Causal inquiry in international relations

OXFORD

CAUSAL INQUIRY IN INTERNATIONAL RELATIONS

Adam R. C. Humphreys | Hidemi Suganami

Causal Inquiry in International Relations

Causal Inquiry in International Relations

Adam R. C. Humphreys

&

Hidemi Suganami

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2024941523

ISBN 9780198823582

DOI: 10.1093/oso/9780198823582.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Acknowledgements

We are very grateful to the University of Reading for providing the funds to publish this book open access, so that it can reach a wider audience.

Contents

List of a Figure and Tables

Introduction

Here is a widely known fact about causation: an object dropped near the Earth's surface will accelerate downwards at approx. 9.8m/s^2; this is caused by gravity.[1]

Here is another widely known fact about causation: not all objects dropped near the Earth's surface accelerate downwards at approx. 9.8m/s^2; indeed, very few of them do!

Take a relatively light, un-aerodynamic object, such as a feather: if dropped, it is likely to be blown around in the breeze before eventually settling some distance from the point directly below where it was dropped, having fallen at no great speed and with no great acceleration. Compare the time this feather takes to reach the ground with the time it would take for a relatively heavy and aerodynamic object, such as a bowling ball, to reach the ground; consider, further, what this indicates about the different rates at which they will have accelerated downwards towards Earth.

Putting these two widely known facts together illustrates something important about gravity: we know what effect it will have on an object dropped near the Earth's surface *in the absence of interference*—it will cause this object to accelerate downwards at approx. 9.8m/s^2—but we are rarely in a position to observe its effect in the absence of interference; in practice, the effect of gravity on an object dropped near the Earth's surface will nearly always be countered, to some extent, by air resistance and by the wind. Moreover, the relative influence of the various factors affecting the object's fall will differ for each object, depending on how massive and how aerodynamic it is, which is why, if a feather and a bowling ball are each dropped from a similar position near the Earth's surface, the latter will fall much more quickly and in a straighter line. The two objects will fall together, each accelerating downwards at approx. 9.8m/s^2, only under very special conditions—that is, when interference is removed.[2]

[1] This rate of acceleration may be derived from Newton's law of universal gravitation and his second law of motion, given the mass and radius of the Earth.

[2] This is beautifully illustrated in the BBC documentary *Human Universe*, Series 1, Episode 4 (2014): see https://www.youtube.com/watch?v=E43-CfukEgs (accessed 18 Aug. 2023).

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0001

More broadly, these two facts illustrate something profoundly important not just about gravity, but also about causation: claims about causation relate to the unfolding of what we will, in this book, term 'propensities'; these propensities unfold *fully* only when the conditions are right for them to do so. We will come to discuss this idea in more detail, but for now let us simply spell out the point in relation to gravity:

- The statement 'gravity causes objects dropped near the Earth's surface to accelerate downwards at approx. 9.8m/s^2' is a propensity statement: it describes the effect which gravity will produce under the right conditions—that is, in the absence of interference;
- When interference is present, this propensity will not unfold fully: due, for example, to the effects of air resistance, objects dropped near the Earth's surface will typically accelerate at less than 9.8m/s^2.[3]

Most of us will have little problem reconciling our 'knowledge' that gravity causes objects dropped near the Earth's surface to accelerate downwards at approx. 9.8m/s^2 with the fact that most objects dropped near the Earth's surface do not actually behave like this. This is because we understand, at least implicitly, that the claim 'gravity causes objects dropped near the Earth's surface to accelerate downwards at approx. 9.8m/s^2' is a propensity statement, not an empirical generalization: it does not *generalize* about what in fact happens to actual objects dropped near the Earth's surface, but rather expresses our theoretical understanding of the effect that gravity would have on such an object in the absence of interference.[4]

The significance of all this for those studying world politics is that it is often *misunderstood* in the discipline of International Relations (IR).

In IR, and especially in the research methods literature, it is common to see causal statements characterized as if they were generalizations, or as if they described regularities that should be observable in the world around us. Consider, for example, Stephen Van Evera's contention, in his widely used guide to research methods, that statements of the form '*A* causes *B*' imply observable patterns of the form 'whenever *A*, then *B*' or 'whenever *A*, then *B*, with probability *x*' (1997: 8; emphasis added).[5] Our discussion so far already provides

[3] Moreover, due to this interference, most falling objects will quickly reach their 'terminal velocity' and stop accelerating altogether.

[4] The idea that scientific laws and theories express understandings of how things work under special conditions which are not routinely observed in the world around us is a central feature of Nancy Cartwright's early work (1983, 1989, 1999).

[5] Our point is not to criticize Van Evera, in particular, for presenting a misleading analysis; this is merely one highly visible example of a widespread problem.

serious grounds for concern about this claim, for the statement 'gravity causes objects dropped near the Earth's surface to accelerate downwards at approx. 9.8m/s^2' is a statement of the form '*A* causes *B*', but it does *not* seem to imply that whenever gravity is present, it will cause an object dropped near the Earth's surface to accelerate downwards at approx. 9.8m/s^2, or even that gravity will cause this to happen with any particular probability. As we will show in Chapter 4, what a statement of the form '*A* causes *B*' implies is that an *A*-type event will produce a *B*-type event *if the conditions are right*: the statement '*A* causes *B*' is, in other words, a propensity statement. The details can wait. For now, the key point is just that existing discussions of causation in IR can be, and often are, profoundly misleading in this regard.

