COUNTERFACTUALS AND COTENABILITY

by

GARY TIMOTHY GLEB
B.A., University of British Columbia, 1975

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in
THE FACULTY OF GRADUATE STUDIES
The Department of Philosophy

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
August 1977

© Gary Timothy Gleb, 1977
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date August 29, 1977
Abstract

If philosophers had an acceptable theory of counterfactuals, counterfactuals would be an extremely useful philosophical tool. Unfortunately, analysis of the truth-conditions of counterfactuals has proved to be a difficult task. I examine Nelson Goodman's attempt in *Fact, Fiction, and Forecast* to develop a criterion of truth for counterfactuals, an attempt which ended with the discovery of a notorious problem, that of cotenability. This problem arises directly out of Goodman's inclusion of what is known as the "cotenability condition" within his tentative criterion. I explore in some detail the evidence that Goodman adduces in favour of this condition, and in doing so, argue that the cotenability condition as it stands is viciously circular. However, I also argue that this evidence is best seen as giving rise to two distinct problems, of which the solution to neither requires use of the cotenability condition. Resolution of the first, I claim, ultimately depends on the notion of explanation. With respect to the second, on the other hand, I contend that if the approach I suggest is borne out, then no modification of Goodman's criterion is necessary, other than that required to resolve the first problem.
# Table of Contents

1. **Introduction**  
   1

2. **Counterfactuals and Conditionals**  
   2

3. **Goodman and the Problem of Counterfactuals**  
   5

4. **On the Road to the Ideal S**  
   9

5. **Cotenability**  
   12

6. **The Problem of the S-related Counterfactuals**  
   15

7. **The Problem of the Undercutting Phenomenon**  
   22

8. **The Missing Link (1)**  
   26

9. **The Missing Link (2)**  
   34

10. **An Approach to the Undercutting Phenomenon Problem**  
    43

**Bibliography**  
   51
Acknowledgements

I would like to thank my supervisor, Michael Beebe, and Steve Savitt for their comments on preliminary drafts of this thesis, as well my wife, Rosslyn, for her care and patience throughout its creation.
1. Introduction

If philosophers had an acceptable theory of counterfactuals, counterfactuals would be an extremely useful philosophical tool. Unfortunately, the above is both true and counterfactual. Says Nelson Goodman:

>a solution to the problem of counterfactuals would give us the answer to critical questions about law, confirmation, and the meaning of potentiality.

([10], p.3)

However, the three decades since his own attempt, which ended in despair, have not produced a generally acceptable theory of counterfactuals.

Goodman's ground-breaking attempt to discern their truth-conditions ended with the discovery of an exceedingly difficult problem, that of cotenability ([10], esp. pp.13-16). Philosophers have disagreed over the nature of this problem, and have advocated different solutions, none of which, I believe, are successful (cf. e.g. Rescher [18]; Sellars [19]). Recently, others have abandoned Goodman's general approach for an analysis couched in terms of possible worlds and a notion of comparative similarity (Lewis[14]; Stalnaker[20]). However, this analysis appears afflicted with grave, even fatal problems arising out of the requisite notion of similarity (Bennett[3]).

In light of the difficulties facing this newer view, I suggest a reappraisal of the Goodman approach. Unlike Goodman, however, I have a more sanguine view of this matter: the cotenability problem,
I believe, can be solved. The key to this problem, I shall contend, is that it is the confused and misshapen product of lumping together two distinct problems, each of which, though formidable in its own right, appears soluble. A real advance, I argue, could be made if the two were pryed apart, thus exposing their distinct natures and preventing the matters raised by each from obscuring those of the other. Indeed, the core of my argument is directed towards this separation. However, in addition, I shall discuss what appear to me to be the most promising routes towards resolving each problem, although I do not claim, in either case, to have a full solution.

2. **Counterfactuals and Conditionals**

What are counterfactuals? They are conditionals of some sort, but what sort?¹ There is no simple answer to this question. It has long been recognized that many so-called "counterfactuals" do not literally deal in "counterfacts", that is, have false antecedents or consequents. (Chisholm[4], p.236). They do, however, carry some presupposition of belief by the utterer that at least the antecedent is false.

¹ My discussion here follows that of Lewis ([14], pp.3-4) closely.
Says Lewis:

Counterfactuals with true antecedents -- counterfactuals that are not counterfactual -- are not automatically false, nor do they lack truth-value ... Granted, the counterfactual constructions of English do carry some presupposition the antecedent is false. It is some sort of mistake to use them unless the speaker does take the antecedent to be false, and some sort of mishap to use them when the speaker wrongly takes the antecedent to be false. But there is no reason to suppose that every sort of presupposition failure must produce automatic falsity or a truth-value gap.

([14], p.3)

Another common feature is that counterfactuals are often stated in the subjunctive. However, while this mode of construction may be a sufficient feature of counterfactuals, that it is not necessary is illustrated by Lewis' non-subjunctive, "No Hitler, no A-bomb" ([14], p.4).

Complicating this issue is M.R. Ayers' contention that no philosophical interest resides in distinguishing counterfactuals from other conditionals. He argues that:

philosophers who are now still attempting to solve the so-called problem of counterfactuals or subjunctive conditionals by finding a special analysis for one or the other of these classes of statement, or by giving a special account of their significance or verification, would do better to widen the scope of their inquiry, and try instead to provide us with a philosophical description, which, if it applies to an empirical conditional, does so regardless of whether the condition is fulfilled or unfilled, and of whether the statement is expressed in the subjunctive or in the indicative mood.

([2], p.264)
Hence, if Ayers is correct, our discussion of cotenability should include some consideration of its relation not just to counterfactuals, but to conditionals in general.

The only way to refute such a claim -- that is, that there are no interesting differences between counterfactuals and, say, indicative conditionals -- is to find interesting differences. As Lewis notes, an example is supplied by Ernest Adams, who contends that of the following conditionals, the first, an indicative conditional, is likely true, and the second, a counterfactual, is likely false:

1. If Oswald did not kill Kennedy, then someone else did.
2. If Oswald had not killed Kennedy, then someone else would have.

([1], p. 90)

With Lewis, therefore, I conclude that:

there are really two different sorts of conditional; not a single conditional that can appear as indicative or as counterfactual depending on the speaker's opinion about the truth of the antecedent.

([14], p.3)

The task which remains -- spelling out a sound set of distinguishing marks for counterfactuals -- is not one I shall undertake. On the assumption that we have a solid grasp of counterfactuals, and that this intuition is philosophically justifiable, I shall proceed with a discussion of cotenability. Our initial set of characteristics -- that counterfactuals carry some presupposition of the antecedent's falsehood, and are usually expressed in the subjunctive -- though rough-and-ready, will have to serve as first approximation.
Following Lewis I shall use '\( \Box \rightarrow \) ' to represent the counterfactual connective and purely grammatical transforms thereof. It is read as "If it were the case that ..., then it would be the case that ....". I shall introduce other symbols and technical apparatus when necessary.

3. **Goodman and the Problem of Counterfactuals**

What is the problem of counterfactuals? Material conditionals are true if their antecedents are false; however, we often deny counterfactuals even though their antecedents are false, or are believed to be false. For example, we assert:

3. There is no moon \( \Box \rightarrow \) there are no solar eclipses.

and reject:

4. There is no moon \( \Box \rightarrow \) there are solar eclipses.

despite the fact that we know that "There is no moon" is false. Counterfactuals, therefore, are not truth-functional. What is the nature of their truth-conditions?\(^2\)

Goodman's guiding intuition is that "a counterfactual is true if a certain connection obtains between the antecedent and consequent" ([10], pp.7-8). This connection, as he sees it, can be analysed in inferential terms. That is, he believes it to be expressible in a truth-criterion of the following form:

---

\(^2\) This, of course, is not the only problem peculiar to counterfactuals. However, with Goodman, I shall ignore those special difficulties associated with counterfactuals with problematic antecedents, such as counteridenticals and countercomparatives (cf.[10], pp.5-6), and will stick to counterfactuals of a straightforward sort, e.g. (3) and (4).
a counterfactual is true if and only if its antecedent, some laws, logical or empirical, and a description of relevant conditions, permit the derivation of the consequent.

Goodman uses 'A' for the antecedent, 'C' for the consequent, and 'S' "for the set of statements of the relevant conditions or, indifferently, for the conjunction of these statements" ([10], p.9). With minor modifications, I shall follow these conventions. In addition let me introduce, as an expository device, the notion of the set of relevant conditions or situation -- call it 'K' -- in which the truth-value of a given counterfactual is to be assessed. Goodman's criterion, so rephrased, looks like this:

a counterfactual is true if and only if A, some laws, and some K-describing S, entail C.

