Notre Dame Journal of Formal Logic
Volume 23, Number 1, January 1982

When is a Fallacy Valid?
Reflections on Backward Reasoning

ROMANE CLARK*

I catch the glint of light on metal through the trees by the drive, remark that I see the family car is there, and go on to infer my son is home. It may be said that taken literally I have misdescribed things. What I see, it may be said, is a flash of light through the trees. Strictly I infer, but do not see, that the car is there. C. S. Peirce was a philosopher who would have characterized it in this way. *All perception*, he thought, is inferential.[1]

I do not want to challenge Peirce in this, although I think it false. (Whatever the truth of the matter, the issue is complex, trading as it does on implicit views of the relation of sensation to perception, and on the relation between seeing things and seeing what is the case.) Let us suppose here that Peirce is right. A natural question then is this: What sort of inference is it when I say I see, but strictly I infer, the car is there? And what are the conditions of its validity and its soundness?

Peirce had an articulate answer. The inference is an *abduction*. Consider our example. It seems surely true that if indeed I infer, rather than see, that the car is there, this inference is very different from the inference that my son is home. The inference that my son is home has, presumably, a classical deductive form. It moves from the tacit, unspoken assumption: If the car is there, then my son is home; and the perceptual premise: the car is there; to the evident conclusion.

But my inference that the car is there cannot be like that. For in this instance I reason backward from what I see, the flash of light on metal, and my seeing it, to a cause the presence of which I believe to be sufficient to

*This paper is a slightly edited version of one presented at the University of Notre Dame, Fall 1980, and benefits from the suggestions of P. Bricker.

Received January 5, 1981; revised January 30, 1981

explain my experience. Knowing the situation, and knowing the way things look in circumstances like these, I infer that the car is in the drive.

This inference has a different, nonclassical, form. There is first, premise 1: p (the puzzling perceptual occurrence, the glint of light through the trees, whose happening motivates the consequent intellectual transaction); I bring to this premise 2: B (a conjunction of bits of relevant background knowledge and belief. B contains information ranging from judgments implicit in my level of visual sophistication in general to certain bits of happenstantial detail in particular: e.g., I know my son has the car this morning. For our example, B will in the simplest instances include a belief that, in circumstances like these and other things equal, the car's being in the drive would cause a flash of light through the trees like this.) I conclude from my premises, p and B, that h. I conclude that the family car is there, this being the hypothesis I draw the truth of which I believe is sufficient to account for that puzzling perceptual happening, p.

This pattern of reasoning is quite common. And it *is* after all a sort of *reasoning*. There is here a texture of structured thoughts leading to a conclusion. Moreover, there's something sensible about it. It is not just silly. But of course reasoning this way, I have sinned deductively. My reasoning is not deductively valid. (p and B might after all quite well be true and yet h false. Perhaps it is not in fact the car but a visiting neighbor's camper whose flash of light on metal I catch.) Peirce insisted that all creativity has its source in sin: reasoning of this general sort is the only creative form of inference. It is the only sort that yields as conclusions *new* hypotheses not covertly asserted in the premises; new hypotheses now to be tested and examined; hypotheses which may determine whole new lines of inquiry. This reasoning is, he thought, quite ubiquitous, present indeed in all perception but in nearly every area of contingent inquiry as well.[2] (It is philosophical commonplace, too. How frequently we reason backward from an epistemological puzzle to an ontological posit.) Peirce, in characterizing this backward, abductive, reasoning which runs from effects to hypotheses about causes sufficient to ensure them, has implicitly answered our title question. When is a fallacy valid? Answer: When it is a good abduction.

When then is an abduction a good one? When is one valid? When is one sound? For the most part Peirce and those who since his time have been captured by his suggestive remarks have turned from asking when abductions are logically sound to answering the question of how competing sufficient hypotheses are to be selected; to questions of their cost, novelty, or simplicity. (Philosophers intrigued with abduction have turned up papers with titles like "The logic of discovery", "The logic of creativity", and have gone on to discuss not the logic of abduction but the significance of nontrivial instances of it. Thus, N. R. Hanson's very interesting book *Patterns of Discovery* [6][3] pays fulsome tribute to Peirce and abduction. But it is not an analysis of abductive patterns. It would have been better titled *Discoveries of Pattern*, for it is in fact an historical tale of scientific discoveries of patterns.)

