras (or one of those other old codgers) having done the hard part? Forgetting the part about universality (of which more later), this is what classical logic allows. For, amazingly enough, the following *Original Universal Converse Hypothesis* is a classical tautology, for all arrow statements I.

OUCH! $\text{¢I} \leftrightarrow (\text{I} \rightarrow \text{¢I})$.

In the left to right direction, OUCH! is just another implicational paradox. In the right to left direction, it is an instance of the tautology
(1) $((p\rightarrow q)\rightarrow(q\rightarrow p))\rightarrow(q\rightarrow p)$.
For those who wonder whether (1) is a tautology (somehow it doesn't look like one, and rightly so), we recall that $A \vee B$ is classically definable as $(A\rightarrow B)\rightarrow B$, whence what (1) comes to is just
(2) $(p\rightarrow q) \vee (q\rightarrow p)$
which has a familiar rationale: either p is false, and hence materially implies everything, or else it is true, and is materially implied by everything. Enough said!

Well here is something else for boolean apologists(4) to try to wiggle out of. At the sentential level, logic purportedly guarantees that, for every implicational proposition *I*, either *I* or its converse is true. While this is scarcely an original complaint, the modification that it induces in the geometry proof sketched above is a little too good to be true. According to OUCH!, whatever follows from its converse is so; whether or not the converse in fact obtains. So, in order to show that *A* implies *B*, one can help oneself to the assumption that *B* implies *A* as a free premise. For example, if we wish to show that if John is a man then he is a mammal, there is "no harm" (as they say) in initially assuming that if he is a mammal then he is also a man. According to common logical practice, this would suffice.

In our view, though, there is not much to be said for OUCH! It does not correspond to a reasonable way of invoking a proposition to prove its converse. To be sure, we might try to locate the unreasonableness in the already dubious principle (2). But then there is the case of intuitionist logic, which forswears (2), while still suffering OUCH! Intuitionist logic validates OUCH! this way: From left to right, faithful to its policy of preserving paradoxes of implication while abandoning the greater part of their motiva-

(4) Among whom we count ourselves, sometimes.

ion,(5) the argument for OUCH! is again immediate. For the converse, it turns out that (1) does not depend so much upon its classical equivalent (2) as one might have thought. For a proof with a "natural deduction" flavour, assume $(p \to q) \to (q \to p)$, and then assume q. It will suffice to show p on these assumptions. Appealing yet again to an implicational paradox (they are ubiquitous in how they support each other), we get $p \to q$ from q, whence by *modus ponens* we get $q \to p$ from our first assumption, whence by *modus ponens* again we get p. So it goes, providing an "independent argument" for OUCH! Which, naturally, also works classically.

Cannot any logic free us from this dreadful theorem? Our next try might be the semi-relevant logic **RM**. While **RM** is not, properly speaking, a finitely many-valued logic, it is possible (see Meyer and Parks [1972]) to view its implicational part as a 3-valued logic. Take the truth-values as T, N, F, with implication and negation defined truth-functionally on T and F. We complete the 3-valued specification of these particles by setting $-N = N$, and $N \to x = x$ and $x \to N = -x$ for all values x. Then the theorems of **RM** in $\to$ alone are just the formulas that never take the value F on any assignment of truth-values to sentential variables.

But among those theorems, alas, is OUCH!(6)

Note, moreover, that many-valued logics in the original style of Lukasiewicz are no better, since they notoriously have (1) as a theorem, even in the infinitely many-valued case.

It is time to try something else. How about a fully relevant logic like **R**? Or some modal logic, perhaps? Here, finally, OUCH! bites the dust. For there is a simple (Kripke-style "worlds") semantical argument that does it in, both for **R** and for **S5**. Since most relevant logics hitherto taken seriously are subsystems of **R**, and most familiar modal logics are subsystems of **S5**,

(5) This seems a regular intuitionist malady: intuitionism, for all its recent popularity as a constructive alternative to classical logic, manages frequently to perpetuate classical anomalies, while abandoning the truth-functional rationale that lends such anomalies such plausibility as they possess.