Beside the conventions mentioned above, I shall follow these:

(i) 'S^i' will be used to represent explicitly individual members or, as the context demands, conjunctions of members of S; and
(ii) 'K_j' will be used to represent distinct situations.

As we shall see, not every K-description will be satisfactory, hence the criterion must be made complex. Various conditions will be added which must be met by any such description before it is accepted as counterfactual-supporting. For the moment, however, let us dub any acceptable K-description for a given counterfactual an 'S-ideal set', or an 'ideal S', to distinguish it from the imposters we shall be considering.

Three points to note. First, one should not assume the necessity of a non-empty S-ideal set for every true counterfactual: some may be true, as Goodman observes, because A and some laws alone entail C
([10], p.17). Call these 'S-independent' counterfactuals.

Second, one should not assume that a counterfactual, if true, is always supported by the same ideal S, that is, that it can be true in situations fitting only one description. It is apparent that counterfactuals may be true for different reasons in different situations. For example,

5. I throw this switch \( \square \rightarrow \) the light turns on may be true in one situation because the light switch is properly connected; and in another, because snapping the switch, though ineffective in itself, will cause, by pre-arranged signal, an accomplice in the darkness to make the real connection.

Finally, one should not believe that a true counterfactual is always supported by a unique S in any K. It seems likely that many non-equivalent S-ideal sets can be found in K if a counterfactual is true, if only because statements of irrelevant fact may be included in an ideal S. However, in the cases to be considered, the following will be assumed: if a counterfactual is true in situation \( K_j \) because of the existence of an associated non-empty S-ideal set, then it will be supposed that there is only one such set -- call it 'ideal S\(_j\)'. There is no reason to believe that this is always true, as I have suggested, but in assuming this I gain simplicity of exposition without, I believe, sacrificing certainty of result in the cases to be considered.
Besides employing a description of $K$, the A–C connection has another significant feature: it is law-dependent. From Goodman's discussion it is clear that he lumps under the category of laws both logical laws, or rules of inference, and natural, physical, or causal laws (henceforth disjunctively referred to as empirical laws). This is implied by his warning that in the usual case the A–C connection is not purely logical because some of the laws employed in the inference from A to C will be empirical in nature ([10], pp.8–9).

A peculiarity demanding explanation is that Goodman apparently countenances the use of empirical laws as rules of derivation. With regard to

6. Match $m$ is scratched $\rightarrow m$ lights

he remarks that:

The principle which permits inference of

That match lights

from

That match is scratched. That match is dry enough. Enough oxygen is present. Etc

is not a law of logic but what we call a natural or physical or causal law.

([10], pp.8–9)

The role attributed to the empirical law here is ambiguous: it could be either that of premise or of inference-rule. However, later on Goodman argues that:
the distinction between [the] connecting principles and the relevant conditions is imprecise and arbitrary; the 'connecting principles' might be conjoined to the condition-statements and the relation of the antecedent-conjunction (A*S) to the consequent thus made a matter of logic. ([10], p.17)

implying that the role initially attributed to the empirical law was that of inference-rule, not premise. This idea, that of treating empirical laws as inference principles, has been advanced in various forms by Schlick, Ryle, Dray, and Toulmin, according to Hempel ([11], pp.354-356). I shall duplicate Goodman's tolerance of this view, although in general I shall treat empirical laws as statements.

4. On the Road to the Ideal S

The completion of Goodman's analysis saddles him with two tasks: (i) discovering the conditions on an ideal S; and (ii) defining empirical laws ([10], pp.8-9). From our previous discussion (Section 3) it should come as no surprise that Goodman argues, as a preliminary step towards (i), that an ideal S contains only true members:

We do not assert that [a] counterfactual is true if the circumstances obtain; rather, in asserting the counterfactual we commit ourselves to the actual truth of statements describing the requisite relevant conditions. ([10], p.8)

On the other hand, an ideal S cannot always contain every statement true of K, since in cases where A is false "among the true sentences is the negate of the antecedent, so that from the antecedent and all true sentences everything follows" ([10], p.9). Indeed, as
it turns out, (i) is more complex a task than it first appears, for cases arise in which certain tentative K-descriptions (a) trivially permit the derivation of C or (b) support false counterfactuals.

The second task, one I shall not undertake, consists of discovering those characteristics which separate empirical laws, as well as a wider class of law-like generalizations, from non-law-like generalizations (cf. [10], pp. 17-27; Chisholm [5]). Instead, I shall take the concepts of empirical law and law-like generalization as undefined (although not undefinable).

After a terse but lucid survey of cases of kinds (a) and (b) Goodman advances the following criterion:

> a counterfactual is true if and only if there is some set S of true sentences such that S is compatible with C and \(-C\), and such that A·S is self-compatible and leads by law to C; while there is no set S* compatible with C and \(-C\), and such that A·S* is self-compatible and leads by law to \(-C\).

([10], p.13)

together with the proviso that:

neither S nor S* follows by law from \(-A\)

([10], p.13)

added in response to a counterexample raised by Parry in [17].

One will notice that Goodman's original intuition has become more elaborate, partially because of a logical feature of counterfactuals and partially because of problem cases falling under (a) and (b). First of all, not only must there be an inference-route from A through some K-describing S to C (call this the criterion's positive condition),

but no such route from A to -C (call this the negative condition).

Why this addition? As Goodman notes, counterfactuals of the form
\[ \phi \rightarrow \psi \quad \text{and} \quad \phi \rightarrow -\psi \]
are contraries: only one can be true in a given K, although both can be false. Hence we must take care that "our criterion not only admits the true counterfactual we are concerned with but also excludes the opposing conditional" ([10], pp. 12-13).

Secondly, the criterion employs the notion of compatibility, which requires explication. As mentioned, Goodman wishes to exclude any K-description which permits trivial derivation of C. One sort of trivial derivation, already described, is that which results when A·S entails P·-P. Another sort, the result of employing an S which entails violation of an empirical law -- call it 'L' -- occurs when A·S entails -L, or equivalently, when A·S·L entails P·-P ([10], pp.10-11). Hence, given Goodman's remarks, we can define the following:

(i) \( \phi \) is logically incompatible with S, and \( \phi \cdot S \) is logically \underline{non-self-compatible} if and only if \( \phi \cdot S \) entails \( P \cdot P \);

(ii) \( \phi \) is physically incompatible with S, and etc. if and only if \( \phi \cdot S \cdot L \) -- but not \( \phi \cdot S \) alone -- entails \( P \cdot P \); and

(iii) \( \phi \) is compatible with S, and \( \phi \cdot S \) is self-compatible if and only if \( \phi \) is neither logically nor physically incompatible with S.

Having discussed the rationale for the amendments to the criterion, especially the conditions on S, I shall presuppose their application but avoid their mention in what follows.
5. **Cotenability**

These modifications aside, the criterion is still defective. Its failure to reject certain K-descriptions which support false counterfactuals is both spectacular and notorious. The rock upon which the criterion founders is the rock of cotenability, and a durable rock it has proved to be.

What is the problem of cotenability? Goodman's presentation of it is brisk and laconic. It will profit us to see the problem as Goodman describes it, then to explore parts of that description in detail.

Says Goodman:

... many statements that we would regard as definitely false would be true according to the stated criterion. As an example, consider the familiar case where for a given match m, we would affirm

\[ \text{[6. Match } m \text{ is scratched} \implies m \text{ lights}] \]

but deny

\[ \text{[7. Match } m \text{ is scratched} \implies m \text{ is not dry}] \]

According to our tentative criterion, statement [7] would be quite as true as statement [6]. For in the case of [7] we may take as an element in our S the true sentence

\[ \text{Match } m \text{ did not light} \]

which is presumably compatible with A (otherwise nothing would be required along with A to reach the opposite as consequent of the true counterfactual statement [6]. As our total A*S we may have

\[ \text{Match } m \text{ is scratched. It does not light. It is well-made. Oxygen enough is present ... etc.} \]

and from this, by means of a legitimate general law, we can infer

\[ \text{It was not dry.} \]
And there would seem to be no suitable set of sentences $S^*$ such that $A\cdot S^*$ leads by law to the negate of this consequent. Hence the unwanted counterfactual is established in accord with our rule.

([10], pp.14-15)

He concludes:

the trouble is caused by including in our $S$ a true statement which though compatible with $A$ would not be true if $A$ were. Accordingly, we must exclude such statements from the set of relevant conditions; $S$, in addition to satisfying the other requirements already laid down, must be not merely compatible with $A$ but 'jointly tenable' or cotenable with $A$. $A$ is cotenable with $S$, and the conjunction $A\cdot S$ is self-cotenable, if it is not the case that $S$ would not be true if $A$ were.