However, it is not difficult to say quite simply when an argument is abductively valid (valid^{ab}). Abductions we have seen have targets. Among their premises is one which records an occurrence or matter of fact which the conclusion is thought to ensure. Let us call the target premise of an abduction its *designated premise*. (p is the designated premise of the simple perceptual example which initiated our discussion.) We say that an argument is valid^{ab} when its conclusion (together with the remaining premises, if any), implies the designated premise. Schematically, we put it this way: $B, p \overset{ab}{\vdash} h$ if and only if $B, h \vdash p$. (Here, p is the designated premise and B is a set, possibly empty, of premises. We have in our earlier example, on one formalization of that inference, a valid abduction: p, h suffices for $p \overset{ab}{\vdash} h$. This is valid since h, if h then $p \vdash p$ in some standard logic. Throughout we use 'p' as a syntactical expression for some given, designated premise, 'h' for an abductive conclusion, 'B' and upper-case letters for sets of formulas.)

Evidently, for interesting applications there is here accordingly no general, determinate specification of abductive validity which can be given in advance. There is none since that will depend ordinarily upon the availability of ordinary deductive proof algorithms and these often are not to be had. Moreover, despite what Peirce seems to suggest, there evidently are uninteresting, noncreative, and trivial valid abductions. $B, p \overset{ab}{\vdash} p$ is one such. It exemplifies as well the further point that an abduction may, on occasion, have the form of a valid deductive inference. (In general, where a deductive inference yields a conclusion which entails one of its premises, we have a valid abduction with that premise as its designated premise. This fact, though trivial, is of some interest, as we shall see, and is often exploited in backward deductive proof searches.)

We may, if we wish, slightly tighten the characterization of abductive validity and rule out thereby certain unnatural, and trivial, inferences which otherwise are sanctioned. It is natural to require that an abduction, $B, p \overset{ab}{\vdash} h$, be such that the members of $B \cup \{h\}$ be consistent. This is natural, for abductions intuitively are meant to be vehicles for turning up reasons why a designated premise is true, and contradictory assumptions are not satisfactory *as such* for that purpose.[4] If we do so tighten our characterization of valid abductions, then of course abductive and standard pools of valid inferences will to that extent diverge (even when restricted to arguments with at least one premise).

Despite such qualifications, abductive validity as so far minimally characterized is a pretty trivial formal affair. Whatever its creative powers in psychological or other terms—those terms which captured the attention of Peirce and the imagination of his followers—abduction is so far a formally trivial matter. It emerges as no more than a kind of syntactical reciprocal of deductive validity. It is rather the need to characterize abductive soundness which forces the nontrivial nature of abduction on us. For it will not do to say that an inference is abductively sound (sound^{ab}) just in case it is valid^{ab} but with the members of $B \cup \{h\}$ all true. Schematically, abductive soundness requires more than just this: $B, p \overset{ab}{\models} h$ if and only if $B, p \overset{ab}{\vdash} h$, where h and the members of B are simultaneously true. Abduction, to be sound, seems to require that the designated premise, p, be true because of h. It requires at least, that if h obtains it will in the circumstances suffice to ensure that p. Evidently, the *truth* of h alone does not guarantee this. h must be true all right, but h must, in the circumstances, be operative as well. For an abduction to be sound, the conclusion must obtain and not be *overridden*; for a conclusion to be *uniquely operative* its designated premise must not be *overdetermined*. An abductive conclusion is uniquely operative relative to a designated premise when that conclusion is the case, is not overridden, and the designated premise is not overdetermined.

We say that h is *overridden relative to p* just in case h obtains, h is sufficient for p other things equal, but there obtains as well a condition, j, such that h and j do not ensure that p.[5]

We say that a designated premise, p, is *overdetermined relative to* an abductive conclusion, h, just in case there obtains, in addition to h, a condition, j, compatible with h and such that each of h and j is separately sufficient to ensure that p. h obtains, and would otherwise ensure that p in circumstances like these, but perhaps it is j on the occasion which brings it about that p; or perhaps each does.

A camper is parked alongside the car, there in the drive. I catch the glint of light on metal through the trees, and I would have done so even if the camper were not there. But the fact is that it is the camper's glint whose flash I detect. The car is there, and the car's being there is sufficient in the circumstances to create the effect I experience. Nonetheless, the flash I see is due not to the car, but the camper. The effect is *overdetermined*. Even though the car is there, if I abductively infer that it is *because* it is there that I experience what I do, I have reasoned unsoundly.