(6) It will suffice to show that any assignment of values to variables gives $q \to p$ the same value as $(p \to q) \to (q \to p)$. This is classically obvious if neither of p,q are assigned N; and since $N \to N = N$, it is also obvious if both are assigned N. If just p is assigned N, $q \to p$ reduces to $-q$ while $(p \to q) \to (q \to p)$ reduces to $q \to -q$, which a quick computation shows reduces to $-q$. A similar thing happens if q is assigned N, ending the argument that RM validates OUCH!

we may take OUCH! to have been refuted in these contexts. We shall give an **R** refutation; an **S5** refutation in the same style is immediate: Recall from Routley and Meyer [1972] that R-models require a ternary relation over worlds and an assignment of truth-values to formulas at worlds. We consider a model consisting of just two worlds, $\varnothing$ and b, with the ternary relation R defined so that $R\varnothing\varnothing\varnothing$, $R\varnothing bb$, $Rbbb$, $Rb\varnothing b$, $Rbb\varnothing$ hold, and otherwise R fails. We note in passing that this particular **R** model structure (actually an **R**+ model structure, since we do not need to define the Routley * operation here) is actually more familiar in a dual algebraic form, due to Church [1951].

We now wish to find an assignment of T,F to variables at each of our worlds $\varnothing,b$ so that OUCH! turns out false on the modelling of Routley and Meyer [1972]. We shall use 'Aw' to mean that formula A is true at world w on the interpretation to be defined, and '$\Rightarrow$' as a metalogical 'if'. The principal semantical clause governing the interpretation of $\rightarrow$ is the following:

(T$\rightarrow$) $(A\rightarrow B)x$ iff $(\forall y,z)(Rxyz \Rightarrow (Ay \Rightarrow Bz))$

In the special case where x is $\varnothing$, application of other postulates on an **R** model structure reduces (T$\rightarrow$) to the following Semantical Entailment Theorem:

SemEnt $(A\rightarrow B)\varnothing$ iff $(\forall x)(Ax \Rightarrow Bx)$

We now use SemEnt to falsify (1), and hence OUCH!, by making $(p\rightarrow q)\rightarrow(q\rightarrow p)$ true at $\varnothing$ but $q\rightarrow p$ false at $\varnothing$. Invoking (T$\rightarrow$), it will suffice for the latter to make q true at $\varnothing$ and p false at $\varnothing$. For the former, reverse the assignment at world b, making p true but q false. In view of $R\varnothing bb$ and $Rbbb$, and applying (T$\rightarrow$), the assignment to b assures that $p\rightarrow q$ will be false both at $\varnothing$ and at b, which, invoking SemEnt, assures that $(p\rightarrow q)\rightarrow(q\rightarrow p)$ is true at $\varnothing$. Thus OUCH! is falsified in the chosen semantic model of **R**, and is accordingly a non-theorem of that system.

Stripped to its essence, we have refuted OUCH! by a *universalizing move*. In the semantics for **R**, and even more so in the semantics of modal logics, $A\rightarrow B$ gets explicated as though it were of the form $(\forall x)(Ax \rightarrow Bx)$, pulling off what Meyer and Routley [197a] called the "Robin Hood ploy" by introducing the "phoney parameter" x. It is now time to take note of what many people who have written on implication have had in the back of their heads, although it tends to have been obscured by a lot of side issues. Despite having been charged with being confused about implication, we find that Whitehead and Russell are among the clearest and most coherent writers on this subject, when comes to the details that really count, and not the side issues. If you don't like material implication, these authors said, you will probably like formal implication better, where the former stands to the latter as $A \rightarrow B$ to $(\forall x)(Ax \rightarrow Bx)$. It is rather remarkable that almost everybody *has* liked formal implication better, while being reluctant to admit that what is liked better was in Whitehead and Russell all along.

For what is the sanitized notion, these days, of the classical doctrine of logical implication? It is that A entails B if whenever A is true, B is also true. What is this 'whenever'? Taken literally, if A and B really do stand in for English sentences, the recipe is nonsense. Sentences of English are either true or false, they are not true or false 'whenever' (the considerations that lead to temporal logics aside, which are not here at issue). So the sanitized classical doctrine only makes sense if our formal A's and B's are only in appearance stand-ins for English sentences. For the purpose of explicating logical implication, they are treated rather as if they were one-place predicates, which take an argument x. Then we can make sense of the 'whenever'. Consider the statement 'A entails B just in case whenever A is true in a model M, so also is B'. The explicating clause then turns out to be something of the form $\forall x(Ax \rightarrow Bx)$. But that, after all, is formal implication. And it is not due to Tarski.