It is important to appreciate, moreover, that the problem is *not* just that some research methods texts offer misleading over-simplifications: the problem goes much deeper than this. An implicit assumption that causal statements express generalizations, or describe regularities which should be observable in the world around us, shapes substantive causal inquiries in quite profound ways.

Consider, for example, one of the most well-known research programmes in IR: the investigation into the causes of the so-called 'democratic peace'—the fact that liberal democracies have never gone to war with one another (at least on certain definitions of what constitutes a 'liberal democracy' and of what constitutes a 'war').[6] In research on the democratic peace, a key analytical tool has been the study of potential exceptions: cases in which it is claimed that democracies did in fact go to war with one another (see, for example, Ray 1993) or that the behaviour of democratic states was inconsistent either (i) with the core idea that 'shared democracy causes peace' (see, for example, Layne 1994; Rosato 2003) or (ii) with certain favoured accounts of *how* shared democracy causes peace (see, for example, Owen 1994).[7] Curiously, however, a key question is rarely discussed: why should such exceptions, if they exist, be considered significant?

The focus on possible exceptions to the 'democratic peace' appears to be motivated by one of two closely linked assumptions. The first is that *if* shared democracy does cause peace, then it should do so without exceptions: this mirrors Van Evera's contention that if '*A* causes *B*', then a pattern of the form 'whenever *A*, then *B*' should be observable. If this is right, then discovering a proven exception would 'falsify' a claim that shared democracy causes

[6] For a recent survey of the state of the art in relation to this research programme, see Reiter (2017).

[7] The central claim of democratic peace theory is that when two states are both democracies, they will not go to war with each other and that their peaceful relations are caused by their shared democracy; we summarize this in the slogan 'shared democracy causes peace'.

peace. The second assumption is that *if* shared democracy causes peace, then there should be a single causal mechanism which operates in every case, thus enabling us to generalize about *how* shared democracy causes peace. If so, then discovering a single case in which a putative mechanism did *not* operate would decisively eliminate this mechanism from the list of possible mechanisms through which shared democracy produces peace.

To get a sense of why these assumptions might be problematic, consider, by way of comparison, the claim 'exercise causes improved mental health'. As with the gravity example we have already considered, most people will have no problem in accepting this statement while also acknowledging that not every instance of exercising will produce improved mental health. In short, most people will intuitively understand that exceptions—cases in which exercise does *not* generate improved mental health—do not undermine the core claim 'exercise causes improved mental health'. Why, then, should cases in which two democracies went to war with one another, or in which a democracy failed to behave in accordance with expectations derived from the theory 'shared democracy causes peace', undermine this theory?[8] It is also common knowledge that exercise can influence mental health in different ways.[9] It would therefore seem odd to demand that there should be a single causal mechanism which operates in every case in which exercise contributes causally to improved mental health. But, if so, is it not also odd to demand that there should be a single causal mechanism which operates in every case in which, it is claimed, shared democracy contributes causally to peace between democracies?

Causal knowledge is potentially very powerful, enabling us to control and manipulate the world around us. However, misleading analyses of causal statements (such as that offered by Van Evera) and problematic assumptions about issues such as the causal significance of exceptions to a pattern (as found in some research on the 'democratic peace') are widespread in IR. They are, moreover, closely linked to what we will, in this book, term 'the culture of generalization': the tendency to prioritize the production of 'general' knowledge over knowledge of 'particular' (also: 'specific', 'unique', 'individual', 'local') facts and events and to define key terms in ways which reflect this goal.

[8] As we explain in Chapter 5, a theory such as 'exercise causes improved mental health' or 'shared democracy causes peace' becomes well established when one or more cases are discovered in which this causal relation is shown persuasively to have unfolded; for the theory to be good, it is not necessary that exercise *always* improves mental health or that shared democracy *always* generates peace.

[9] For example, it might contribute to better sleep, improved mood, or enhanced social contact.

We accept, of course, that 'general' knowledge can be extremely valuable, but the tendency to construe causal statements as generalizations and to associate causal claims with a picture of the world in which regular patterns are readily observable is quite misleading (cf. Abbott 1988). In order to develop reliable causal claims, it is necessary to be realistic about the complexity of the world around us, to be precise about what causal statements assert and imply, and to recognize the limits of what our causal inquiries reveal. On all these matters, the culture of generalization points in the wrong direction. A central message of this book is that causal relations concern the unfolding of propensities and hence that even when we succeed in developing reliable causal knowledge it remains very difficult to 'generalize' about what should be observable in the world around us.

Aims

This book emerges out of a conviction that causal inquiry, when properly understood, has a crucial role to play in the study of world politics, but that in the discipline of IR important aspects of causal inquiry are widely misunderstood. The book has three broad aims.

First, we aim to demonstrate what is problematic about some established ways of thinking about causation and some deeply entrenched understandings of what causal statements assert and imply; we aim, thereby, to provide a sounder basis on which to design, conduct, and evaluate causal inquiries.

Second, we aim to provide a more adequate account of the basic logic of causal inquiry and what it implies for a range of methodological questions.

Third, we aim to enhance collective understanding of the contribution that causal inquiry can make to empirically rich and critically aware scholarship about world politics and to highlight some of the challenges that are likely to arise, even in high-quality work, when conducting causal inquiries in a discipline such as IR.

In pursuing these aims we will draw on a wide range of existing scholarship. However, one significant obstacle that those seeking to understand these issues must overcome is that existing discussions, in IR, of causation and causal inquiry have developed in two quite different directions.