Or, a camper is parked alongside the car, there in the drive. I catch the glint of light on metal through the trees. Again I infer that the car is there. In this instance, however, not only is the glint I catch the camper's and not the car's, but I would *not* have experienced the flash were the camper not there. For although the car is there, my son, preparatory to a paint job, has masked the chrome and reflecting surfaces with tape. Again, h is true and, in the circumstances, h is sufficient for the truth of p, other things equal. But other things are not on this occasion equal. h has been overridden, the masking has ruled out what otherwise would in these circumstances be perceptually evident given the presence of the car in the drive. Again, if I infer in circumstances like these that the car is in the drive, I have reasoned unsoundly. The car's presence does not ensure my seeing what I do.

A sound abduction, then, requires a conclusion which is operative. It must obtain and not be overridden. What is not obvious is whether it must also be uniquely operative, and its designated premise not overdetermined. There is a distinction to be drawn between what is true *because* of another thing and what, in the circumstances, *ensures* that it be true. Accordingly, there is a stronger and a weaker version of abductive soundness available.

A *weakly sound* abductive inference is one with an operative conclusion. An operative abductive conclusion, h, relative to some designated premise, p, is one such that: (i) h obtains; (ii) h suffices for p in the circumstances, other things equal; and (iii) other things are equal.

A *strongly sound* abductive inference is one with a uniquely operative abductive conclusion. Such a conclusion, h, relative to a designated premise, p, must satisfy the conditions (i)-(iii) plus the condition (iv): h is in the circumstances the only condition which suffices for p. The designated premise of a strongly sound abductive inference is said to obtain *because* of its conclusion.

We seek now to characterize weak abductive soundness, specifying thus a relevant sense of "ensures". The qualifying clauses, "in the circumstances" and "other things equal", are crucial. Depending upon the particular occasion, a condition generally sufficient to bring about something may fail to do so. What seemed earlier a trivial variant of standard formal characterizations of deductive validity now seems, given the notion of weak abductive soundness, intractably nonstandard. We wish the relation of h to p, of abductive conclusion to designated premise, to be in general defeasible but, in the circumstances, undefeated. How are we to express this?

We have two main tasks. One is to characterize what it means to say that one thing ensures another. The second is to characterize what it means to say "other things equal". We need to know when other things are *not* equal; we need to know how the clause which expresses this modifies our characterization of "ensures".

The logic of "ensures" presupposes, I believe, a conditional logic. To say categorically and without qualification that one thing ensures another is to say two things. It is to say that the one thing obtains, and to say also that the other would be the case, were the one. It is this second conjunct which requires a logic of conditionals. These conditionals may, as here, be read subjunctively but if so not necessarily as contrary-to-fact. We formalize conditionals using the slash. We write: a/b, and we read, variously, "a's being the case would ensure b," or "b would be the case, were a," or "a is sufficient to bring it about that b".

The logic of the slash is perhaps best understood as a modal logic the semantics for which can be articulated in a version of so-called "world theory". I make no brief here for philosophical insights afforded by world theory, but it is a nice way to turn up certain formal properties of a system for conditionals. What is important here is that it makes available direct comparisons of semantic assumptions and theorems which flow from the characteristic axioms they validate.[6] Thus to each sentence containing our conditional slash, the formalization perhaps of an assertion about what ensures what, we now provide a unique transcription. This is a formula which no longer contains the slash and which is simply an expression of standard first-order logic now supplemented with some specific predicates and constants and with some special assumptions governing these. Logical truths containing the slash have transcriptions which are theorems in this extended, first-order logic.

The version of world theory exploited here is not one which is standard in the recent literature.[6] It turns on a three-termed relation, one of the form: '$Raww'$'. To assert something, a formalization of which is a/b, is to say something true just as the following transcription is so: $(w)[Raow \rightarrow Twb]$. To say that if a were the case, b would be, is to say that b is true of all worlds sufficiently like ours with respect to the proposition a. (In this formula, 'o' is a constant, taken to designate the actual world; 'T' expresses the "truth-in-a-world" predicate.) On plausible assumptions about R, (assumptions, e.g., that in all worlds sufficiently like a given world with respect to a proposition, that proposition is true; or the assumption that a world is sufficiently like itself with respect to any proposition true in it; on assumptions like these), we can generate a system of first-order theorems as transcriptions of statements expressing the logic of conditionality. In this logic, the transitivity of conditionality, as expressed with the slash, now fails as does contraposition. Nor can negation be freely confined across the slash: $\sim(a/b) \rightarrow (a/\sim b)$ is not a theorem; even if it is not true that a suffices for b, still it doesn't follow then that a suffices instead for $\sim b$.