Unfortunately, this classical program has not got implication right, even after sanitization. The ills induced by taking implication materially in the first place live on, even after the universalizing move. For example, not even making a suppressed $\forall$ explicit can save sentences that are always false from inheriting the magical properties that go with material $\rightarrow$. But the universalizing move is nonetheless one in the right direction. For it is characteristic of 'implies' statements that they are asserted as matters of law and not simply as matters of fact. We believe this is equally true of garden variety 'implies' statements, whose ground is not logical law but a law of nature or of morals. That logical implications are matters of law and not just matters of fact is signalled by the old saw that, if A is logically true, it is so as a matter of form. Put simply, the p's and q's and x's and y's that appear in logically true statements yearn to be bound to universal quantifiers.

All of this suggests that, perhaps, nobody has yet come up with a decent theory of logical implication. One candidate theorem for such a theory, if we take seriously this stuff about form, would be $A \to A'$, where A' is any substitution instance of A. For, if A holds in virtue of its form, it would seem obvious that its instance A' would thereupon hold in virtue of its form. (Evidently one could approximate such a theory by taking some standard theory, introducing propositional quantifiers, and defining a new formal $\Rightarrow$ by setting $A \Rightarrow B$ equal by definition to $\forall p_1 \ldots \forall p_n(A \to B)$, where $\to$ is the implication of the old system and $p_1, \ldots, p_n$ are all the sentential variables that occur in either A or B.)

With these thoughts in mind, let us return to our geometrical example, and think again about how the Pythagorean theorem was used to establish its converse. One would like to think of the argument as of the form $(R \to P) \to (P \to R)$, which we might put idiomatically as 'If to be right is to be Pythagorean, then to be Pythagorean is to be right'. Had our argument really been of this form, we would have been caught squarely in the OUCH! anomaly. Thinking classically, intuitionistically, or many-valuedly, the truth of $R \to P$ would have been indifferent to that of $P \to R$. But the fact is that when we *applied* $R \to P$, we *switched triangles*. It was the triangle *ABC* that we were trying to show a right triangle, having assumed it Pythagorean. But it was the triangles *ADB* and *ADC* that we constructed to be right triangles, thereupon concluding by the Pythagorean theorem that they were Pythagorean, and using that fact in our proof of *PR*.

So our argument would seem to have been one that established $\forall x(Rx \to Px) \to \forall x(Px \to Rx)$. Should we rest with that? Not quite, we believe. In the first place, how different are *ABC* and *ADB*? They are, after all, similar; which means that they are the same in the respects crucial for the proof. But let us waive that point. Our idiomatic rendering of $P \to R$ was 'to be Pythagorean is to be right'. If the reader will allow us to speak that way, it is an implication that stands between predicates, not between sentences. And one principle that seems to us important is this: *Logic is the same in all categories*. From a truth-functional and standard set-theoretic point of view, one gets just enough sense that this is the case to make the proposition sound appealing: corresponding to & as a truth-functional connective is $\cap$ as a set-theoretic operation; corresponding to $\vee$ there is $\cup$; and the laws relating these seem to be analogous. With negation things begin to diverge, while still retaining the feeling of sameness: the negation of a sentence is certainly a sentence, but the true complement of a set (that is, the collection of all things that don't belong to it) is only a set in highly non-standard set theories. And how, to continue, should we relate $\to$ and $\subseteq$? Even their grammar is non-analogous; as an operation on truth-values,

$\rightarrow$ yields a truth-value; but as an operation on sets, $\subseteq$ yields... a truth-value. Yet, so far as they go, the laws still stay the same: $A\&(B \vee C) \rightarrow (A\&B) \vee C$ still has the same flavour as $a \cap (b \cup c) \subseteq (a \cap b) \cup c$.