On the one hand, there is an extensive and sophisticated methodological literature, focused on showing how specific methods for generating and analysing empirical evidence may be applied to generate 'causal inferences'. Moreover, there is an abundance of applied work, much of it high quality, which engages in the detailed business of producing causal knowledge

in accordance with the guidance found in this methodological literature.[10] As we have indicated, however, some of the claims made in the methodological literature about what causal statements assert and imply is quite misleading. Consequently, while the discipline benefits from extensive knowledge and guidance about how to apply particular methods in the service of causal inquiry, collective understanding of how causal findings can be most accurately presented and of the kind of reasoning required to evaluate and improve on them is much less satisfactory. This manifests itself most obviously in overblown claims about how 'generally' the knowledge produced by particular instances of causal inquiry holds.

On the other hand, recent years have seen a significant renewal of interest in philosophically informed engagement with questions about causation (see, for example, Patomäki and Wight 2000; Kurki 2008; see also Jackson 2011). This work has tended to be quite critical of the mainstream methodological literature and of causal inquiries which adhere to the guidance this literature provides, but it has, for the most part, focused on metaphysical questions, especially questions about whether causation is, and should explicitly be acknowledged to be, a real feature of the world existing independently of human thought. Indeed, the most striking claim to have emerged out of this renewed philosophical engagement with questions about causation in IR is that in order to improve our causal inquiries it is necessary to embrace causal and scientific realism (see, for example, Wendt 1999; Bennett 2013).[11]

For the most part, these two literatures have talked past each other. Advocates of causal and scientific realism have made a persuasive case that realism provides a viable and coherent basis on which to conduct causal inquiries, but they have not demonstrated convincingly that it is superior to plausible alternatives (see Chernoff 2002; Humphreys 2019) and they have not spelled out in detail what difference an embrace of causal realism might make for the conduct of causal inquiry in IR. In short, increased interest in getting to grips with philosophical puzzles about causation has not yielded a clear methodological payoff.[12] Consequently, this work has been largely ignored in the research methods literature and by those actually conducting causal inquiries.

A central aim of this book is to bridge this gap: to show how thinking in a careful, philosophically informed way about causation and causal

[10] We will analyse some specific examples in Chapter 8.

[11] We unpack these terms more fully below, and in Chapters 1–3. It is, however, important to note even at this early stage that causal and scientific realism have nothing to do with 'realism' in IR theory; indeed, realism in IR theory is often said to be philosophically anti-realist (see, for example, Joseph 2010).

[12] One exception to this general pattern is Heikki Patomäki's work on historical analogies, iconic models, and scenario building (see, for example, Patomäki 2002 [esp. part II], 2008, 2017).

inquiry can generate powerful methodological insights. Although we will not ourselves advocate causal and scientific realism as a vehicle for improving causal inquiry, we will pay careful attention to the methodological implications of its proponents' arguments. We will also consider established philosophical debates about the relationship between the principal kinds of causal statements and consider what these may reveal about the basic logic of causal inquiry. Our goal is to offer an account of causal inquiry which is rigorous and philosophically informed, but which also speaks directly to some of the methodological problems which routinely arise in the conduct of causal inquiry in IR.

Useful background ideas

Before we start to develop our argument in detail, it may be helpful to indicate how the approach we will take in this book relates to two existing distinctions which are often invoked in discussions of causation in IR and which are hence liable to inform readers' expectations about the form that a philosophically informed engagement with the topic of causal inquiry might take. These are, first, Hollis and Smith's (1991a) distinction between causal explanation and interpretive understanding, which also informs the so-called 'third debate' in IR and, second, the distinction drawn by so-called 'critical realists', among others, between causal realism and causal idealism (see, for example, Patomäki and Wight 2000). We will also briefly touch on the principal existing 'theories of causation' and outline two key assumptions which will structure our analysis.

Previous approaches to causation and causal inquiry in IR

One of the most well-known frameworks for thinking about what causal inquiry consists in and for situating it in relation to other forms of inquiry is Hollis and Smith's (1991a) distinction between 'Explanation' and 'Understanding'.[13] Hollis and Smith contend that in IR, and in the study of the social world more broadly, 'there are always two stories to tell'—causal ('Explanation') and interpretive ('Understanding')—and that these two stories are not easily combined (1991a: 7). On their account, 'Explanation' treats 'the human realm as part of nature'; it is 'modelled on the methods of natural

[13] This distinction did not originate with Hollis and Smith. For an earlier account see, among others, Von Wright (1971).

science and is usually described as [involving] a search for causes' (1991a: 1, 3). By contrast, 'Understanding' seeks to reveal what 'events mean, in a sense distinct from any meaning found in unearthing the laws of nature' (1991a: 1). This vision of the social sciences as drawing uneasily on two categorically distinct modes of inquiry has proved influential. There are, however, two quite serious problems with the picture of causal inquiry that Hollis and Smith present.

The first problem is that Hollis and Smith fail to provide a clear analysis of what causal inquiry consists in. What little they do say is largely contained within a discussion of mid-twentieth century positivism (see 1991a: 45–55). Thus, for example, they state that the 'broad idea' behind causal inquiry is that 'events are governed by laws of nature which apply whenever similar events occur in similar conditions', and they assert that the aim of causal inquiry is to 'detect the regularities in nature' (1991a: 3, 50). This is to be achieved, they suggest, by building up 'a solid body of successful generalizations' which 'can serve as a reliable source of explanations' (1991a: 50–2).