Other pertinent results turn up. We learn, for example, that a kind of over-determination fails: $((a/b) \rightarrow [(a \ \& \ c)/b])$ is not a theorem. This seems right. If the match were struck, it would light, but that doesn't mean that if the match were struck and under water at the time, it would do so.

More notable perhaps are results like the ones to follow where the present system diverges from others like those of D. K. Lewis [8] or R. C. Stalnaker [11]. Joint truth, for instance, does not guarantee conditionality. It is not a theorem that $(a \ \& \ b) \rightarrow (a/b)$; from the facts that the cat is on the mat and Christmas falls on Thursday, we should not be tempted to suppose that were Christmas to fall this year on Thursday, the cat would be on the mat. Lewis however argues the other side of the question in developing his system of conditionals in which this is a theorem.

Again, it is not theorem here that $(a/b) \vee (a/\sim b)$, and no theorem that $[(a/(b \vee c)) \rightarrow [(a/b) \vee (a/c)]]$. These failures of theoremhood strike me as just right relative to occurrences of possibly unrelated matters of fact, most especially so with some sense of 'ensures' in mind. They are, however, theorems in other systems for the conditional. We may note here, anticipating some later comment, that on plausible assumptions we have $[(a \ \& \ b)/c \rightarrow a/(b/c)]$ as theorem but not its converse. Exportation, but not importation, holds.

The relative strength of a system for conditionals like this to others based on a fixed selection function which picks a world maximally similar to a given one is this: the theorems here are included there, but not conversely, while the theorems here include, but are not all included in, the set of those generated by a system based on the material necessity of conditionals based upon a fixed selection function.[7]

If the antecedents of the conditionals required in expressing the logic of *ensures* were always operative, a system like the one briefly sketched here might be a candidate for expressing what goes on in abductive reasoning. But of course these antecedents are not always operative. Often, perhaps usually, they are not. Typically one thing ensures another not only if it is the right sort of thing and actually occurs, but only if nature conspires as well. Other things in the circumstances must not interfere. Accordingly, our conditional requires qualification, a rider expressing that other things be equal. It requires that for the antecedent to be operative, it must not be overridden. Earlier, we said that a overrides b with respect to c just in case in the absence of a, b is sufficient for c but is not so in a's presence, i.e., $(((b \ \& \sim a)/c) \ \& \sim((b \ \& \ a)/c))$. If b is operative relative to c, there can obtain no such overriding circumstance, a. This is to say, $\sim(\exists a)(a \ \& \ ((b \ \& \sim a)/c) \ \& \sim((b \ \& \ a)/c))$, or equivalently, $(a)((a \ \& \ ((b \ \& \sim a)/c) \rightarrow ((b \ \& \ a)/c))$. If b is on the occasion operative, everything which obtains then is, in conjunction with b, sufficient to ensure that c. Let us abbreviate this formal clause, 'o' (for "other things equal"). We have then that b ensures c, other things equal just in case $b \ \& \ (b/c)$, given that o. Where, exactly, does the rider, o, ride the conditional it modifies?

We noted earlier that a stutter of slashes like $(b/(o/c)$ is implied by, but does not imply, $((b \,\&\, o)/c)$, and that it does not (lacking further, implausible assumptions about "sufficient similarity" across worlds) imply $(o/(b/c))$. Since these are different things to say, it matters where exactly 'o' goes. This is a matter of logical detail finer than my intuitions can discriminate. But, with some diffidence, I suggest we write: $(b/(o/c))$; and then read: if b were the case in this, our actual world, then (whether other things are in fact equal or not) if they were equal, c would be the case.

We have now a characterization of 'ensures': to say b ensures c, other things equal, is to say formally something like this: $b \,\&\, (b/(o/c))$. Given this, we have now the resources for a formal characterization of weakly sound abductions. The abduction, $B, p \overset{\text{ab}}{\vdash} h$ is weakly sound just in case the abduction is valid, the members of $B \cup \{h\}$ are true, h ensures p other things equal, and other things are equal, i.e., o is the case. Finally, an abduction, like that schematized above, is strongly sound just in case it is weakly sound and p obtains *because* of h. p obtains *because* of h just in case h ensures that p and is the only condition which, on the occasion, does so. Perhaps abductions are in fact only rarely strongly sound. If so, we are only rarely entitled to assert, simply, that one thing occurs solely because of another.