([10], pp.15)

But this requirement is disastrous, since it demands that selecting $S$ (and presumably $S^*$ as well) requires determining the truth-values of an indefinite number of counterfactuals, involving us, as Goodman notes, "in an infinite regressus or circle" ([10], p.16). This is the cotenability problem.

Is this requirement necessary? The answer, I believe, is 'No!', but establishing this will require patient effort. For the moment, however, let us briefly consider Goodman's match argument. He notes that in the 'familiar case' we affirm (6) and reject (7). Why?

Let us examine our unwashed intuitions. With regard to (6), we know that in the right circumstances, that is, when match m is dry,

---

3 Why? Just as we do not wish to establish a false counterfactual with a non-cotenable $S$ which allows inference of $C$, we do not want to reject a true counterfactual on the basis of some non-cotenable $S^*$ which permits derivation of $-C$. 

oxygenated etc., scratching the match will cause or causally explain the lighting of m. It is upon this fact, which is causal in nature, that we affirm (6).

Why then do we reject (7) in the same set of circumstances? Clearly, that a match is oxygenated, scratched, and unlit does not cause or causally explain its 'non-dryness'. Some sort of causal fact, present for (6), is absent for (7), given the circumstances as described. This is not to deny, of course, that there are imaginable situations in which (7) would be true. One could suppose, in addition to the features of the 'familiar case', the presence of a special fire-prevention device, which, upon detecting the noise of a match being scratched, directs a powerful jet of water towards the noise's source. In this situation it is plausible to affirm (7); but here again we base this judgment on the presence of features by which the match's potential wetness could be caused or causally explained.

Therefore, if these insights are on track, the solution to the difficulty raised by Goodman's match argument, for at least some counterfactuals, is to be found in causation and related areas. Goodman, of course, does not take this route. Rather, he hopes to analyse the A-C connection in an inferential manner, using only notions such as logical and physical compatibility, as well as that of empirical law. However, since Goodman's criterion constitutes an analysis of counterfactual connection, and since the criterion establishes both (6) and (7) when it should only establish the first, it follows that the
conditions described in the 'analysans' cannot be sufficient for the 'analysandum', though their necessity is unchallenged. Some feature of the connection, some vital link, escapes the criterion as it stands. Our intuitions identify it as causation or some related notion, at least in the 'familiar case'; Goodman, on the other hand, identifies it as cotenability. In the next two sections, we shall examine precisely why Goodman is mistaken.

6. The Problem of the $S$-related Counterfactuals

Let us look closely at Goodman's match argument. Let $\phi$ be "Match $m$ is scratched", $\psi$ be "$m$ lights", $S^1$ be "$m$ is dry", $S^2$ be "$m$ is oxygenated", and $S^3$ be "$m$ is well-made". Moreover, let our general causal law, presumably:

$$L. \text{Scratching dry, oxygenated, well-made matches causes them to light}$$

be interpreted as:

$$L'. (x) (\text{Match } x \text{ is scratched } \land x \text{ is dry } \land x \text{ is oxygenated } \land x \text{ is well-made } \rightarrow x \text{ lights})$$

Our $K$, that is, the 'familiar case', is such that $\phi$ and $\psi$ are false while $S^1, S^2, \text{ and } S^3$ are true.

$L'$ can be regarded either (i) as the logical form of $L$, or (ii) as an entailment of $L$. The match argument works regardless of which view we adopt. If (i), the argument proceeds as described; if (ii), since (a) $L$ entails $L'$, and (b) $L$ is one of our true causal principles, we can simply infer $L'$, and the argument proceeds as above. This indicates that the argument itself, and therefore any solution, does not turn in any obvious way on a particular conception of the logical form of causal laws.

$L$ should actually be interpreted as: $L''. (x) (x \text{ is a match } \land x \text{ is scratched } \land x \text{ is dry } \land x \text{ is oxygenated } \land x \text{ is well-made } \rightarrow x \text{ lights})$. However, I employ $L'$ for simplicity of exposition.
Goodman assumes that:

6. March \textit{m} is scratched. \textbf{m} lights

is true in this K, implying that \{s^1, s^2, s^3\} constitutes an ideal K-description, and that there is no \(S^*\) permitting inference of \(-\psi\). However, as Goodman notes, since \(-\psi\) is actually true in K, \{s^2, s^3, -\psi\} is also acceptable to the criterion as it stands. Since \{s^2, s^3, -\psi\} and \(\phi\), together with \(L'\), entail \(-s^1\), and there is no appropriate \(S^*\),

7. Match \textit{m} is scratched. \textbf{m} is not dry

is established as well. Moreover, (7) is not alone. A number of equally acceptable K-descriptions, each formed through the union of \{-\psi\} and some subset of \{s^1, s^2, s^3\}, establish a range of bizarre counterfactuals in K: e.g.

8. Match \textit{m} is scratched. \textbf{m} is not oxygenated

9. Match \textit{m} is scratched. \textbf{m} is not well-made (?)

and even

10. Match \textit{m} is scratched. \textbf{m} is not dry or \textit{m} is not oxygenated or \textit{m} is not well-made (??)

(taking \{-\psi\} alone as the K-description), all of which are intuitively false.

In order to fully convey the implications of this argument, let me introduce the notion of an 'S-related' counterfactual. As we have seen, if a counterfactual \(\phi \square \rightarrow \psi\) is true in a given K, it is usually because some non-empty ideal S exists which, together with \(\phi\) and some laws, entails \(\psi\). Given such an S and \(\phi \square \rightarrow \psi\), we can construct an S-related counterfactual by (i) retaining \(\phi\) as
antecedent, and (ii) employing the negate of some $S^i$, where $S^i \in S$, as consequent. The general form of an S-related counterfactual of is:

$$\phi \square \neg S^i,$$

where $S^i$ is a member, or conjunction of members, of $S$. For any $K_j$ in which $\phi \square \psi$ is true because of the presence of an ideal $S_j$, we can construct an S-related 'family' of counterfactuals.

Of course, as we noted in Section 3, there may be many distinct situations in which $\phi \square \psi$ is true, hence there will be a distinct $K$-description $S_1 \ldots S_n$ for each such $K_1 \ldots K_n$, provided $\phi \square \psi$ is true for different reasons in every $K_j$. Thus $\phi \square \psi$ may possess many distinct S-related families, each associated with a particular $K_j$. However, this complication will be ignored in what follows.

The defect illustrated by the match argument may now be described as follows: if (i) $\phi \square \psi$ is true in $K_j$, (ii) $\phi \square \psi$ is not S-independent (is supported by some non-empty ideal $S_j$), and (iii) $\psi$ is false in $K_j$, then there exist criterion-acceptable $K$-describing sets -- formed through the union of $\{\neg \psi\}$ and some subset of the ideal $S_j$ -- which support each of the $S_j$-related counterfactuals in $K_j$. Unfortunately, the latter ought to be false, at least in the match example, and -- as we shall see -- in the general case as well. Let us call this the "S-related counterfactuals" problem.

What deficiency does this betray in the criterion? Why does the criterion support the true $\phi \square \psi$ as well as its $S_j$-related counterparts in $K_j$? Let us review Goodman's diagnosis. With respect to the
unacceptable (7) he says:

The trouble is caused by including in our S a true statement [i.e. \(-\psi\)] which though compatible with A would not be true if A were.

([10], p.15)

That is, \(-\psi\) should not be included in any K-description because it is not cotenable with \(\phi\).

It must be granted that there is something plausible about Goodman's remedy. Nevertheless, I believe it is wrong, although demonstrating this will require pulling the roots of this illusory plausibility painstakingly, one-by-one.

How well does the cotenability condition fare as a solution to the S-related counterfactuals problem? As we have already observed, Goodman admits by adding this condition to the criterion it appears we find ourselves in a regress or circle because "we can never explain a counterfactual except in terms of others" ([10], p.16). However, he doesn't quite close the door on this issue: he considers the possibility that the criterion could be revised:

so as to admit first those that depend on no conditions other than the antecedent, then use these counterfactuals as the criteria for the cotenability of relevant conditions with the antecedents of other counterfactuals, and so on.

([10], p.17)

The motivation behind this suggestion, however, is unclear. That is, Goodman may perceive the problem facing himself in at least two ways, each of which admits a different interpretation of this suggestion.