Having presented this account, Hollis and Smith immediately acknowledge the problems with it. They point out, correctly, that positivism's 'heyday is over' (1991a: 46), and they discuss several influential critiques of the positivist vision of science which they have presented (see 1991a: 55–66). More specifically, they warn that 'there is more to "cause" than a constant and predictable correlation between events' (1991a: 50), and they indicate that causal explanation 'involves an appeal to causal laws and not solely to generalizations' (1991a: 63); they even suggest that 'things' may possess causal powers (see 1991a: 63, 200, 207). However, all these hints are left unelaborated; they never flesh out what causal inquiry might consist in *other than* building up a body of generalizations to be used as a source of explanations. They indicate that this image of causal inquiry may be deficient, but they do not articulate a clear alternative.

The second problem is that Hollis and Smith's insistence that causal inquiry is part of 'Explanation' and that 'Explanation' and 'Understanding' cannot be combined is overly restrictive. It precludes the possibility that investigations into the meanings that events hold for people can contribute to a causal account of why they acted as they did: within Hollis and Smith's framework, reasons cannot be causes (cf. Davidson 1963). It also appears to preclude the possibility that the meanings events have for people may themselves have identifiable causes: within their framework, it does not appear possible to provide a causal account of how people arrive at their understandings of the world and how it works. More broadly, the positivist picture of causal inquiry which Hollis and Smith present makes it hard to recognize

what Milja Kurki (2008: 11) terms the 'causal nature' of 'aspects of social life' such as 'rules, norms, ideas, reasons, discourses', and 'structures of social relations'.

Despite these problems, it is important to appreciate how widely the positivist vision of causal inquiry which Hollis and Smith articulate has been accepted in IR. As Kurki (2008) emphasizes, it is embraced not only by many of those who advocate and conduct causal inquiries but also by many critics. This is worth briefly elaborating.

In broad terms, IR's so-called 'third debate' (Lapid 1989) involved the 'dissent' of interpretive and critically oriented scholars from what they characterized as the 'positivist' mainstream of the discipline (see, for example, George and Campbell 1990; Smith 1996).[14] Among other things, this included a rejection of causal inquiry. For example, David Campbell famously asserted his opposition to 'cataloging, calculating, and specifying the "real causes" of an event or process' (1998b: 4). Importantly, Kurki (2008: 124–44) emphasizes that causal inquiry has often been rejected by these interpretive and critical scholars based on an explicit or implicit assumption that it must proceed in a positivist fashion: associating causal inquiry with the attempt to establish empirical generalizations and use these as a source of explanations, interpretive and critical scholars have viewed it as incompatible with their goals (see also Kurki and Suganami 2012).

Our purpose here is not to criticize interpretive and critical scholars who have rejected the predominant account of causal inquiry with which they were confronted. The important point is that the picture of causal inquiry presented in accounts such as Hollis and Smith's is an impoverished one: it is not a good guide to what causal inquiry consists in and it is not a good guide to how causal inquiry relates to other forms of inquiry. We will therefore not take this account as our starting point in this book, despite its undoubted influence in IR.

In more recent years, philosophically oriented discussions of causation have often privileged a second distinction: that between 'causal realism', on the one hand, and 'causal idealism', on the other hand. The broad idea which has motivated the introduction of this distinction is that it will be possible to move beyond the unhelpful opposition between 'positivism' and 'post-positivism' only by recognizing what the two positions in fact have in common and considering what alternative possibilities this excludes.

[14] The binary opposition between 'positivist' scholarship on the one hand and interpretive and critical (or 'post-positivist') scholarship on the other hand has also been framed as an opposition between 'rationalism' and 'reflectivism' (see Keohane 1988) and between 'explanatory' and 'constitutive' theory (see Smith 1995).

According to critical realists, positivists and post-positivists both embrace a form of 'anti-realism', or 'idealism', about causation; the alternative to this is (causal and scientific) 'realism'.[15] The critical realists contend, in short, that in order to develop a more satisfactory account of causal inquiry it is necessary to adopt a realist stance on causation.[16]

According to critical realists such as Heikki Patomäki and Colin Wight (2000), the positivist attempt to discover robust empirical generalizations has led to a focus on observable features of the social world and to a neglect of unobservable features such as social structures as well as ideas, rules, discourses, and practices. In essence, critical realists argue, positivists reduce reality to what can be observed, either neglecting unobservable features of the social world entirely or treating references to them instrumentally. Meanwhile, post-positivists reduce reality to what is constructed through discourse, sometimes even going so far as to assert that 'nothing exists outside of discourse' (Patomäki and Wight 2000: 217, citing Campbell 1998a: 24–5). Critical realists argue that the alternative to focusing on what is experienced, or what is constructed through discourse, is to 'go beyond … appearances and inquire into the nature of things' (Patomäki and Wight 2000: 218)—that is, to investigate what exists in the world independently of human thought and experience. They contend that 'the world is composed not only of events, states of affairs, experiences, impressions, and discourses, but also of underlying powers, structures, and tendencies that exist, whether or not detected or known through experience and/or discourse' (Patomäki and Wight 2000: 223). On their account, the task of science is to investigate these underlying structures, powers, and tendencies in order to discover whether 'things are really as described' (Patomäki and Wight 2000: 218).

According to critical realists, the need to move beyond appearances is nowhere more urgent than in relation to causation. They argue that the positivist approach to causation, outlined by Hollis and Smith, is essentially 'Humean' (see Patomäki and Wight 2000: 219–20; Kurki 2008: 6; see also Hollis and Smith 1991a: 49): because 'what can be meaningfully said of the world' is limited to what can be experienced, positivists reduce causation to 'the constant conjunction of events' (Patomäki and Wight 2000: 220). It is this same, 'Humean', conception of causation, critical realists argue, that is

[15] Although the terms 'idealism' and 'anti-realism' sometimes appear interchangeable, they are not synonymous. As we will show in Chapter 3, one can be an anti-realist in regard to causation without being a causal idealist.