The earlier characterization of abductive validity, so incompletely sketched and so apparently trivial, can now be given a bit more substance and the notion of abductive soundness can be brought in line with validity in a more usual way. An inference, $B, p \overset{\text{ab}}{\vdash} h$, is abductively valid just in case $B, h \vdash p$ is valid in a suitable conditional logic, perhaps one like that used above to characterize the conditional conjunct of 'ensures'. Given this, we can restate weakly sound abductions as ones which are, in our present altered sense, abductively valid and such that the members of $B \cup \{h\}$ are true. Thus, in the simple example which initiated this paper, my inference that the car is in the drive is an abductively sound one if my conclusion is true and if, given the glint I catch, my background assumptions contain such truths as: other things equal, if the car were there, I'd experience just what I do, but, in fact, the circumstances are perceptually normal and other things *are* equal. The inference is an abductively sound one because the applied inference of conditional logic now metalinguistically characterized as having the form $(h/(o/p))$, $h, o \vdash p$ is itself a valid one with true premises.

Abduction, we know, is ubiquitous in philosophy, and the scandal of metaphysics. Russell, for one striking example, in his little introductory work, *The Problems of Philosophy*, races from epistemological problems to ontological hypotheses leaving us breathless and wondering if other posits might not have sufficed as well. For the relativity and errors of perception—for illusion, the facts of perspective, and hallucination—Russell hypothesizes our direct awareness of sense data; for the universality and necessity of a priori knowledge, he hypothesizes intuitive knowledge of universals; he posits facts to explain truth and error. Abduction is the scandal of metaphysics not because it is unreasonable, or even invalid, but because it is so often uncheckable. The reasoning may be valid but typically there is little evidence that it is sound. There is little ground for thinking that what may be sufficient to explain an epistemological puzzle is either necessary, or if not necessary, true and operative; there is little ground for thinking that it dominates other competing, sufficient explanations.

On the other hand, it is not so often noticed that abduction is itself also exploited in certain characterizations of deductive systems, or in proof searches, and can be employed with demonstrable soundness.

It is natural, at least for philosophy, to wonder what a proof is *for*? Why do we want or need one? One answer might be that a suitably simple characterization of deductive proofs should provide a kind of standard of formal evidence. A proposition is formally evident, we might say, just in case it is either self-evidently true, an instance of, say, the law of excluded middle, or else, if it is not, then it is the terminal entry in a sequence constituting what might be called a *Cartesian proof.*[8] The motivating idea behind Cartesian proofs is that, when one exists, it rests upon the logically simplest truths as axioms and proceeds in accord with the simplest rules of inference by the most discrete steps possible, to the demonstrated proposition. A proposition so demonstrated might be thought "clear and distinct", a formally evident one. A Cartesian demonstration is thus not only formally, but also epistemically and perhaps even psychologically, compelling.

Systems of abductive, backward reasoning can be exploited to turn up Cartesian proofs. This is evident in Gentzen-like adaptations of abduction for elementary, first-order logic and for modal and other extensions expressible there.[9] The abductive systems yield a proof algorithm for, and provide a useful vehicle for demonstrating the completeness and soundness of, first-order logic.

Given an arbitrary nonaxiomatic formula of first-order logic, an epistemological standard of Cartesian clarity of its formal intelligibility is that there exists a Cartesian proof of it. Accordingly, we embark on a search for such a proof. We ask, abductively, what would be sufficient, but simpler, to ensure this formula? On the answer, if it is not itself an axiom but of sufficient complexity to accommodate the question, we repeat the procedure. Otherwise the process terminates. Should the process turn up sequences of formulas, the disjunctions of members of which are instances of the law of excluded middle, we say that the original formula which initiated the process is a theorem. It is, if we assume the soundness of the system, a logical truth of first-order logic. The original formula has been reduced through a sequence of steps, each of which is sufficient to ensure the earlier stage from which it is immediately derived and its reduction has terminated with an axiom. Reductions like this are thus a sequence of abductive inferences each of which reasons from a designated premise to something which implies it of lesser formal complexity.

To obtain a proof algorithm from the process, it remains to specify determinate rules in specified patterns of applications for the reduction. To further obtain in this way a Cartesian proof, these rules must be themselves of maximal simplicity and discreteness.