(i) On the one hand, he may not be convinced that, in the final analysis,
the cotenability condition as it stands produces a **vicious** circle
or an **infinite** regress, despite appearances to the contrary. That is,
he may believe that given the cotenability condition as it stands,
checking for cotenability involves no circle, but only an arbitrarily
long regress, terminating with S-independent counterfactuals, a fact
to be made apparent through an appropriate revision of the criterion.
(ii) On the other hand, Goodman may be certain that the cotenability
condition as it stands produces vicious circularity or infinite regress,
hence that any solution to the S-related counterfactuals problem on the
basis of cotenability must involve a radical overhauling of the notion
of cotenability as well as his treatment of counterfactuals.

Is (i) an open alternative? The answer is "No": it can be shown
that the cotenability condition as it stands produces vicious circularity.
Why? Let us first determine what cotenability amounts to. If Goodman
actually adheres to (i), then he is convinced that judging the truth-value
of a counterfactual depends, in part, upon determining that all the $S^i$
in S are cotenable with A: that is, that for all $S^i$ in S, it is not the
case that $S^i$ would not be true (or the case) if A were true (or the case).
Now, ignoring for the moment "It is not the case that", "If A were the
case, then $S^i$ would not be the case", is logically equivalent to,
"If A were the case, then \(-S^i\) would be the case". Hence the cotenability requirement is equivalent to:

for all \(S^i\) in S, it is not the case that \(-S^i\) would be the case if A were the case

or, employing the \(\square \rightarrow \) symbol:

for all \(S^i\) in S, it is not the case that A \(\square \rightarrow -S^i\).

The requirement, in other words, amounts to demanding that the S-related counterfactuals of \(\phi \square \rightarrow \psi\) be rejected as false before \(\phi \square \rightarrow \psi\) is accepted as true.

If there is any doubt about this, consider the following. When discussing the law of excluded middle, that is,

Either S or not-S

for, say, "Jack fell",

Either Jack fell or Jack did not fall,

we assert:

(i) If it were not the case that Jack fell, then it would be the case that Jack did not fall

and

(ii) If it were the case that Jack did not fall, then it would not be the case that Jack fell.

But (i) and (ii) entail:

(iii) It would not be the case that Jack fell if and only if it would be the case that Jack did not fall

provided nothing lurks in the innocuous-looking grammatical moves from "were" to "would" and vice-versa. This suggests that

(iv) It would not be the case that S if and only if it would be the case that not-S

is analytic.
Cotenability's vicious circularity can now be demonstrated. Remember that in the match example \( \phi \) stands for "Match m is scratched", \( \psi \) for "m lights", \( S^1 \) for "m is dry", \( S^2 \) for "m is oxygenated", and \( S^3 \) for "m is well-made". (i) In determining the truth-value of \( \phi \square \rightarrow \psi \) in the 'familiar case', given the criterion as it now stands, we must establish that, besides meeting other requirements, all the members of \( \{S^1, S^2, S^3\} \) are cotenable with \( \phi \), before we can accept this K-description as S-ideal. (ii) However, establishing cotenability is equivalent to rejecting as false those S-related counterfactuals constructed from \( \{S^1, S^2, S^3\} \), that is (7), (8), (9), (10) etc. (Cf. pp. 15-16) (iii) In turn, rejecting those counterfactuals in the 'familiar case' means demonstrating that certain K-descriptions which support them are illegitimate because, as Goodman observes, they include a certain truth, that is, \( -\psi \), among their members which is non-cotenable with \( \phi \). (iv) However, since demonstrating that \( -\psi \) is cotenable with \( \phi \) would mean showing that it is not the case that \( \phi \square \rightarrow (-\psi) \), demonstrating that \( -\psi \) is not cotenable with \( \phi \) means showing that it is not the case that (it is not the case that \( \phi \square \rightarrow (-\psi) \), or equivalently, that \( \phi \square \rightarrow \psi \). We have come full circle, and the circle is vicious: it cannot be broken by any S-independent counterfactual.

What about alternative (ii)? Although I admit that the argument offered above does not rule out, in any straightforward way, this interpretation of Goodman's suggestion, that is, the possibility of radically overhauling the cotenability condition and Goodman's treatment
of counterfactuals in terms of some step-by-step procedure based on S-independent counterfactuals, I am afraid that I must agree with Goodman when he says that "this idea seems initially rather unpromising in view of the formidable difficulties of accounting by such a step-by-step method for even so simple a counterfactual as

If the match had been scratched, it would have lighted"

(10], p.17).

7. The Problem of the Undercutting Phenomenon

We have seen that the cotenability condition as it stands places hopeless demands on the criterion. As a solution to the S-related counterfactual problem, therefore, its employment seems akin to using an H-bomb to capture a rabbit: just as the H-bomb would destroy the rabbit, the original cotenability condition destroys the possibility of determining the truth-value of a given counterfactual. However, as we saw earlier, our intuitions tell us that in the match example, the S-related counterfactual (7) is false while (6) is true because the former lacks in full the underlying A-C connection possessed by the latter. Indeed, in many cases this appears to be a feature of the S-related counterfactuals of a given φψψ when φψψ is true. For example, suppose that Martin, a friend, leaves the key in the ignition of his Alfa Romeo. Provided that the Alfa is in working order, it is

7I thank Steve Savitt for pointing this out to me.
true that:

11. The key is turned \[ \square \Rightarrow \] Martin's Alfa starts 

and false that:

12. The key is turned \[ \square \Rightarrow \] the Alfa's starter malfunctions since turning the key, in this situation, would start the car but not bring about a starter malfunction. On the other hand, if our intuitive judgments about (6) and (7), or (11) and (12), were based on the original cotenability condition it is impossible that they could be made at all. In view of the formidable difficulties facing the cotenability approach, therefore, let us set it aside, and turn instead to a search for the causal feature possessed by (6) and (11) but lacked by (7) and (12). For convenience, let us dub this feature of the A-C connection underlying (6) and (11) the "missing link".

That there is something plausible about Goodman's introduction of the cotenability condition was noted earlier (Section 6). However, given the relatively straightforward observations made about the causal nature of the missing link in Section 5, this residual plausibility remains somewhat mysterious.

The source of this plausibility, I believe, is a feature of the S-related counterfactuals I have hitherto avoided discussing, a feature that Goodman may have observed, but failed to identify clearly. Indeed, this feature is such that it appears to force use of the cotenability condition independently of any consideration of the S-related counterfactuals problem or of the missing link. What is it?
We noted that in the familiar case (7) and the rest of the S-related counterfactuals are false, while (6) is true. Indeed, isolating the basis for this fact and modifying the criterion accordingly is the gist of the S-related counterfactuals problem. However, we have also observed that in other, less familiar situations, at least some of the S-related counterfactuals would be true. An example of this was our fire-prevention device case: in that situation (7) was true. But notice this as well: in the same situation it appears correct to assert that (6) in this situation can be explained this way:

"Since the fire-prevention device is present, if the match were struck, it would be wet. But wet matches normally can't light, even if they are scratched. Hence, it is not the case that if the match were scratched, it would light".

The upshot of this is that in certain situations the truth of (7) "undercuts" the truth of (6). Does this "undercutting" phenomenon hold for other counterfactuals as well? Remember the situation discussed earlier in which we decided (11) is true and (12) is false. Imagine that this situation is now altered in the following way: suppose that an envious acquaintance places a small bomb on the Alfa's starter, wired to the ignition switch. In this new situation we would assert (12) and deny (11), reasoning thus: "Since the bomb will go off when the key is turned, destroying the starter, if the key were turned, the Alfa's starter would malfunction. But cars without starters cannot turn over. Hence, it is not the case that if the key were turned, the Alfa would start". Consideration of this and other cases, I believe, will bear out
the following contention: in many situations K, if an S-related counterfactual of $\phi \Rightarrow \psi$ is true, then its truth undercuts the truth of $\phi \Rightarrow \psi$. In Section 10 I shall attempt to provide an analysis of this phenomenon, in the course of which I shall try to isolate some of the features of the situations in which it can occur.

Our original problem, as demonstrated by the match argument, was that the criterion automatically establishes e.g. (7) when it establishes the true (6), a problem we hope to solve not with cotenability, but by finding the missing link. However, we have discovered a feature of S-related counterfactuals which, independent of our first problem, threatens reinstatement of the cotenability condition: the undercutting phenomenon. Why? The undercutting phenomenon entails that in many situations K, if an S-related counterfactual of $\phi \Rightarrow \psi$ is true, its truth undercuts the truth of $\phi \Rightarrow \psi$. Hence, when determining the truth-value of $\phi \Rightarrow \psi$ in any such K, we must ensure that its S-related counterfactuals are false. One method of ensuring this is Goodman's: that is, the cotenability condition, at least in its original form. We have in this, I believe, an explanation of the original cotenability condition's intuitive plausibility, despite its absurdity as a solution to the specific difficulty raised by the match argument. Let us call our new difficulty the problem of the "undercutting phenomenon".

