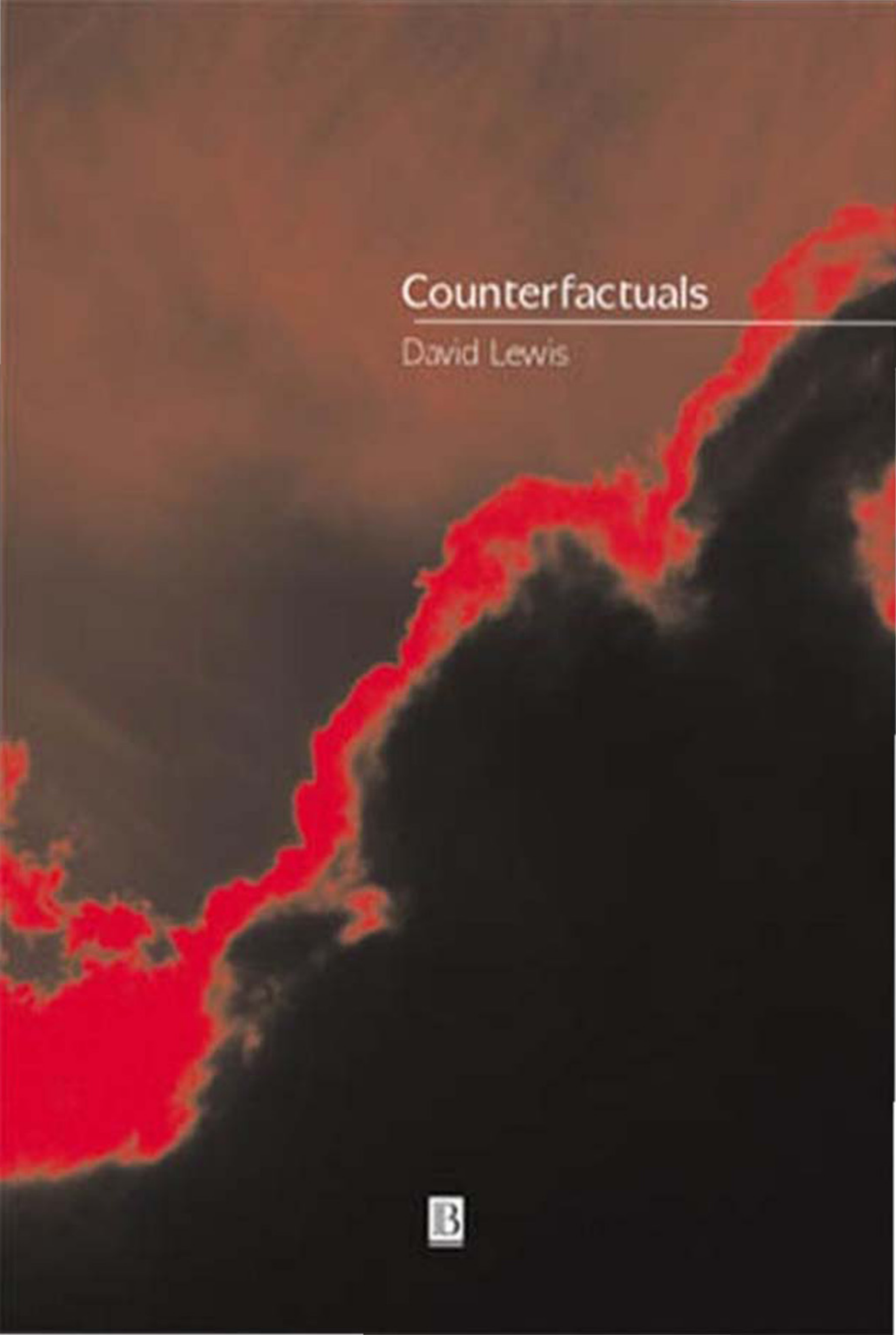
Counterfactuals

David Lewis



Counterfactuals

Counterfactuals

DAVID LEWIS

Blackwell
Publishing

BLACKWELL PUBLISHING
350 Main Street, Malden, MA 02148-5020, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

First published by Basil Blackwell Ltd 1973
First publishing in the USA by Harvard University Press 1973
Reissued by Blackwell Publishers 2001
Reprinted 2003, 2005

Library of Congress Cataloging-in-Publication Data

Lewis, David K., 1941-
Counterfactuals / David Lewis.
p. cm.
First published: 1973.
Includes bibliographical references and index.
ISBN 0–631–22495–5 (hardback.: alk. paper) — ISBN 0–631–22425–4 (pbk. : alk. paper)
1. Counterfactuals (Logic) I. Title.