It is probably not necessary or even desirable here to try to specify in any great detail the nature of an abductive search for Cartesian proofs. But to fix in a general way a sense of how the process goes, consider some formally complex sentence, p, expressible in first-order logic. If p is sufficiently complex it will not be obvious whether it expresses a logical truth. But since p is a logically complex sentence, a certain finite linear sequence of expressions, there will be some dominant logical operator or the negation of such within the scope of which the remaining expressions lie. The sentence is, we may say, "mainly a conjunction" or "mainly a negative existential" or some such. (If we were to adopt the convention of writing these formulas in Polish-like prefix form, then all occurrences of logical operators of larger scope in a formula would lie to the left of any of lesser scope which fell within their scope.) We seek now a set of abductive rules by which to reduce any such logically complex sentence, by attacking its dominant (in prefix notation, left-most) operator, positive or negative. There is one, but only one, rule for each *type* of positive or negative logical constant. Moreover, each rule applies to just *one symbol occurrence* of an operator, positive or negative, of dominant scope. In this way, there is exactly one unique rule for abductively reducing a conjunction, precisely one for negative existentials, and so on. Precisely one rule applies to exactly one logical constant at each point in the reduction.

The rules have further important characteristics. They specify how to reduce a given formal sentence to which the rule applies to certain sequences of sentences no member of which has any greater logical complexity than has the original, now reduced, sentence. They are moreover such that if any member of the resulting sequences is true, then so too must be the original to which the rule was applied. Each reduction step must be a valid abduction of maximal simplicity and discreteness, relative to the original sentence which, as its designated premise, is the target of the reduction. It must be a valid abduction, for the rules each guarantee that the disjunction of the reduction products classically imply the designated premise.

A trivial truth-functional example is this: suppose p is a material conditional, perhaps of the form $\rightarrow \& ab\ c$, in prefix notation. There exists then precisely one rule specifying that p is to be replaced by the sequence $\sim \& ab$, c. No sentence in this resulting sequence is more complex than p in the sense of involving more logical operators than p in its full expression. Further, if any sentence in the resulting sequence should be true, so too must be the conditional. (If $\sim \& ab$ should be true, so too would be $\rightarrow \& ab\ c$.) So long as the result of a rule application like this is a sequence some member of which is a logically complex sentence, the process may be repeated. In this instance, a negative conjunction, $\sim \& ab$, occurs, and so it may in turn be reduced. There is just one rule applying to negative conjunctions, which specifies that they be replaced in a determinate way. In this application, the rule applied to the sequence $\sim \& ab$, c specifies that this be replaced by $\sim a, \sim b$, c; a sequence which satisfies the requirements of simplicity and abductive validity relative to its designated premise which were mentioned above.

Thus, the general form of an arbitrary stage in a sequence of application of our abduction rules is this: A, p, B. A is here a sequence of formulas, perhaps empty, any member of which would be an atomic sentence or the negation of an atomic sentence. A contains the logically simple products of previous abductions, if any. p is the designated premise relative to the abduction to be performed at this stage in the reduction process. It is the left-most unreduced sentence of any formal complexity. In prefix notation its left-most symbol, positive or negative, determines the one rule which can be applied to it. B contains the formulas, if any, which remain as yet unreduced in addition to p. If the sentences of our logic are expressed in prefix notation, we move through sequences of formulas attacking as designated premise always the left-most sentence whose left-most operator dictates the application of the one rule which applies to it. If repeated applications of our rules yield at some stage a sequence of formulas to which no rules apply, the reduction process is terminated. At any stage, including terminal ones, one can determine *by inspection* of the finite number of members whether or not that sequence is *axiomatic*, that is, whether or not there occurs in it both a sentence and its negation. The disjunction of the members of such an axiomatic sequence is indeed literally an axiom, being an instance of the law of excluded middle. These are of course remarkably simple logical truths, and candidates for self-evidence. (We note that, of course, there are nonlogical truths whose reductions never terminate.)

Evidently if the sequences resulting from abductive reduction all, at some stage, turn out to be axiomatic, then the original reduced sentence must itself be a truth of logic. For the abduction rules each have the property that they pick out reduction products which transmit truth upwards to the stage from which they are abductively derived. The final axiomatic sequences of course must contain a true sentence on any interpretation.