So it is not only programming language designers and set theorists and other metaphysicians who are divided on the question whether, in the end, everything should be treated as of the same category, or whether things split up in some ontologically fundamental way into stuff of different categories. Logicians also have this problem. And whatever there is really, it is at least useful to pursue both approaches. For example, Curry in his early work on combinatory logic made it typeless. But these days, there is a flourishing industry in typed combinatory logic. Conversely, the Whitehead-Russell solution to the logical paradoxes was to cause types to proliferate; in a sense, Gödel struck back with the reflection that things not intended as numbers (such as sentences and open formulas) can at any rate *be numbered*, unexpectedly vindicating the Pythagorean view that there is just one kind of thing, viz. numbers, in a way that produced new puzzles.

Amidst this jumble, there is a device which will produce some unifying understanding. It is what Whitehead and Russell called typical ambiguity — which, put otherwise, is just the principle that we enunciated above, that logic should be the same in every category.(7) So far as 'implies' is concerned, in particular, let us note that in English it does not merely stand between sentences. It stands equally well (and often better, as a matter of colloquial speech) between gerunds, infinitives, and nominal constructions of various sorts that signal a connection between properties. Being green does imply being coloured and that is all there is to it. To parse this as 'Always, x is green implies x is coloured' (or, among those who think it important, 'everything is such that if it is green then it is coloured') not only does violence to our native tongue, it loses the thought that universality is already built into 'implies', and what is universal need not be the logician's "all possible worlds" or "all models", it may simply be a connection between universals themselves.

Traditional logicians did not err, we think, in taking "Men are mortal; Greeks are men; so Greeks are mortal" and "Men are mortal; Socrates is a man; so Socrates is mortal" as instances of the same inference scheme in BARBARA. Nor do we think that any error was committed by those who

(7) After all, if there is only one category the principle evidently needs no defence; if there are many, we need to know what operates in all of them. And what is a more fitting candidate for this role than logic?

assimilated to this scheme the *hypothetical syllogism* "if q implies r then if p implies q then p implies r". They are all, at root, the same thing. It is a tragedy that we have been encouraged by modern logical theory to think of these kindred principles as exhibiting different logical grammar. We disagree. It appears to us that logic, that ancient and distinguished subject, has still a great deal of work to do. Far from its major problems having been solved (as we are so often assured they have been), the question is whether they have even been posed correctly.

At any rate, the impulse on which we acted initially — to parse our informal argument as establishing something of the form $(R \rightarrow P) \rightarrow (P \rightarrow R)$ — was not unreasonable. That is what, taken straightforwardly, it does establish, on the idiomatic translation suggested a page or two back. If readers are still inclined to balk about what they may take to be a mixing of logical categories, we can put the whole expression back into a vocabulary wholly sentential: we might read that sentence "That that *ABC* is right implies that *ABC* is Pythagorean implies that that *ABC* is Pythagorean implies that *ABC* is right". Where 'implies' is 'relevantly implies', we take this sentence to be true — not true as a matter of logic, but true as a matter of geometry, for reasons already sufficiently elucidated. And it does not collapse simply to the assertion of its second half, as it would either materially or intuitionistically.

However, the left to right half of OUCH! — which we have hitherto been treating as true by a paradox of implication — survives even in quite weak relevant and modal logics. For consider the following argument, in natural deduction style:

1.	$A \rightarrow B$	Hypothesis
2.	$B \rightarrow A$	Hypothesis
3.	$A \rightarrow A$	1,2, transitivity of $\rightarrow$
4.	$A \rightarrow B$	3,1, transitivity of $\rightarrow$
5.	$(B \rightarrow A) \rightarrow (A \rightarrow B)$	2-4, conditional proof
6.	$I \rightarrow (¢I \rightarrow I)$	1-5, conditional proof, taking I as $A \rightarrow B$.

All the steps are relevant, and the argument is unexceptional when put into one of the official natural deduction systems for relevant logics like **E**, **R**, and **T** — or modal logics like **S3**, **S4**, **S5**, for that matter. So, given that the hypotheses were used in the argument, $I \rightarrow (¢I \rightarrow I)$ isn't an implicational paradox after all.

But there is still one further point, something that caused some concern among some early commentators on the system **E**. This is the cycling and reiteration in the argument above, that leads from $A \to B$ back to itself. This is reflected formally by the fact that the hypothesis $A \to B$ is used twice, once to get step 3 and then to get step 4.