Are we therefore stuck with cotenability? Is there no way of determining that S-related counterfactuals are false without stooping to "explain a counterfactual ... in terms of others". The answer to both
these questions, I believe, is "No": the approach to take with this new problem is quite simple, and the greater portion of it is easily solved; but more about this later. We shall return to this problem in Section 10, after devoting two sections to a discussion of the S-related counterfactuals problem.

8. The Missing Link (1)

Goodman's leading insight about counterfactuals is that they are supported by an antecedent-consequent (A-C) connection. As we have seen, this connection cannot fully be analysed by the stock of notions Goodman wishes to employ: some important feature, rooted in causation, escapes the criterion. This we dubbed the 'missing link'. I shall contend that this link amounts to a residual form of A-C connection which the Goodman criterion, given its limited set of concepts, cannot capture. Nevertheless, I think it can be explained in terms of inference-routes from A to C. For now, however, I shall refer to it as the causally-grounded, or 'causal' A-C link.

Presumably, the causal aspect of this connection has something to do with the use of causal laws in the inference-route from antecedent to consequent. Let us define as 'causal counterfactuals' those which, when true, are supported by inference-routes employing only (i) logical and (ii) causal laws (as opposed to other kinds of empirical laws and law-like generalizations). Therefore, it is for causal counterfactuals that the causal link is yet unanalysed.
Not all counterfactuals are causal, however. Does the S-related counterfactuals problem afflict them as well? Let us explore an example. Suppose that an application form provides two blanks, the first for one's Christian name, and the second for one's surname. I have written "John" in the first blank, but nothing in the second. Therefore, it is true that:

13. I write "Black" in the second blank \( \rightarrow \)
I have written "John Black".

The relationship between antecedent and consequent here is not causal: writing "Black" does not cause, or causally explain, the writing of "John Black". However, by admitting a certain truth — that I have not written "John Black" — into a criterion-acceptable situation-description, as in the match case, it is apparent that the false:

14. I write "Black" in the second blank \( \rightarrow \)
I have not written "John" in the first blank

can be established as well.

As we have seen, characterizing the missing link will resolve our problem for causal counterfactuals alone. Once this task is complete, therefore, we must consider the prospects for a general solution.

How are we to understand the nature of the missing causal link? The word "cause" often means the production of one event by another. Can the link underlying e.g. (6) but not (7) in the 'familiar case' be explained in terms of this sense of "cause"?

It will be remembered that, in our discussion of Goodman's match argument (Section 6) we saw that both:
6. Match \( m \) is scratched \( \square \rightarrow m \) lights

and

7. Match \( m \) is scratched \( \square \rightarrow m \) is not dry

are supported by sets of \( K \)-describing sentences -- call them '\( S^I \)' and '\( S^H \)', respectively -- each of which is equally acceptable to the criterion as it stands. \( S^I \) is \{"m is dry", "m is oxygenated", "m is well-made"\}, and \( S^H \) is \{"m is oxygenated", "m is well-made", "m does not light"\}.

The suggestion under consideration amounts to this. Given the causal laws of our world, we know that, with respect to (6), (i) any possible event described by \( A \cdot S^I \) -- that is, "Match \( m \) is scratched \( \cdot m \) is dry \( \cdot m \) is oxygenated \( \cdot m \) is well-made" -- causally produces the lighting of \( m \). On the other hand, we know that, with respect to (7), (ii) any possible event described by \( A \cdot S^H \) -- that is, "Match \( m \) is scratched \( \cdot m \) is unlit \( \cdot m \) is oxygenated \( \cdot m \) is well-made" -- cannot by itself causally produce \( m \)'s wetness.

How can this insight be adapted to the inferential nature of the criterion? Let us divide the classes of all causal statements into two kinds: causal reports, and causal laws. Causal reports, which I shall call simple causal statements, are appropriate for the description of instances of causal relations. Examples of these are:

15. The scratching of match \( m \) caused it to light

and

16. The fact that match \( m \) was scratched caused it to be the case that \( m \) lights
(Note: we shall see later that (15) and (16) have distinct, non-equivalent, logical forms). Causal laws, on the other hand, are general assertions of kinds of causal relations, for example,

L. Scratching dry, oxygenated, well-made matches causes them to light.

One expects the following logical relationship to hold between these kinds of statement: that from causal laws and sets of descriptive sentences — sentences describing particular circumstances — one can infer simple causal statements.\(^8\)

Our original insight can now be cast in inferential terms. Given L, it seems plausible that from a description of an event of type (i), that is, A'S\(^i\), we can infer not only "m lights", but also the simple causal statement, "The fact that match m is scratched * m is dry * m is oxygenated * m is well-made causes (ed)\(^9\) it to be the case that m lights". However, while we can infer "m is not dry" from L and a description of an event of type (ii), that is, A'S\(^ii\), we cannot infer, "The fact that m is scratched * m is unlit * etc. caused it to be the case that m is not dry".

---

\(^8\)Davidson discusses this matter to some extent in \([8]\), pp.90-92.

\(^9\)The tense of the verb "to cause" here is the simple present, whereas in the causal law L the verb is untensed. To prevent confusion, I shall use "caused", which is the verb's simple past, in simple causal statements, retaining the untensed form in causal laws.
This reasoning suggests a condition which, though restricted to causal conditionals, appears to resolve the S-related counterfactuals problem: that is, A·S, besides meeting the criterion's other conditions, must permit not only the derivation of C, but also the simple causal statement, "The fact that A·S caused it to be the case that C". (Presumably, an appropriate condition must be added to the negative side of the criterion as well.) Moreover, the nature of the causal link thus formulated is purely inferential, based on the assumed derivability of simple causal statements from general causal laws and descriptive statements. However, one large problem remains: filling in the details of this kind of inference.

Davidson has argued that causal relations occur between events, which he contends are unrepeatable individuals of some sort ([7], p.25). This conclusion is based on a general feature of the language with which we describe the properties and relations of events: that is, linguistically speaking, events correspond to singular terms, not sentences ([7], p.25). On this view, the role of the expression "caused" in a singular causal statement such as

15. The scratching of match m caused it to light

is seen, not as a non-truth functional sentential connective, as in

16. The fact that match m was scratched caused it to be the case that m lights

but as a two-place predicate, linking two singular terms or definite descriptions, as in

17. (\exists x)(Is \text{scratched}(m,x)) \text{ caused } (\forall y)(\text{Lights}(m,y))
which may be read, somewhat sloppily, as "The x which is a scratching of m caused the y which is a lighting of m" ([6], p.87). However, if this view is correct, how can we infer a causal statement of the form, "The fact that $\phi$ caused it to be the case that $\psi$", where $\phi$ and $\psi$ are sentences, from any group of causal laws and descriptive statements, when simple causal statements such as (15) do not express connections between facts -- which correspond to sentences -- but between events -- which correspond to singular terms?

Nevertheless, we can find a place for statements of the form, "The fact that $\phi$ caused it to be the case that $\psi$". An interesting feature of (17) is that, in order to construct definite descriptions for events, it was necessary to construe "is scratched" and "lights" as two-, not one-place predicates. Someplace must be found for a variable ranging over events. Indeed, Davidson has defended the view that certain predicates, usually associated with action or change, are best regarded as having an additional place for events ([8], p.93). He regards:

18. Caesar died

for example, as an existential statement constructed from a two-place predicate, "died", and a singular term, "Caesar", that is,

19. $(\exists x)(\text{Died(Caesar,x)})$

or, loosely put, "There exists an event x such that Caesar died it" ([9], p.83). This view -- that some verbs contain "one more place than we generally think, a place for events" ([9], p.83) -- allows Davidson to explain a number of the logical features of the language of action,
change, and causation (cf.[6]; [7]; [8]). Let us call such verbs 'event-predicates'.

An upshot of this is that we can find an interpretation of some causal statements of the form, "The fact that $\phi$ caused it to be the case that $\psi$". This is accomplished by interpreting "Match m is scratched" as

$$ (\exists x) (\text{Is scratched}(m,x)); $$

"m lights" as

$$ (\exists y) (\text{Lights}(m,y)); $$

and (16) as the complex existential statement

$$ (\exists x)(\exists y)[(\text{Is scratched}(m,x)) \cdot (\text{Lights}(m,y)) \cdot (\text{Caused}(x,y))] $$

that is, as "There exist events x and y such that x is a scratching of m, y is a lighting of m, and x caused y" (cf.[6], pp.86-87). Here, of course, "caused" retains its construal as a two-place predicate. Therefore, this interpretation of statements of the form, "The fact that $\phi$ caused it to be the case that $\psi$", is restricted to those statements $\phi$ and $\psi$ which contain, in an appropriate fashion, an event-predicate such as "is scratched" or "lights".