Moreover, an abductive reduction of this sort supplies explicit instructions for the creation of a Cartesian demonstration of its originating sentence. The axioms of the deductive demonstration are the disjunctions of the axiomatic sequences which resulted from the reduction. Rules of inference are applied in order as indicated by reversing the steps in the reduction. Each abductive rule for these reductions has a matching, simple rule of deductive inference. Flipped over, each abductive step specifies a rule of deductive inference for the disjunctions which resulted. Turning over the abductive rule for the reduction of conditionals, as applied earlier in our example, we infer $\rightarrow \& ab\ c$ from $\vee \sim \& ab\ c$. Repeated applications of inference rules which appropriately match the rules of reduction and which are applied sequentially in reverse order will, in a finite number of steps, yield the original sentence which was reduced. It is in this way thus shown to be a theorem. (In practice, recognizing that all this is so, it suffices to establish the reduction to axiomatic sequences and then claim the hand. The winning tricks, the axioms, are spread there for all to see.)

Cartesian proofs afford a certain insight into the certainty, universality, and intuitive understanding which, as philosophers, we like to claim for truths of logic. The axioms which initiate such proofs are, we noted, of one simple, restricted type, the truth of which literally can be detected by inspection. Each is an instance of excluded middle. The paucity of axiom types is matched by a richness in primitive rules of inference of maximal simplicity. These rules convey not merely truth but epistemic conviction from premise to conclusion. There is a rule for each type of logical concept, and each rule applies to just a single type of logical concept (positive or negative). Moreover, each rule applies to but one symbol occurrence (positive or negative) of that type of concept. Finally, each application introduces in its conclusion the most limited possible complexity relative to its premises. (The matching reduction, we noted, never detours in its backward path through formulas of greater logical complexity than those to which the reduction rules are applied.) Cartesian proofs are thus demonstrations of maximal simplicity, discreteness, and clarity.

Students understandably wonder if the logically complex axioms of *Principia* with their nested clauses are logical truths, or if proof by *reductio* or by "separation of cases" is logically sound; the students' acquisition of the logical concepts expressed in the logical constants involved in these cases by no means alone settles the matter. By contrast, one who commands the logical concepts involved cannot understandably—and cannot justifiably—question an instance of excluded middle, and for the most part Cartesian rules of inference seem simply to define the logical concepts they govern. They are rather a standard by which command of the concept is determined. Failures here are not failures of application but of understanding; one lacks, not misapplies, the relevant notion.

A full characterization of Cartesian proofs for first-order logic would require further features, particularly features turning on the need to cycle through instantiations resulting from applications of the quantifier reduction rules. This can be done, but perhaps enough has been said to give the gist of the central point: that Peircean abductions, applied to standard logic, can be regimented to provide a proof algorithm of particular epistemological simplicity. Backward reasoning cannot only be shown to be sound by determinate, reasonable standards but being so, it can be used for constructive tasks such as searching out Cartesian demonstrations where they exist.

The trick in this case, of course, was to find a complete system of rules such that for each designated logically complex premise precisely one abductive rule was available. Accordingly, for each there is but one specific hypothesis which can be invoked which is logically sufficient to ensure the truth of the designated premise. No competition of hypotheses is possible here. A designated premise might be overdetermined, but its ensuring hypothesis could not be overridden. Quite by contrast, the perhaps unhappy fact is that in ordinary life, and in metaphysics, there may be many hypotheses which, if true, would suffice to explain some puzzling occurrence. And of course an hypothesis which does obtain, and does ordinarily suffice to ensure the truth of a designated premise, may in particular circumstances be overridden. So the abductions used to turn up Cartesian proofs are very special. In the general case, but not here, alternative sufficient hypotheses to a favored one may in fact be operative and other things may, in the circumstances, not be equal. I have much more confidence in the fact of the abductive basis for Cartesian demonstrations than I have in the supposed philosophical point of it. But in any case it does have nice formal properties, whatever its philosophical merit, and more to the point here, it is an articulate example of abduction at work in a determinate and evidently sound way. Despite the fact that abductive rules can be mapped into deductive ones, when stood on their heads, there is a creative point to the process of their regimented applications. Using them, we can *find* Cartesian demonstrations where these are possible. Using just their deductive images, we can evaluate such demonstrations, but scarcely know how to regiment and order their application effectively to turn up the available set of theorems. Deductive proof from premises is open-ended; abductive reduction is here determinate.

The general case, however, where there may be competition between sufficient hypotheses, where other things cannot be assumed equal, is the interesting and difficult case. Here soundness requires, if we are right, a nonstandard characterization of conditionality and sufficiency. It is, I believe, the difficulty in establishing the soundness of a particular abductive inference, not its validity, which makes perceptual, and metaphysical, abductions so very difficult but so very intriguing. The problems here are not so much centered on the reasoning as on finding plausible ways to screen and discriminate the truth from among competing explanations.