Now there are logics which do not accept this iterated use of hypotheses. The logics in question reject the "contraction" principle $(A \to .A \to B) \to .A \to B$, which in effect justifies the iteration at the axiomatic level. This principle corresponds to the **W** combinator so the logics are called **W**-free logics, and corresponding to each of the relevant systems **R**, **E** there is a **W**-free version **R-W**, **E-W**, etc.

Thus **R-W**, now called **C**; **E-W**; and **P-W** (known to some as **T-W**) do not accept this iterated use of hypotheses. In **C** and **E-W**, there are sneaky ways to prove $I \to (\text{¢}I \to I)$ anyway, but **P-W** and all of its subsystems are more demanding on the point. The implicational fragment of **P-W** may be axiomatised with these schemes:

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

We could complete the formalisation with the *modus ponens* rule $\to$E as sole rule, but it is more interesting to use instead an equivalent set of rules suggested by Dwyer. These rules correspond to the transitivity principles licensed by (B) and (B'). They are

(Rule B) $B \to C \Rightarrow (A \to B) \to (A \to C)$
(Rule B') $A \to B \Rightarrow (B \to C) \to (A \to C)$
(Rule BB') $B \to C$ and $A \to B \Rightarrow A \to C$

(The intended sense of the $\Rightarrow$ is that when what is on its left is a theorem (or theorems, in the BB' case), so the right side is also a theorem.)
While these three (Dwyer) rules make a certain amount of sense of the system **P-W**, in that proofs appeal only to principles themselves justified by the axioms of the system, we may alternatively formulate **P-W** with just modus ponens, i.e.,

(Rule $\to$E) $A \to B$ and $A \Rightarrow B$

Even weaker implicational logics have been of interest to relevant logicians, in two directions. The semantical direction explored in, for example Routley and Meyer [1973] considers the result of dropping even the axioms (B) and (B'), perhaps replacing them with alternative axioms, while saving the rules (B), (B'), (BB') and (E). This approach (not that it can't be continued further) hits a certain natural bottom in the system $\mathbf{B}_{\rightarrow}$, whose theorems are exactly the instances of axiom (I). (In that case all the rules mentioned are admissible, although uninterestingly so.)

The other direction of weakening of **P-W**, on which we wish to dwell a bit, drops the question-begging axiom (I), while retaining the axioms (B) and (B') and the corresponding rules. This system, which we have called **S** (for syllogism, since all its principles are syllogistic) has the marvellous property that its implicational theorems are exactly those of **P-W**, except that no instances of $A \rightarrow A$ are provable therein. Thus **S** begs no questions, a result we established in Martin and Meyer [1982], building in part on work of Powers and Dwyer.

How does (1) fare, then, in our very favourite system **S**? We note, to begin with, that we get a rule form of the argument from I to $¢I \rightarrow I$ even in this very weak system. All that is really involved in showing this are the three transitivity rules: Suppose that $A \rightarrow B$ has been proved in any system with these rules. Then by rule (B) we have $(A \rightarrow A) \rightarrow (A \rightarrow B)$ also as a theorem, and by rule (B') we have $(B \rightarrow A) \rightarrow (A \rightarrow A)$ as a theorem. Combining this pair of theorems by rule (BB') yields the conclusion $(B \rightarrow A) \rightarrow (A \rightarrow B)$. Thus the rule $I \Rightarrow ¢I \rightarrow I$ is admissible in any theory (which need not be a logic) that admits the three transitivity rules.

Note that this argument is somewhat different from our previous one. There, we assumed $A \rightarrow B$ on hypothesis, and employed conditional proof. The result was a theorem in the stronger relevant, modal, and standard logics. Here we are showing a certain rule admissible; to do so, we must have a theorem in hand, after which our argument appeals to other rules of the system in a categorical rather than a hypothetical sense. Nonetheless, the two sorts of argument are conceptually connected. For it is not merely the business of logic to determine its own theorems, but to signal the rules under which the theories that respect this logic are taken to be closed. Thus, in favouring a logic at least as strong as **P-W** or **S**, we take (B) and (B') as theorems of logic in order to record the insistence that all theories (not just special ones) may reasonably be expected to conform to the syllogistic transitivity principles recorded in the (B) and (B') rules.