BC199.C66 L48 2000
160—dc21

00-059899

A catalogue record for this title is available from the British Library.

Printed and bound in the United Kingdom
by MPG Books Ltd, Bodmin, Cornwall

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on
Blackwell Publishing, visit our website:
www.blackwellpublishing.com

IN MEMORY OF RICHARD MONTAGUE

Contents

Preface

The principal changes in this revised printing are in Section 6.1, where I have corrected two major errors in my discussion of completeness results for the *V*-logics. Both of them were spotted by Erik C. W. Krabbe in 1976. I am most grateful to him for finding the trouble, and also for very helpful correspondence about alternative methods of repair. One error was in my construction of the canonical basis on pages 127–130: I falsely claimed that the set of co-spheres of cuts around a given index would be closed under unions.* In order to ensure such closure, it is necessary to construct the canonical basis differently. The other was in the axiom system for **VC** given on page 132. I left out the rule of Interchange of Logical Equivalents; however I tacitly appealed to this rule in proving completeness, so my proof did not apply to the axiom system I had given.

In addition I have corrected minor errors on pages 35, 55 and 129, also spotted by Krabbe; removed misprints; and brought some references up to date.

I have had more to say about counterfactuals and related matters. These further thoughts might appropriately have been added to this book; but since they are to be found elsewhere, I have been content to add an appendix giving citations and abstracts.

David Lewis
1986

* Erik C. W. Krabbe, 'Note on a Completeness Theorem in the Theory of Counterfactuals', *Journal of Philosophical Logic* 7 (1978): 91–93.

Acknowledgements

I am grateful to Kit Fine, Hans Kamp, David Kaplan, Richard Montague, J. Howard Sobel, Robert Stalnaker, Richmond Thomason, and many other friends and colleagues for encouragement and for valuable discussions about counterfactuals over the last five years.

I am grateful also to the American Council of Learned Societies for financial assistance, and to Saint Catherine's College, Oxford, for hospitality, during the year when most of this book was written.

David Lewis
Princeton, June 1972

1. An Analysis of Counterfactuals

1.1 Introduction

'*If kangaroos had no tails, they would topple over*' seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over. I shall give a general analysis of counterfactual conditionals along these lines.

My methods are those of much recent work in possible-world semantics for intensional logic.* I shall introduce a pair of counterfactual conditional operators intended to correspond to the various counterfactual conditional constructions of ordinary language; and I shall interpret these operators by saying how the truth value at a given possible world of a counterfactual conditional is to depend on the truth values at various possible worlds of its antecedent and consequent.

Counterfactuals are notoriously vague. That does not mean that we cannot give a clear account of their truth conditions. It does mean that such an account must either be stated in vague terms—which does *not* mean ill-understood terms—or be made relative to some parameter that is fixed only within rough limits on any given occasion of language use. It is to be hoped that this imperfectly fixed parameter is a familiar one that we would be stuck with whether or not we used it in the analysis of counterfactuals; and so it will be. It will be a relation of comparative similarity.

Let us employ a language containing these two counterfactual conditional operators:

$$\Box\!\!\rightarrow$$

* See, for instance, Saul Kripke, 'Semantical Considerations on Modal Logic', *Acta Philosophica Fennica* **16** (1963): 83–94; Richard Montague, 'Pragmatics', in R. Klibansky, *Contemporary Philosophy* (La Nuova Italie Editrice: Firenze, 1968): 102–122, reprinted in Montague, *Formal Philosophy* (Yale University Press: New Haven, 1974); Dana Scott, 'Advice on Modal Logic', in K. Lambert, *Philosophical Problems in Logic* (D. Reidel: Dordrecht, 1970); and David Lewis, 'General Semantics', *Synthese* **22** (1970): 18–67.

read as '*If it were the case that* ___, *then it would be the case that* . . .', and

$$\Diamond\!\!\rightarrow$$

read as '*If it were the case that* ___, *then it might be the case that* . . .'. For instance, the two sentences below would be symbolized as shown.