However, (22) betrays a feature which raises a difficult problem for our present approach: it entails the existence of both a 'cause-event' and an 'effect-event'. In general, treatment of "caused" as a two-place event-predicate restricts our interpretation of "The fact that $\phi$ caused it to be the case that $\psi$" to $\phi$s and $\psi$s which assert the existence of events. But what about those counterfactuals whose antecedent $\phi$ and consequent $\psi$ are causally linked,
in some fashion, yet whose causal link seems to depend on the possibility of the absence of an event being responsible or accounting for the absence of another? For example, there are imaginable situations in which it would be true that:

23. Match m is not scratched \( \square \rightarrow m \text{ does not light.} \)

Clearly there is some positive sense in which not scratching m could be potentially responsible for m's non-lighting. Yet the causal link underlying (23), unlike that underlying (6), seems dependent on the absence of events, not their presence.

Our precise difficulty is this. According to the suggested condition, if (23) is true, then we ought to be able to infer from "Match m is not scratched", some S, and some causal laws the statement "The fact that match m is not scratched 'S caused it to be the case that m did not light". However, if the latter adequately expresses the causal link underlying (23), then it ought not to entail either the existence of a cause-event (unless S does) or an effect-event, although it should express a positive causal connection of some kind. Therefore, it is difficult to interpret along the lines of (22), whose assertion of causal connection is intimately tied to the existence of two events: a cause and an effect. Hence our problem, in short, is this: to assert causation, in this sense of "cause", one needs events.

Perhaps the absences of events can be treated as events in themselves? This seems unnatural: "Match m is not scratched" seems best expressed by:
24. \(-\exists x\)(Is scratched(m,x))

that is, "No scratching exists", rather than by:

25. \(\exists x\)(Is not-scratching(m,x))

that is, "Some non-scratching exists". Indeed, (24) appears to express a kind of state or condition. Nevertheless, Davidson hints at this approach to absences, and Mackie, in a slightly different context, attempts it ([6], p.93; [15]). However, this idea raises a host of serious ontological problems (cf. Kim [13], pp.51-54).

Therefore, I shall pursue a more promising notion: that we have been investigating the wrong sort of causal connection.

9. The Missing Link (2)

At the end of [6], Davidson makes a pregnant observation about the use of the word "cause". After outlining his view that in many contexts "cause" should be treated as a two-place event-predicate, he says:

This is not to say there are no causal idioms that directly raise the issue of apparently non-truth-functional connectives. On the contrary, a host of statement forms, many of them strikingly similar .... challenge the account [previously presented]. Here are samples: 'The failure of the sprinkling system caused the fire', 'The slowness with which controls were applied caused the rapidity with which the inflation developed', 'The collapse was caused, not by the fact that the bolt gave way, but by the fact that it gave way so suddenly and unexpectedly', 'The fact that the dam did not hold caused the flood'. Some of these sentences may yield to the methods I have prescribed .... but others remain recalcitrant. What we must say in such cases is that in addition to, or in place of, giving what Mill calls the 'producing cause', such sentences tell, or suggest, a causal story. They are, in other words, rudimentary causal explanations. Explanations typically relate statements,
not events. I suggest therefore that the 'caused' of the sample sentences in this paragraph is not the 'caused' of straightforward singular causal statements, but is best expressed by the words 'causally explains'.

([6], p.93)

This suggestion — that the word 'cause', when it means 'causally explains', serves as a sentential connective — has been explored by others, in particular, Mackie ([16], Chap. 10). Can it help us?

Two difficulties previously encountered disappear immediately. First of all, explanations connect statements. Hence we no longer face the logical problem of moving between sentences and singular terms, or the ontological problem of moving between states or facts, and events. Secondly, while on our earlier view it was problematic to assert that the non-existence of one event caused the absence of another, there is nothing difficult about asserting that the absence of the first causally explains the absence of the second. All that is required is an account of how the absence of the first, that is, the fact that m was not scratched, resulted in the absence of the second, that is, the fact that m did not light, given the web of causal relations in nature.

Can this idea assist us with our central concern -- the S-related counterfactuals problem -- as well?

Explanations, says Hempel, are responses to the question, "Why is it the case that p ?" ([11], p.334). However, they are not the only form of response to a why-question; why-questions which ask, "Why is it believed that p ?", which Hempel terms "reason-seeking or epistemic", demand not explanations but "evidence or grounds or reasons in support of the given assertion" ([11], pp.333-335).
Let us imagine, for the moment, a situation much like the 'familiar case', differing only in that match \( m \) has been scratched and is lit. Were we to ask, "Why is it the case that \( m \) lights?", we could reply: "\( m \) lights because it is a dry, oxygenated, well-made match, and such matches light when scratched". This account, though brief, is entirely adequate a response. On the other hand, let us imagine a slightly different situation: our oxygenated and well-made match is scratched but does not light, because it is wet. Were we now to ask, "Why is it the case that \( m \) is not dry?", our reply, if adequate, must indicate circumstances in \( m \)'s past which rendered it damp. It is not sufficient to merely point out that \( m \) is oxygenated, well-made, scratched, and unlit: these facts do not explain why \( m \) is not dry.

At best, they could be adduced in favour of one's belief that \( m \) is not dry, if that belief were the subject of a reason-seeking why-question. One could reply: "Well, \( m \) must be so, since it didn't light when scratched, and since \( m \) is oxygenated etc".

Why do these two sets of facts differ in explanatory power? A hint was given in our second case: what we look for in the causal explanation of a fact \( F \), besides causal laws, are facts which are causally prior to \( F \) and responsible for it. The facts adduced in the first case are such; those in the second are not. Hempel concurs with this observation. He says:

\[
\text{in the context of explanation, a 'cause' must be allowed to be a more or less complex set of circumstances and events, which might be described by a set of statements } C_1, C_2, \ldots, C_k.\]
And, as is suggested by the principle "Same cause, same effect", the assertion that those circumstances jointly caused a given event implies that whenever and wherever circumstances of the kind in question occur, an event of the kind to be explained takes place. Thus causal explanation implicitly claims that there are general laws -- let us say $L_1, L_2, \ldots, L_r$ -- in virtue of which the occurrence of the causal antecedents mentioned in $C_1, C_2, \ldots, C_k$ is a sufficient condition for the occurrence of the phenomenon to be explained.

([11], pp.4348-349)

However, Hempel appears to take too narrow a view of the kind of state which can explain, or be explained, in a causal explanation. As we saw earlier, the absence of an event -- which is presumably a state, not an event -- can also explain, or be explained. Therefore, we must widen our notion of explanatory causal antecedents to cover not only the occurrence of events, but the absence of events as well.

Clearly, these insights have implications for the $S$-related counterfactuals problem. In our discussion of the familiar situation $K$ involving match $m$, we saw that this problem arose because the criterion allows selection of a set of $K$-describing sentences -- call it $'S'$ -- which supports

6. Match $m$ is scratched $\rightarrow$ $m$ lights

and also another -- call it $'S''$ -- which supports the false

7. Match $m$ is scratched $\rightarrow$ $m$ is not dry.

Now we see that by imagining two situations, one described by "Match $m$ is scratched" (A), $S'$, and "$m$ lights" (C), the other by A, $S''$, and "$m$ is not dry", we discovered that $A \cdot S'$ possesses a feature lacked by $A \cdot S''$: in the first situation, $A \cdot S'$, together with our causal law (L), explains the
fact that C, while in the second, A·S'' and L do not explain the fact that m is not dry. But how can we adapt this insight to speak directly to our problem? After all, the situation in which the truth-values of (6) and (7) are to be assessed, that is, K, is unlike either imaginary situation in that A is false in K.

Crudely put, an explanation of empirical fact is constructed in some fashion from a set of descriptive sentences and empirical laws. Hempel distinguishes two types of explanation: 'true' and 'potential'. In the former, all the members of the set of descriptive sentences and laws must be true; in the latter, although other requirements must be met, they need not be so ([12], pp.247-249). When p describes an actual state or event, therefore, one expects a true explanation in response to "Why is it the case that p?"; however, when p describes a possible state or event the best one can expect is a potential explanation based on a description of a possible set of antecedents.