[16] Critical realism owes much to the philosopher Roy Bhaskar (see especially 2008), whose arguments we will consider in Chapter 2. On 'critical realism' in IR see, inter alia, Patomäki and Wight 2000; Patomäki 2002; Wight 2007; Kurki 2007; Joseph and Wight 2010; Jackson 2011: esp. 72–111. For a critical perspective see, among others, Chernoff 2007b; Suganami 2013; Humphreys 2019.

rejected by critical and interpretive scholars (Patomäki and Wight 2000: 220; Kurki 2008).[17]

The critical realist solution is to shift the focus of causal inquiry away from an effort to identify constant conjunctions of events—or to identify robust empirical generalizations capable of being used as a source of explanations—and towards an effort to identify the underlying 'causal powers' which generate these patterns (Patomäki and Wight 2000: 223). They insist, moreover, that this shift can encompass the interests of interpretive and critical scholars, for they acknowledge that social objects are constituted in part by social practices, and by ideas about those practices, and hence they also recognize that causal inquiry must encompass investigation into these practices and ideas; they argue, moreover, that causal inquiry must adopt a *critical* perspective on these practices and ideas, exposing where people's ideas and beliefs are incorrect (Patomäki and Wight 2000: 224–5). The aspiration of at least some critical realists is therefore to reunite a 'divided discipline' by bridging the gap between positivists, who are committed to causal inquiry, and interpretive and critical scholars, who reject it (Kurki 2006).

We share the critical realists' scepticism about the positivist picture of causal inquiry presented by Hollis and Smith, among others. We also share the critical realists' conviction that causal inquiry in the social sciences must encompass investigations into ideas, rules, discourses, and practices, and take seriously the influence of unobservable social structures. However, we are unconvinced about the value of the binary opposition between causal 'realism' and causal 'idealism' which the critical realists have introduced.

The critical realists contend that mainstream approaches to causal inquiry in IR are positivist, and that positivism is 'idealist' in the sense that it reduces causation to constant conjunction and that it regards the idea that things have real causal powers as no more than an idea. However, the ubiquitous invocation of the mantra 'correlation is not causation' suggests that most of those engaged in causal inquiry in IR do *not* simply reduce causation to constant conjunction. Moreover, the aim of causal inquiries in IR is often to enable policymakers successfully to manipulate the world around them in order to achieve their policy goals; this suggests that causation is construed as a productive power.

We are also unconvinced that it is necessary to make a forced choice between causal 'idealism' and causal 'realism'. On the critical realists' account,

[17] We will explore Hume's thought on causation, and some of the ways in which it has been (mis)understood, in Chapter 1.

Hume was an idealist because he reduced causation to constant conjunction, but, as we will see in Chapter 1, the real picture is much more nuanced; this opens up the possibility that there are positions other than 'idealism' and 'realism' on which causal inquiry in IR might be based.

We will investigate these issues in much more detail over the first three chapters of this book. For now, the main point is that despite the recent flurry of arguments extolling the benefits of causal and scientific realism as a basis for causal inquiry in IR, we will not adopt causal realism as the starting point for our own investigations. Rather, we will adopt a more cautious and critical stance, seeking to understand more precisely what the options are and, importantly, what difference the choice between them might make for the conduct of causal inquiries in IR.

Philosophical discussions of causation

These debates within IR have taken place against the backdrop of related developments within philosophy. In particular, the rise and subsequent decline of 'positivism' has significantly shaped the ways in which philosophers have dealt with questions about causation (see, for example, Cartwright 2002). We will touch on some of these issues in Chapter 3. At this stage, the crucial point to note is that philosophers have approached the topic of causation and causal inquiry in a range of different ways. These are encapsulated in what are sometimes described as 'theories of causation' (see, for example, Beebee, Hitchcock, and Menzies 2009b: 5; Reiss 2009: 21). These theories might be considered an obvious starting point for a book on causal inquiry, though philosophers disagree about which theories are the most important (see, for example, Cartwright 2007: 11–12, Beebee, Hitchcock, and Menzies 2009b: 5–7; Reiss 2009: 21–6; see also Brady 2008: 218–19). There is, however, a serious problem with taking such theories as a starting point: there is often considerable ambiguity about what, precisely, they are trying to do.

According to Beebee, Hitchcock, and Menzies (2009b: 7), 'theories' of causation 'attempt to provide a definition or analysis of causation'. According to Julian Reiss, meanwhile, these theories offer accounts of 'what causation consists in or what we mean by the word *cause*' (2009: 21). But what does it mean to offer a 'definition or analysis of causation'? Does it mean providing a definition of the concept of 'cause', thereby clarifying what it *means* to say that one thing caused another? Or does it mean analysing what causation *is*: analysing whether causation is a feature of the world existing independently

of human thought and, if so, what it consists in? Although Reiss runs these two things together, they are clearly quite different.

Reiss in fact expresses scepticism about whether 'a universal theory' of causation is possible: this, he explains, is because he believes that all plausible accounts are 'subject to counterexamples' (2009: 21). However, what Reiss appears to mean by this is that if one takes any 'theory' of causation and uses it as a basis on which to distinguish causal relations from non-causal relations, it is always possible to come up with examples where it fails (see 2009: 21–6). But this seems to introduce yet another dimension of what a theory of causation is concerned with, for providing an account of how causal relations can reliably be picked out on the basis of empirical evidence is quite different from providing an account of what causation *is* or of what the word 'cause' *means*.