If Otto behaved himself, he would be ignored.
Otto behaves himself $\Box\!\!\rightarrow$ *Otto is ignored*

If Otto were ignored, he might behave himself.
Otto is ignored $\Diamond\!\!\rightarrow$ *Otto behaves himself*

There is to be no prohibition against embedding counterfactual conditionals within other counterfactual conditionals. A sentence of such a form as this.

$$((\psi \Box\!\!\rightarrow ((\chi \Box\!\!\rightarrow \psi) \Diamond\!\!\rightarrow \phi)) \Diamond\!\!\rightarrow \chi)$$
$$\Box\!\!\rightarrow (\phi \Box\!\!\rightarrow (\psi \Diamond\!\!\rightarrow ((\chi \Box\!\!\rightarrow \phi) \Diamond\!\!\rightarrow (\phi \Box\!\!\rightarrow \psi))))$$

will be perfectly well formed and will be assigned truth conditions, although doubtless it would be such a confusing sentence that we never would have occasion to utter it.

The two counterfactual operators are to be interdefinable as follows.

$$\phi \Diamond\!\!\rightarrow \psi =^{\text{df}} \sim(\phi \Box\!\!\rightarrow \sim\psi),$$
$$\phi \Box\!\!\rightarrow \psi =^{\text{df}} \sim(\phi \Diamond\!\!\rightarrow \sim\psi).$$

Thus we can take either one as primitive. Its interpretation determines the interpretation of the other. I shall take the 'would' counterfactual $\Box\!\!\rightarrow$ as primitive.

Other operators can be introduced into our language by definition in terms of the counterfactual operators, and it will prove useful to do so. Certain modal operators will be thus introduced in Sections 1.5 and 1.7; modified versions of the counterfactual in Section 1.6; and 'comparative possibility' operators in Section 2.5.

My official English readings of my counterfactual operators must be taken with a good deal of caution. First, I do not intend that they should interfere, as the counterfactual constructions of English sometimes do, with the tenses of the antecedent and consequent. My official reading of the sentence

We were finished packing Monday night $\Box\!\!\rightarrow$ *we departed Tuesday morning* comes out as a sentence obscure in meaning and of doubtful grammaticality:

If it were the case that we were finished packing Monday night, then it would be the case that we departed Tuesday morning.

In the correct reading, the subjunctive 'were' of the counterfactual construction and the temporal 'were' of the antecedent are transformationally combined into a past subjunctive:

If we had been finished packing Monday night, then we would have departed Tuesday morning.

Second, the 'If it were the case that ___' of my official reading of $\Box\!\!\rightarrow$ is not meant to imply that it is not the case that ___. Counterfactuals with true antecedents—counterfactuals that are not counterfactual—are not automatically false, nor do they lack truth value. This stipulation does not seem to me at all artificial. Granted, the counterfactual constructions of English do carry some sort of presupposition that 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. Some or all sorts of presupposition, and in particular the presupposition that the antecedent of a counterfactual is false, may be mere matters of conversational implicature, without any effect on truth conditions. Though it is difficult to find out the truth conditions of counterfactuals with true antecedents, since they would be asserted only by mistake, we will see later (in Section 1.7) how this may be done.

You may justly complain, therefore, that my title 'Counterfactuals' is too narrow for my subject. I agree, but I know no better. I cannot claim to be giving a theory of conditionals in general. As Ernest Adams has observed,* the first conditional below is probably true, but the second may very well be false. (Change the example if you are not a Warrenite.)

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

Therefore there really are 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.

* 'Subjunctive and Indicative Conditionals', *Foundations of Language* 6 (1970): 89–94.

The title 'Subjunctive Conditionals' would not have delineated my subject properly. For one thing, there are shortened counterfactual conditionals like '*No Hitler, no A-bomb*' that have no subjunctives except in their—still all-too-hypothetical—deep structure. More important, there are subjunctive conditionals pertaining to the future, like '*If our ground troops entered Laos next year, there would be trouble*' that appear to have the truth conditions of indicative conditionals, rather than of the counterfactual conditionals I shall be considering.*