Adopting this terminology allows us to describe the causal link underlying (6) but not (7) as follows: in A, S', and L we have the materials with which to construct a potential causal explanation for the lighting of m, while in A, S'', and L, we do not for m's wetness.

It appears, therefore, that resolution of the S-related counterfactuals problem, at least for causal counterfactuals, requires (i) an adequate analysis of causal explanation, including (ii) a description of those characteristics which make a set of circumstances explanatory causal antecedents. Lacking both (i) and (ii), however, we can but note that some A-C inference-routes presently accepted by the criterion
employ materials (A, S, and causal laws) which are potentially explanatory of the fact asserted by C; that others do not; and that final amendment of the criterion depends on a full analysis of the difference. In lieu of this, let us call paths of the first type 'explanatory' inference-routes.

It is worthwhile, but not necessary to hitch these observations to the Deductive-Nomological (D-N) model of scientific explanation. As Hempel observes, explanation, as normally understood, is a pragmatic concept:

Very broadly speaking, to explain something to a person is to make it plain and intelligible to him, to make him understand it. Thus construed, the word 'explanation' and its cognates are pragmatic terms: their use requires reference to the persons involved in the process of explaining. In a pragmatic context we might say, for example, that a given account A explains fact X to person P. We will then have to bear in mind that the same account may well not constitute an explanation of X for another person P2, who might not even regard X as requiring an explanation, or who might find the account A unintelligible or unilluminating, or irrelevant to what puzzles him about X. Explanation in this pragmatic sense is thus a relative notion: something can be significantly said to constitute an explanation in this sense only for this or that individual.

([11], pp.425-426)

However, Hempel contends that the fact that a notion X is context-relative does not entail that a related, context-free notion X' cannot be abstracted: proof is one such notion ([11], p.426. The D-N model is Hempel's attempt to repeat this feat for scientific explanation.

What is the D-N model of explanation? A D-N explanation, as conceived by Hempel, has an "explanans" (which does the explaining)
consisting of fact-stating sentences, \( C_1 \ldots, C_k \), and empirical laws, 
\( L_1 \ldots, L_k \), as well as an "explanandum", \( E \) (which expresses the phenomenon to be explained). To qualify as an explanation of \( E \), the explanans must meet at least the first three of the following adequacy-conditions (briefly sketched):

- the explanans must (i) entail \( E \);
- (ii) employ empirical laws in the derivation of \( E \);
- (iii) have empirical content;

and (iv) consist of true sentences ([12], pp.247-249).

As mentioned earlier, an explanation meeting (i) through (iv) is a true explanation; and that meeting only (i) through (iii), a potential explanation.

The D-N model, therefore, constructs explanations from materials abundantly supplied by the Goodman criterion. The truth of a causal counterfactual, on Goodman's view, depends on the derivability of \( C \) from \( A \) (which presumably describes some empirically possible fact, if the counterfactual is causal in nature), \( S \), and some causal laws. Indeed, if one regards \( A, S \), and the causal laws as a tentative explanans, and \( C \) as explanandum, the features of the Goodman criterion satisfy adequacy-conditions (i) through (iii). Thus, it is appealing to view the criterion, from the point of view of the D-N model, as attempting to determine, at least for causal counterfactuals, the features of counterfactual-supporting potential explanations; and, specifically, to regard the criterion as demanding no more than that a causal counterfactual be judged true if and only if \( A, S \), and some causal laws appropriately
explain C; and no such explanation exists for -C.

On this view, given our earlier discussion, the defect in the criterion betrayed by the S-related counterfactuals problem stands out starkly: it does not adequately characterize the explanans of a counterfactual-supporting explanation because, as we saw in the match example, it does not invariably pick out sets of causally explanatory antecedents for the phenomenon described by C. What features characterize these sets? I do not know, although I do know that the answer must lie with analysis of the notion of cause and causal explanation.

Hitherto I have focussed discussion on causal counterfactuals. Can our results be extended to other kinds as well? It will be remembered, in the application-form case (Section 8), that:

13. I write "Black" in the second blank
I have written "John Black"

was illegitimately established in a situation K in which:

14. I write "Black" in the second blank
I have not written "John" in the first blank

is true. K was such that "I have written 'John' in the first blank" (S) is true, while "I write 'Black' etc." (A) and "I have written 'John Black'" (C) are false.

If I ask of K, "Why would it be the case that I have written 'John Black' if I were to write 'Black'?", I direct my listener, in effect, to imagine a possible situation described by A, S, and C, and request him to account for the happenings therein. The reply here is straightforward: "Well, if both 'John' and 'Black' are inscribed in the appropriate blanks,
the result is 'John Black'". This reply, based on A and S, is a potential explanation of C: thus, A and S explain C.

On the other hand, asking "Why would it be the case that I have not written 'John' if I were to write 'Black'?", directs the listener to a situation described by A, -S, and -C. Here the reply is not simple: it is inadequate to point out that despite the fact that "Black" has been inscribed, "John Black" is not, hence "John" has not been written. These facts, that is, A and -C, do not explain why "John" was not written in the possible situation, though they may support our belief that this must be so.

These brief observations, which parallel those made earlier about (6) and (7), suggest a surprising conclusion. Our original intuition about the S-related counterfactuals problem was that its solution for causal counterfactuals such as (6) and (7) lay with the notion of causation. Further examination, however, indicated that the key idea was causal explanation. Now our evidence suggests that the general solution to the S-related counterfactuals problem lies with explanation alone; the notion of causation drops out of the picture, except for causal counterfactuals. Thus, if this is true, two related questions remain: (i) can a general notion of counterfactual-supporting explanation be made out in a form amenable to a Goodman-style criterion?; and (ii) if so, what general characteristics do the sets of descriptive conditions adduced in explanations of this sort share with causal antecedents?
10. **An Approach to the Undercutting Phenomenon Problem**

We noted, in a situation described in Section 7, that both:

11. The key is turned \( \rightarrow \) Martin's Alfa starts

and

12. The key is turned \( \rightarrow \) the Alfa's starter malfunctions

are not true. Specifically, thanks to the antics of a bomb-happy friend, we observed what was called the "undercutting phenomenon": that is, (11) is false because (12), an S-related counterfactual, is true. This phenomenon threatened reinstatement of the original cotenability condition, a difficulty we dubbed the "undercutting phenomenon" problem.

How are we to deal with this issue without inviting vicious circularity, the Horseman of Cotenability? The only way, I believe, is to take the undercutting phenomenon seriously, and turn it to our advantage. Let us begin by noting that the undercutting phenomenon appears to occur only in situations \( K \) in which the \( S^1 \) in the true S-related \( \phi \rightarrow -S^1 \) is a member of the particular \( S \) with which we wish to establish. \( \phi \rightarrow \psi \). Suppose, for example, that in addition to the car's starter Martin's Alfa has a special mechanism - call it "X" - that will also turn over and start the car when the key is turned. Imagine, furthermore, that \( K \) is such that the key is not turned, and that both the car's starter and X are in working order. In this \( K \) we would affirm (11) and deny (12). Of course, in any situation in which the key is actually turned there may be some dispute about which mechanism was causally responsible for the starting of the car. Depending on which mechanism is so regarded, however, it is obvious that (11) may be
supported by either of two distinct S-ideal sets in K.

Now let us modify K. Suppose that Martin's colleague places a bomb on the starter, but, because he is ignorant of the car's special feature, does nothing which interferes with X. That is, let us assume that the bomb is very small, and incapable of damaging anything other than the starter. In this situation we would now affirm: (11) even though we also affirm (12). Why? Roughly put, because there is an inference route, independent of the existence of the starter, but dependent upon that of X, which is capable of supporting (11). That is, it appears that (12) does not undercut (11) in K, even though the S-related (12) is true, because the consequent of (12) is not the denial of any sentence in the particular ideal S with which we support (11). Thus, in our description of the undercutting phenomenon as the fact that, in many K, if an S-related counterfactual of $\phi \rightarrow \psi$ is true, then its truth undercut that of $\phi \rightarrow \psi$, it must be understood that the expression "S-related counterfactual" refers only to those constructed from the particular K-describing S with which we intend to establish $\phi \rightarrow \psi$ in K.

Given this proviso, let us look closely at (11) and (12). The truth of either, in any situation, is partially a function of certain causal laws governing the motion of electricity, the combustion of hydrocarbons, etc. -- in short, both are causal counterfactuals. Moreover, these causal laws dictate a number of causal relationships between various states-of-affairs.

Imagine, for a moment, that X does not exist and that Martin's friend has indeed placed a bomb on the car's starter. The key is turned
and boom! the starter disintegrates. The turning of the key, along with some facts and causal laws, explains this result. But there is another result as well, one which is partially explained by the starter's destruction: because this vital component no longer works, and no other means of starting the car by turning the key exists, the car does not start.