This ambiguity about what theories of causation seek to do extends not only across these theories, taken together, but also to individual theories. Consider the two theories which are probably best known in IR: the regularity view and the counterfactual approach.

According to Stathis Psillos, the 'kernel' of what he terms 'the *Regularity View of Causation*' is the idea that 'the constitutive elements of causation are spatio-temporal contiguity, succession, and regularity (constant conjunction)' (2009: 131–2; emphasis original). This idea is closely associated with the work of David Hume, who is probably the philosopher best known in IR for his writing on causation. However, it requires a lot of unpacking. Does it amount to a claim about what causation *is*—a claim to the effect that causation *is* no more than spatio-temporal contiguity, succession, and regularity, and hence that the idea of causal powers is no more than an idea; or is it a claim about what kind of evidence will be required to identify a relation between events of two kinds as a causal relation? Should it even be understood as indicating something about what is meant by the term 'cause': that when we say one event caused another what we *mean* is that these two events were spatio-temporally contiguous and that events of these kinds regularly succeed one another? This kind of ambiguity indicates the need for a close and careful engagement not only with the idea that there is a (single) regularity view of causation, but also with Hume's thought, with which this view is closely associated. We will provide this in Chapter 1.

The central claim of counterfactual theories of causation is that 'causation is to be understood in terms of counterfactuals of the form "if event *c* had not occurred, event *e* would not have occurred"'; when such a counterfactual is true, *c* is a cause of *e* (Beebee, Hitchcock, and Menzies 2009b: 6; see also Paul 2009). As with the regularity view, however, this immediately raises further

questions. Is the central claim of this approach that causation *is* a relation of counterfactual dependence—that counterfactual dependence is all there is to causation as it exists in the world? Or is the claim that when we can identify a relation of counterfactual dependence, then we have identified a causal relation? Or, perhaps less plausibly, is the claim that when we say that one event caused the other, we are saying that the two events stand in a relation of counterfactual dependence?[18]

The underlying problem here is that philosophers have, quite rightly, been interested in a wide range of different, though interconnected, questions about causation, to which they have given a wide range of quite different answers (see Beebee, Hitchcock, and Menzies 2009b: 1; Brady 2008: 218). This makes 'theories' of causation a difficult place to start, even if we were willing simply to hitch our wagon to one particular approach, which we (the authors of this book) are not.

That said, it is, of course, necessary to start *somewhere*. Our starting point will be that the concept of cause describes a relation of *production* or *bringing about*: to say that one event 'caused' another event is to say that it contributed to the production, or bringing about, of this other event. In our view, this is what is meant by the term 'cause' when causal claims are advanced in IR. If it is claimed, for example, that 'revolutions cause wars', this does not *mean* that revolutions are regularly succeeded by wars, or that revolutions and wars stand in a relation of counterfactual or probabilistic dependence; what it *means* is that revolutions *bring about* wars.[19]

In stipulating that this is the *meaning* of the verb 'to cause', we are not (yet) taking a view on two further questions: first, whether causation, thus understood, is real—that is, whether relations of production are part of the world existing independently of human thought—and second, how causes, thus understood, can be identified. We will discuss both these questions in detail in this book, focusing on the first question in Part I and the second question in Part II.

[18] One could also present the same kind of analysis in relation to other theories of causation, such as probabilistic theories, the central claim of which is that 'causes raise the probability of their effects' (Beebee, Hitchcock, and Menzies 2009b: 6; see also Williamson 2009).

[19] By contrast, John Gerring (2005: 167) argues that the 'core, or minimal, definition of causation held implicitly within the social sciences is that a cause raises the probability of an event occurring'. However, this also needs unpacking. It is true, as Gerring (2005: 169) indicates, that *X*-type events cannot be understood as causes of *Y*-type events unless the occurrence of an *X*-type event raises the probability of a *Y*-type event occurring, given appropriate background conditions. Yet, as Gerring also acknowledges, to be a cause something must '*generate*, *create*, or *produce* the supposed effect' (2005: 170). What this reveals is that Gerring's proposed 'definition' of causation is focused on the *epistemic* question of what criteria something must fulfil in order to be identified as a cause, rather on the *semantic* question of what the concept of 'cause' means. In stipulating that the concept of 'cause' describes a relation of production, we are concerned with this *semantic* question, on which our position is closely aligned with Gerring's.

In offering this stipulative definition of the meaning of the concept of 'cause' we are, however, resisting a fashionable view to the effect that there are *multiple* concepts of cause (see, for example, Cartwright 1999: 104–5; Cartwright 2004; Reiss 2009). The motivation for this claim is that it is not possible to identify one single way in which causal relations can be reliably identified. As Cartwright puts it (1999: 118–19): it is a mistake 'to think that there is any such thing as *the* causal relationship for which we could provide a set of search procedures … each causal relation may be different from each other, and each test must be made to order' (see also Reiss 2009: 20). This may be the case, but we see no reason to believe that it entails pluralism *about the concept of causation*—about what the verb 'to cause' *means*. In our view, what it appears to entail is merely pluralism about how causal relations are to be identified.[20] Rom Harré (1964: 353) draws an important distinction between, on the one hand, the claim that a concept, such as the concept of 'cause', has multiple *meanings* and, on the other hand, the claim that there are multiple sets of *criteria* which, if satisfied, allow us legitimately to assert that a causal relation is present. Our view is that while there may be multiple different criteria which can be used, in different situations, to justify a causal claim, there is just a single, central concept of cause: it is a concept of production.