1.2 Strict Conditionals

We shall see that the counterfactual cannot be any strict conditional. Since it turns out to be something not too different, however, let us set the stage by reviewing the interpretation of strict conditionals in the usual possible-world semantics for modality. Generally speaking, a strict conditional is a material conditional preceded by some sort of necessity operator:

$$\Box(\phi \supset \psi).$$

With every necessity operator $\Box$ there is paired its dual possibility operator $\Diamond$. The two are interdefinable:

$$\Diamond\phi =^{\mathrm{df}} {\sim}\Box \sim \phi, \quad \text{or} \quad \Box\phi =^{\mathrm{df}} {\sim}\Diamond \sim \phi.$$

If we like, we can rewrite the strict conditional using the possibility operator:

$${\sim}\Diamond(\phi \;\&\; {\sim}\psi).$$

Or we could introduce a primitive strict conditional arrow or hook, and define the necessity and possibility operators from that.‡

A necessity operator, in general, is an operator that acts like a restricted universal quantifier over possible worlds. Necessity of a certain sort is truth at all possible worlds that satisfy a certain restriction. We

* Notation: sentences of our language are mentioned by means of lower-case Greek letters ϕ, ψ, χ *et al.*; sets of sentences by means of Greek capitals. Logical symbols and the like are used autonymously, and juxtaposition of names of expressions signifies concatenation of the expressions named. Possible worlds are mentioned by means of the lower-case letters h, i, j, k; sets of worlds by means of capital letters; and sets of sets of worlds by means of script capitals.

‡ In this section only, I use the unmarked box and diamond to stand for *any* arbitrary paired necessity operator and possibility operator. When next they appear, in Section 1.5, they will be reserved thenceforth for a specific use: they will be the 'outer' necessity and possibility operators definable in a certain way from the counterfactual (or they will be analogously related to operators analogous to the counterfactual). The dotted box and diamond, $\boxdot$ and $\diamonddot$, will be likewise reserved when they appear in Section 1.7.

call these worlds *accessible*, meaning thereby simply that they satisfy the restriction associated with the sort of necessity under consideration. Necessity is truth at all accessible worlds, and different sorts of necessity correspond to different accessibility restrictions. A *possibility operator*, likewise, is an operator that acts like a restricted existential quantifier over worlds. Possibility is truth at some accessible world, and the accessibility restriction imposed depends on the sort of possibility under consideration. If a necessity operator and a possibility operator correspond to the same accessibility restriction on the worlds quantified over, then they will be a dual, interdefinable pair.

In the case of *physical necessity*, for instance, we have this restriction: the accessible worlds are those where the actual laws of nature hold true. Physical necessity is truth at all worlds where those laws hold true; physical possibility is truth at some worlds where those laws hold true.

In the case of physical necessity, which possible worlds are admitted as accessible depends on what the actual laws of nature happen to be. The restriction will be different from the standpoint of worlds with different laws of nature. Let i and j be worlds with different laws of nature, and let k be a world where the laws of i hold true but the different laws of j are violated. From the standpoint of i, k is an accessible world; from the standpoint of j it is not. Accessibility is in this case—and most cases—a relative matter. It is the custom, therefore, to think of accessibility as a relation between worlds: we say that k is *accessible from* i, but k is not accessible from j. We say also that i stands to k, but j does not stand to k, in the *accessibility relation* for physical necessity and possibility.

In general: to a necessity operator $\Box$ or a possibility operator $\Diamond$ there corresponds an accessibility relation. The appropriate accessibility relation serves to restrict quantification over worlds in giving the truth conditions for $\Box$ or $\Diamond$. For any possible world i and sentence ϕ, the sentence $\Box\phi$ is true at the world i if and only if, for every world j such that j is accessible from i, ϕ is true at j. Likewise $\Diamond\phi$ is true at i if and only if, for some world j such that j is accessible from i, ϕ is true at j. More concisely: $\Box\phi$ is true at i if and only if ϕ is true at every world accessible from i; $\Diamond\phi$ is true at i if and only if ϕ is true at some world accessible from i. It follows that the strict conditional $\Box(\phi \supset \psi)$ is true at i if and only if, for every world j such that j is accessible from i, the material conditional $\phi \supset \psi$ is true at j; that is, if and only if, for every world j such that j is accessible from i and ϕ is true at j, ψ is true at j. More concisely: $\Box(\phi \supset \psi)$ is true at i if and only if ψ is true at every accessible ϕ-world. ('ϕ-world', of course, abbreviates 'world at which ϕ is true', and likewise for parallel formations.)