What is the moral of this story? We noted earlier that whenever a causal counterfactual such as (12) is true, there is an explanatory inference-part from "The key is turned" (A) to "The Alfa's starter malfunctions". What our story suggests is that when (12) is true in a situation K there may also be an explanatory inference-route from A to "Martin's Alfa does not start" (-C). Given Goodman's criterion (with the modifications suggested in Section 9), if this is so, then (11) cannot be true in K when (12) is true, since (11) can be true, only if (i) there is an appropriate route from A to "Martin's Alfa starts" (C), and (ii) no such route from A to the negation of this sentence. Therefore, this idea -- that when the S-related (12) is true there is an appropriate path from A to the denial of (11)'s consequent -- if true, would explain the undercutting phenomenon in K.

Can we demonstrate this to be the case? Yes, but the demonstration will require some groundwork. First of all, demonstrating the existence of an explanatory inferential path from A to -C in K when (12) is true involves two distinct considerations: showing that (a) there is an inferential path simpliciter from A to -C; and that (b) this path is explanatory. For the moment we shall set aside (b) and discuss (a), in
the course of which we shall establish a set of conditions on the K in which (a) holds.

Our first condition will be put forward in the following fashion and then modified progressively. Let us pretend, for the moment, that (i) a state described by "The key is turned • the starter is in working order" (A'S) constitutes the only set of possible causal antecedents for the fact that the Alfa starts, ignoring momentarily both the complexities of automobiles, and the possibility of starting the automobile in other ways.

By condition (i) I mean the following: we are to suppose that, given the causal laws of the world, there is only one possible state — described by A'S — that can causally account for the fact described by C; or, in other words, that the causal make-up of the world is such that the truth of A'S is a necessary, as well as sufficient condition for the truth of C.\(^{10}\) The following material equivalence expresses this fact:

\[
F \land (A'S) \equiv C. 
\]

Moreover, demonstrating (a) requires a second important condition on K: that (ii) the use of F in constructing a set of inferences from A to ¬C does not violate any of the criterion's requirements (ignoring the question of explanatory power as well). It should be noted, however, that we shall be modifying condition (i) by stages, each of which will

\(^{10}\)I am not trying here to analyse the notion of causation in terms of conditions; nor do I presuppose such an analysis (cf. Mackie[15]). F is simply intended to express a certain relationship between certain states wrought by the web of causal laws.
demand introduction of a distinct F-like truth. Condition (ii) is intended to cover these as well.

Given that K meets conditions (i) and (ii), the demonstration of (a) when (12) is true is as follows:

I. Assume that (12) is true. Therefore, ex hypothesi, there exists an explanatory inference path from A to "The Alfa's starter malfunctions" (-S).

II. By a simple logical law, we can infer (-S v -A), or -(A·S), from -S.

III. From -(S·A) and F we can infer -C. Q.E.D.

How important and credible are conditions (i) and (ii) ? Condition (ii) is both crucial and plausible. Without F (and as we shall later see, the F-like truths as well), derivation of -C is impossible. However, derivation of C does not seem to depend on any violation of physical or logical law.

On the other hand, condition (i) seems essential, but is not credible. It plays a crucial role in establishing F, yet ignores the complexities of automobile mechanics. In reality, first of all, the set of causal antecedents required to start an Alfa by turning the key is much larger -- there must be sufficient gasoline, the fuel pump must run etc.; and secondly, there are distinct alternative methods of starting an Alfa -- 'push-starting' and 'hot-wiring' -- that do not depend on a key and/or starter -- otherwise, car thievery would be in a sorry state. Let us assume in the first case, that all the other circumstances required to start the Alfa, aside those described by A·S, are expressed by R; and, in the second case, that the alternative ways of starting the car are expressed by sentences X,Y,Z,... which are
disjoined to form a single sentence, T.

The first complication, I believe, is not serious. (A'S) may not be sufficient for C, but (A'S'R) is. If (A'S'R) describes the sole set of causal antecedents for the car's operation — ignoring the second complication — we can construct an F-like sentence:

$$F^1. \quad (A'S'R) \equiv C$$

and, with only slight modification, perform the inference required by step II. It is as simple to infer (¬S v ¬α(A'R)), or (¬A'S'R), from ¬S, as it is to infer (¬A'S). Step III is then easily performed, employing (¬A'S'R) and F^1.

The second complication, however, raises greater difficulties. In this case, (A'S'R) is no longer necessary for C; however, the disjunction of (A'S'R) and T, i.e. ((A'S'R) v T), is. That is, it is extremely plausible that if the car starts, it is because the Alfa's key was turned etc. or it is 'push-started' etc. or it is 'hot-wired' etc. for an indefinite (albeit not infinite) number of alternatives. Once again we can construct an F-like sentence:

$$F^2. \quad ((A'S'R) v T) \equiv C$$

F^2, however, has different logical features from its predecessors.

In order to demonstrate (a) employing F^2, we must rely on a third condition: that (iii) ¬T is true in K, and no criterion condition excludes employing ¬T as a K-describing truth in the inference route from A to ¬C. Given this condition, the requisite modification in the inference route from A to ¬C is made in the following way. Step I remains the same. Then, as in the case of F^1, we infer (¬A'S'R) from
-S. But now comes an important change. Since -T is true of K, we infer -C from F² via -(A'S'R) and -T. On the other hand, without -T we cannot perform this last step.

Thus far I have contended that given certain conditions on K there will be an inference route simpliciter from A to -C when (12) is true. But what of (b)? Is this route causally explanatory? Here we encounter a serious difficulty: we do not have a working account of causal explanation, hence cannot rigorously distinguish routes which are explanatory from those which are not. Thus our guide in this matter must be intuition.

What does intuition say? In Section 9 we determined explanatory power by considering possible situations described by the sentences involved in the inference routes. In our present case, the possible situation corresponding to the inference route in question is that given at the beginning of this section: the tale of the bomb-crazy colleague. It will be remembered that this story suggested that the fact that the Alfa's starter malfunction because the key is turned, together with the absence of other means of starting the car by turning the key, explains the fact that the car does not start.

In summary, therefore, our examination of the implications of that story has proved rewarding, for we have discovered that, given a number of conditions on K, an explanatory inference route from A to -C appears to exist when (12) is true, hence (11) is false. The conditions discovered are these: first, that some F-like truth exist; secondly, that use of this truth be criterion-acceptable in the route from A to -C; and thirdly,
that if the F-like truth is $F^2$, then $-T$ must be true and criterion-acceptable as well. It should be noted, however, that this list is tentative. Without extended and arduous discussion, it is impossible to say with certainty that these conditions are adequate as they stand, or that they are necessary as well as sufficient for the occurrence of the undercutting phenomenon in a given situation $K$. Nevertheless, there is reason for optimism here. If further examination bears out the sort of analysis of the undercutting phenomenon advanced here, an upshot is that no modification of the criterion, beyond that discussed in Section 9, is required to resolve the undercutting phenomenon problem.

Why? If our analysis is correct, then for any $\phi \rightarrow \psi$, $S$-related counterpart $\phi \rightarrow {}^1_S$, and situation $K$ meeting the conditions listed above, if $\phi \rightarrow {}^1_S$ is true in $K$, then there exists an explanatory inference route from $\phi$ to $-\psi$. Hence, given the criterion as it stands, if $\phi \rightarrow {}^1_S$ is true in $K$, then $\phi \rightarrow \psi$ must be false. Moreover, the contrapositive of our analysis' result is this: if there is no such route from $\phi$ to $-\psi$ in $K$, then $\phi \rightarrow {}^1_S$ is not true, that is, is false. Hence, given the criterion as it stands, if $\phi \rightarrow \psi$ is judged true, then $\phi \rightarrow {}^1_S$ must be false. In neither case, therefore, must the criterion concern itself directly with the truth-value of $\phi \rightarrow {}^1_S$ when determining the truth-value of $\phi \rightarrow \psi$ in $K$. Specifically, use of the cotenability condition is unnecessary.
II. Summary

I have attempted to demonstrate four theses: first, that the problem of cotenability is really two problems, not one; secondly, that the cotenability condition is required as a solution to neither; thirdly, that the first, the S-related counterfactuals problem, will be resolved, at least for causal counterfactuals, by an adequate account of causal explanation, and that a similar approach may be successful for all counterfactuals; and finally, that the second, the undercutting phenomenon problem, may require no further modification of Goodman's criterion. However, work remains on both these problems.
Bibliography