Our attitude is a little different, though, about two other old saws — namely, the identity axioms (I), and the modus ponens rule ($\rightarrow$E). As for

$A \to A$, we see no good reason to take it as ever corresponding to a good inference (or any inference, for that matter). This gives a direction to our view of the relationship between provable arrow-statements of logics and rules admissible in corresponding theories. To be sure, if one wants to put it that way, the *rule* to infer A from A is always admissible, in any given theory. But this does not make it a good inference. That is, inferences give rise to rules, but there is no reason to suppose that every admissible rule is backed by a logically valid sentence of the form $A \to B$.(8) As for ($\to$E), it is extremely doubtful that this is a principle under which one wants to close arbitrary theories. Relevant logicians have already made this point where the $\to$ in question is material $\to$ (bringing on themselves a "disjunctive syllogism" debate, more distinguished on all sides — especially the other one — by heat than light). Our position, which has the virtue of both simplicity and self-evidence, is that whatever $\to$ one has in mind, one should trust *modus ponens* for this $\to$ in just those concrete theories for which it is trustworthy.(9) But, having done so elsewhere, Martin and Meyer [199a], we won't go on about this here either.

All of this leads us to make some final points about converses, with relevant logics particularly in mind, and **S** and **P-W** more particularly in mind. It is not difficult to see that, where its formal properties are concerned, our converse operator works to some extent like a negation operator. To begin with, ¢¢I is the very same sentence as I, for all sentences I. The standard theorem that most got our goat, (¢$I \to I$)$\to I$, has a form in which *reductio ad absurdum* is often put ("What is implied by its negation is true"). The other classical form of the goat-getter, I ∨ ¢I, is a kind of excluded middle. On the other hand, systems like **S** which resist that sort of stuff suggest another use for (¢$I \to I$), as a kind of modal operator that we might write as □I (recalling the modal definition of 'necessary A' as 'not-A implies A'). While we don't want to push that analogy too far, in these terms what we have just got through showing is that the rule to infer □I from I is after all admissible even in quite weak relevant logics. (In stronger logics, $I \to$□I is itself a theorem, which is one reason for not pushing the modal analogy.)

Now it often happens — and we dwelt on an example — that one wishes

(8) Those who hold otherwise tend to have already confused the two, a confusion otherwise known as "establishing semantic completeness".

(9) It is a little more intricate saying just what theories these are.

to prove both a given implication and its converse, and even to use one to help establish the other. In this case, having shown $A \leftrightarrow B$, we take A and B to be equivalent sentences, and to express the same proposition from the viewpoint of the theory at hand. We have one more philosophical demand to make; in this case, the theory at hand must never be logic. The sweet reasonableness of our demand comes from reflecting once more on the identity sentences $A \to A$. These sentences, and only these, are their own converses as a matter of logical form. Such sentences, corresponding as they do to a universally condemned fallacy while nonetheless being almost universally approved, can never be theorems of the One True Logic. And it is easy to see that this is an instance of a more general consistency requirement on ¢; namely, if I is a theorem of logic, ¢I must never be a theorem. (For we can just apply the transitivity rule to get $A \to A$ from $A \to B$ and $B \to A$.) Interestingly, the **P-W** problem as originally posed by Belnap (see Anderson and Belnap, vol 1 [1975]) sought a solution that imposed the general requirement except in the special case: Show that not both I and ¢I are theorems of (pure implicational) **P-W** unless I = ¢I. It is smoother and more coherent to ditch the exception, showing of **S** that it is ¢-consistent in all cases, of which the identity case is the crucial one. Showing this in Martin [1978] settled the original Belnap conjecture also.