It is important to flag one final assumption which will structure our discussions in this book: we will assume that the things linked by putative causal relations—what are sometimes called 'causal relata' (see, for example, Ehring 2009)—are distinct, spatio-temporally located events and not, for example, objects, or facts, or variables, or event-types.[21] In other words, when we advance a statement of the form '*a* caused *b*', *a* and *b* are specific, separate events (or sets of events), located in space and time. Having said that, we construe the term 'event' extremely broadly, so as to encompass not only short-term changes, but also conditions, states of affairs, and longer-term processes, as well as omissions and the absence of change (see Kim 1975: 48; Ducasse 1975: 115; Lewis 1993: 184). Our aim is not to restrict, a priori, the kinds of things that can be causes or effects, but rather to facilitate accurate analysis of what particular causal statements assert and imply. This will be particularly important in Part II of this book.

[20] Reiss (2009) argues that what he terms 'evidential pluralism' collapses into 'conceptual pluralism', but we do not find his analysis convincing.

[21] Although we contend that a variable, as such, cannot be a cause (or effect), we accept that an event, which on our account can be a cause (or effect), may be summarized as the change in the value of a variable (see Woodward 2003: 112).

Plan of the book

Our analysis and argument will develop in two parts.

In Part I of the book—'Metaphysics'—we will examine the contention, prominently advanced by critical realists, among others, that in order to improve and expand the scope of causal inquiries in IR it is necessary to embrace causal and scientific realism. We will examine the key arguments of the two principal philosophical figures invoked in this debate: David Hume (1711–76) and Roy Bhaskar (1944–2014). We will also assess the case for causal and scientific realism and consider what alternative positions may be available. Crucially, we will pay close attention to the methodological implications of these arguments, identifying some important lessons for the conduct of causal inquiry but also questioning whether a commitment to 'realism' about causation is really required.

Chapter 1 will focus on Hume's well-known arguments about causation and on the so-called 'regularity view of causation', which is closely associated with his thinking. We will show that although there is an important sense in which Hume was a regularity theorist, he did not endorse the versions of the 'regularity view' with which he is most often linked. Relatedly, although Hume is often identified as a causal 'idealist', and hence as a key proponent of the view that causal 'realists' reject, his arguments are more subtle than this characterization suggests; indeed, they point towards a way of thinking about causation which is neither 'idealist' nor 'realist'. In addition, we show that Hume offers some important insights into the significance of observed regularities as evidence for the presence of a causal relationship and into the kind of evidence which would, if available, be even more valuable—that is, experimental evidence.

Chapter 2 will consider Bhaskar's argument for causal realism, which has been a touchstone for advocates of causal and scientific realism in IR and which emphasizes the central role that controlled experiments play in scientific discovery. To paint a very crude picture: Bhaskar argues that the scientific practice of controlled experimentation is intelligible only on the assumption that causal powers are a real feature of the world as it exists independently of human thought. We will analyse this argument closely. Ultimately, we will find it unconvincing: while we accept that causal realism is a possible view to hold and that it constitutes a coherent basis upon which to conduct causal inquiries, we do not believe that Bhaskar succeeds in showing, through his analysis of scientific practice, that causal powers are real. However, we will endorse some important conceptual and methodological

principles which are linked to Bhaskar's arguments, and which we will carry forward into our own analysis of the conduct of causal inquiry in Part II of this book.

In Chapter 3, we will wrap up this part of our analysis by considering whether, as Hume's arguments (properly understood) appear to imply, there is an alternative to causal idealism and causal realism, and whether this alternative provides a viable basis on which to conduct causal inquiries. We will do this by exploring the thinking of the philosopher Bas van Fraassen, who is little known in IR but who is one of the most prominent critics of causal and scientific realism in contemporary philosophy of science. We will show how, on Van Fraassen's account, it is possible to remain agnostic in relation to the metaphysical questions which realists raise about the reality of both causation and unobservable theoretical entities (such as sub-atomic particles, or social structures). Moreover, we will show that this agnosticism is not detrimental to substantive scientific and causal inquiries: the differences between 'realists' and 'agnostics' do not have any direct methodological corollaries. This finding deflates the power of the realist–idealist dichotomy which has informed so much of the recent philosophically informed work on causation in IR. We will show that it is not necessary for those conducting causal inquiries in IR to accept the metaphysical doctrine of causal realism; rather, the key to improving our understanding of causal inquiry is to adopt a more explicitly methodological focus.

This is what we will do in Part II of this book: 'Methodology'. Here, we will offer a new account of the underlying logic of causal inquiry, developed from a careful analysis of the relationship between the two principal kinds of causal statement: statements which refer directly to specific events, located in space and time, and statements which abstract from specific events to describe causal propensities. We will show how the logic of causal inquiry flows out of a simple, but rarely understood, fact: that the only kind of causal statements that can be directly supported by empirical evidence are those which refer to specific events that have already occurred. Having laid out this logic, we will offer a detailed account of what it implies for causal explanation and causal reasoning and show how our analysis can illuminate some substantive examples of high-quality causal inquiries in IR.