In particular, as an especial philosophical application and final example, it is perhaps worth remarking that this account of the nature of abduction is itself an exercise in abduction. We have reasoned backward from a puzzling fact—the widespread employment in philosophical inquiry of arguments which are deductively fallacious—to an attempt to characterize an adequate explanation of the phenomenon. We have tried to sketch minimal formal standards by which abductions can be evaluated as valid or sound, and their employment justified. I wish I could say more about what is important about abduction and the competition of sufficient hypotheses. I wish I could formulate an articulate formal *system* of abduction. But even a sketch like this is something. It seems to me at least to override an obvious competitor in explaining our ubiquitous use of these forms of inference; the view that these are just logical lapses—irrational applications of the fallacy of asserting the consequent.

NOTES

1. See, e.g., [9], Vol. V, paragraphs 181-192.
2. Cf. [9], Vol. V: paragraphs 189ff; Vol. II: paragraphs 636ff; Vol. VI: paragraphs 469ff.
3. Cf. also [12].
4. Abductions seem to require a relation between the hypothesis concluded and its designated premise which is stronger than that of classical implication but distinct from that characterized by "relevance logics" where entailments from contradictory assumptions are not proscribed. For relevance logics, see [1].
5. See, e.g., [3], esp. pp. 147-149.
6. I sketch and employ a system like this in [5]. Much of the detail is directly due to or results from discussions with Robert Binkley, with whom a joint presentation of the system was made at MacMaster University in Spring, 1969.

   When, with respect to a proposition, worlds may be thought to be sufficiently similar will depend, no doubt, on the content of the proposition. It is not something which could be fixed in advance for any proposition whatever. Thus, with respect to a proposition with contemporary economic import, it would be natural to suppose that worlds sufficiently similar would at least be ones consistent with the facts both stated and implied by the proposition which preserve those laws of economics compatible with the proposition. I don't think it counts against a version of world-theory like this that it cannot specify in advance and with generality the conditions of sufficient similarity with respect to any and all propositions, pertinent to any area of inquiry.

7. This is due to R. Binkley. See Note 6 above.

8. This discussion lifts from my [4], especially pp. 97-100.

9. See, e.g., [2], [7], and [10].

REFERENCES

[1] Anderson, A. R. and N. D. Belnap, *Entailment*, Vol. 1, Princeton University Press, Princeton, New Jersey, 1975.

[2] Binkley, R. and R. Clark, "A cancellation algorithm for elementary logic," *Theoria*, vol. 2 (1967), pp. 79-97.

[3] Chisholm, R., "The ethics of requirement," *American Philosophical Quarterly*, vol. 1, No. 2 (1964), pp. 147-153.

[4] Clark, R., "Rationality and the axiological orientation of the universe," *International Journal for the Philosophy of Religion*, vol. 1, no. 2 (1978).

[5] Clark, R., "Prima-facie generalizations," in *Conceptual Change*, eds., G. Pearce and P. Maynard, 1973.

[6] Hanson, N. R., *Patterns of Discovery*, Cambridge University Press, Cambridge, England, 1961.

[7] Kanger, S., *Provability in Logic, Stockholm Studies in Philosophy*, Vol. 1, Almqvist and Wiksell, Stockholm, Sweden, 1957.

[8] Lewis, D. K., *Counterfactuals*, Harvard University Press, Cambridge, Massachusetts, 1973.

[9] Peirce, C. S., *The Collected Papers of Charles Sanders Peirce*, Vols. I-VI, eds., C. Hartshorne and P. Weiss, Harvard University Press, 1931-1935; Vols. VII-VIII, ed., Arthur Burks, Harvard University Press, Cambridge, Massachusetts, 1958.

[10] Prawitz, D., *Natural Deduction, Stockholm Studies in Philosophy*, Vol. 3, Almqvist and Wiksell, Stockholm, Sweden, 1965.

[11] Stalnaker, R. C., "A theory of conditionals," pp. 98-112, in *Studies in Logical Theory*, ed., Nicholas Rescher, *American Philosophical Quarterly* Monograph Series, No. 2, Blackwell Publishing Company, Oxford, England, 1968.

[12] Weiss, P., "The logic of creative process," in *Studies in the Philosophy of C. S. Peirce*, ed., Weiner and Young, Harvard University Press, Cambridge, Massachusetts, 1952.

Department of Philosophy
Indiana University
Bloomington, Indiana 47405