The role of sentences $A \to A$ in **S** now gets rather tricky. From a conceptual perspective, they function simultaneously both as least truths and greatest falsehoods. (This is not accurate, strictly speaking, but it will do for an initial picture.) We can give this definite technical content. On the 3-valued metastory told in Martin and Meyer [1982] and Meyer and Martin [199a], it is natural to take all non-theorems of **P-W** as definitely false; and all theorems of **S** as definitely true. (It is also natural, since we have not told the reader all that is up our sleeve, to take **P-W** and **S** as the same system, viewed from slightly different perspectives.) The middle ground is occupied by those sentences that are theorems of **P-W** but not of **S**. These are exactly the formulas of the form $A \to A$. On our technical story, these turn out to have an intermediate truth-value (which, depending on one's inclination to be known for shocking or sound views, one can think of as both true and false, neither true nor false, or just neuter).

This semantic picture makes a certain amount of syntactic sense as well. Theorems of **S** never imply anything of the form $A \to A$. Conversely, every theorem of **P-W** (and hence of **S**) is implied by something of the form $A \to A$. What happens about non-theorems is a little less clear, save that it is of course sufficient to be a non-theorem of **S** if there is an $A \to A$ that one entails. But some non-theorems, just p for example, do not entail anything of the form $A \rightarrow A$. It would be interesting if **S** had the following property: whenever B is a non-theorem, there is some substitution instance B' of B which implies an $A \rightarrow A$. If the property held, this would complete the story on which the $A \rightarrow A$ mark the division point between logical sheep and goats. But, fearing a quick counterexample, we won't quite make that a conjecture. Finally, the relation among the $A \rightarrow A$ themselves is quickly told (cf. Meyer, Martin and Dwyer [1983]): $A \rightarrow A$ entails $B \rightarrow B$ in **P-W** just in case A is a subformula of B; and in **S**, just in case A is a proper subformula of B.

So, having begun with a geometrical proposition which, being true, collaborated in the proof that its converse is true, we have reached the conclusion that — at least at the level of sound implicational logic — this never happens. So, while we made our last territorial demand in logic a while back, we have thought of another one. Provable entailment, as a relation of pure logic, must be accorded its direction. It must never just wander around in circles, deriving a conclusion B from an assumption A when A might just as well be derived from B. That sort of aimlessness, getting nowhere except where one has already been, is of course consistent with the idea that the purported logical truths are all tautologies, and hence empty truisms. It is, note, equally consistent with the view that the purported logical truths are empty falsisms for the most part. But we have defended here an older idea, the Aristotelian opinion that, in pure logic, the conclusion of a valid argument must always be something other than its premiss. Accordingly, we beg no questions. To stay consistent, we never establish the converse of a logically true entailment.

Australian National University
University of Canberra

References

Anderson, A.R. and Belnap, N.D. Jr., 1975 *Entailment. The logic of relevance and necessity*. Vol. I. Princeton University Press

Belnap, N.D. Jr., 1975 *Grammatical Propadeutic*. In Anderson and Belnap [1975], pp. 473-492.

Church, A., 1951 'The weak theory of implication'. In: *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, ed. A. Menne, A. Wilhelmy and H. Angsil, Munich, pp. 22-

37.

Curry, H.B., 1963 *Foundations of Mathematical Logic*, New York, McGraw-Hill.

Martin, E.P., 1978 *Solution to the P-W Problem*. Ph.D. Thesis, Australian National University.

Martin, E.P. and Meyer, R.K., 1982 'Solution to the P-W problem.' *The Journal of Symbolic Logic*, vol. 47, pp. 869-886.

Meyer, R.K. and Martin, E.P., 199a *S for syllogism*, unpublished ms.

Meyer, R.K., Martin, E.P. and Dwyer, R., 1983 'The fundamental S theorem — a corollary.' *Notre Dame Journal of Formal Logic*, vol. 24, pp. 509-516.

Meyer, R.K. and Parks, R.Z., 1972 'Independent axioms for the implicational fragment of Sobocinski's three-valued logic.' *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 18, pp. 281-295.

Meyer, R.K. and Routley, R., 197x *Extensional Reduction II*, unpublished ms.

Routley, R. and Meyer, R.K., 1973 'Semantics of Entailment-I.' In: *Truth, Syntax and Modality*, ed. H. Leblanc, Amsterdam, North-Holland, pp. 199-243.

Routley, R. with Plumwood, V., Meyer, R.K. and Brady, R., 1982 *Relevant Logic and their Rivals, Part I*. Ridgeview Publishing, Atascadero, California.