We begin, in Chapter 4, by analysing the two principal kinds of causal statement and showing how the 'culture of generalization' impedes accurate appreciation of the differences between them. We will show that whereas causal statements are often distinguished according to whether they are 'singular' or 'general', these labels are highly misleading. The key characteristic

of statements, such as 'the short circuit caused the fire', which are often labelled 'singular', is not that they refer to 'single' events—many do not—but rather that they refer to specific events, located in space and time. The key characteristic of statements such as 'short circuits cause fires', which are often labelled 'general', is not that they are generalizations—they are not—but rather that they abstract from specific events to describe causal propensities. We will also explore the implications of these two kinds of statement, showing that, contrary to what is often claimed in IR, neither kind of statement implies that a regularity of the form 'whenever a short circuit occurs, a fire will occur' should be observable in the world around us. As we have noted already, causal statements concern the unfolding of propensities: hence, what the statement 'short circuits cause fires' *implies* is that a short circuit will produce a fire *under the right conditions*.

Building on this analysis, Chapter 5 lays out the logic of causal inquiry. We start from the simple observation that the only kind of causal statements that can be directly supported by empirical evidence are those which refer to specific events that have already occurred, such as 'the short circuit caused the fire', for it is only statements of this kind that can be checked against empirical evidence. We then explore what kind of evidence will offer persuasive support for a statement of this kind, arguing that what is required is evidence which rules out competing accounts of what caused the effect event—in this case, the fire. We contend that this is true regardless of which method for generating and analysing empirical evidence is used. Finally, we show how knowledge of the causal relations linking specific events which have already occurred can potentially be drawn upon to generate further causal explanations as well as causally informed predictions and policy prescriptions.

Chapters 6 and 7 each offer more detailed discussion of an important theme which emerges out of our analysis of the logic of causal inquiry. One of these themes is that causal inquiry revolves around the provision of empirical support for statements, such as 'the short circuit caused the fire', which may be understood as offering an explanation for the occurrence of the effect event—in this case, the fire. Chapter 6 therefore examines the topic of causal explanation. We ask what makes a causal explanation satisfactory, pointing out that it is not enough for an explanation to provide information that is reliable; the information provided must also be relevant, helping to resolve whatever puzzle motivated the request for an explanation. This 'relevance' requirement stems from the important, but often neglected, 'pragmatic' component of causal explanation, which we explore in detail. We

also correct a series of misunderstandings and 'red herrings' that have often plagued discussions of causal explanation in IR.

Chapter 7 explores a second important theme emerging out of our analysis of the logic of causal inquiry: that generating causal knowledge involves reasoning about when a causal claim should be accepted on the basis of the available evidence, and about the range of conditions under which a known causal relationship can be expected to unfold. These two issues are often discussed under the headings of 'internal' and 'external' validity, respectively, but we go beyond existing analyses in two ways. First, we explore the kind of reasoning required to support a causal inference with high internal validity: we show that this reasoning is neither inductive, nor deductive, but abductive. We lay out what is involved in using abductive reasoning to support causal claims and consider when causal judgements arrived at by means of abductive reasoning will be more, or less, reliable. Second, we indicate some of the ways in which existing discussions of external validity may be misleading and explain why it is hard to generate reliable knowledge of how causal relationships will unfold in situations that have not yet been studied.

Part II of the book concludes with Chapter 8, in which we examine four high-quality examples of causal inquiry in IR, showing how the analysis we have developed in this part of the book can enhance understanding of how reliable causal conclusions may be generated using a range of methodological approaches and of how existing causal knowledge can be expanded and made more robust.

Putting these pieces together, this book does two big things.

First, it deflates the significance of metaphysical debates for discussions of causal inquiry in IR, showing that although such debates hold perennial interest, they do not always have substantive methodological corollaries. In short, and contrary to what has often been claimed in recent work on causation in IR, it is not necessary to embrace causal and scientific realism in order to improve and expand causal inquiries in IR.

Second, the book provides a new account of the logic of causal inquiry, showing how understanding of key aspects of this logic has been hamstrung by the pervasive influence of the 'culture of generalization'. We show that despite the widespread emphasis on the pre-eminent value of 'general' knowledge, all causal knowledge in fact arises out of the provision of empirical support for concrete causal statements referring to specific events that have already occurred, and that this empirical support is provided by ruling out competing accounts of how specific events were brought about. This emphasis on the importance, for causal inquiry, of facts about specific events, located in space and time, may appear surprising to some readers. If so, this

reflects the hold that the culture of generalization exerts over our collective understanding. We not only defend our position in detail, but also spell out its implications and show how our analysis can illuminate and strengthen the conduct of causal inquiry. We thereby demonstrate how a philosophically informed approach to questions about causation and causal inquiry in IR can generate a substantial methodological payoff.

The conclusion to the book will first provide a brief synopsis and then spell out what our analysis suggests about the place of causal inquiry within the discipline of IR, including how it relates to normative, interpretive, critical, and historical inquiry.

PART I

METAPHYSICS

1

Hume and the Regularity View of Causation

One of the key issues that anyone conducting causal inquiries in IR will have to consider is the relationship between causation and observable regularities, or correlations. True, 'correlation is not causation' as '[n]early everyone knows' (Haas 1974: 59). Yet it is also commonly supposed that causation has *something*—rather than *nothing*—to do with correlation, or the regular conjunction of events of particular types. Indeed, the search for patterns of covariation has often been accorded a prominent place within the methodology of causal inquiry, not only in IR, but in the social sciences more broadly (see, for example, King, Keohane, and Verba 1994; Jackson 2016: 58–82). In what respect, then, is causation linked to observable regularities? Finding the right answer to this question is integral to this book's exploration of causal inquiry, not only because of its methodological implications but also because of the various confusions which, we will show, have sedimented around this subject. The aim of this chapter is to remove these confusions and, thereby, to arrive at a clearer account of the relationship between causation and regularities.

An obvious starting point is to take a critical look at the so-called 'regularity view (or theory) of causation'—a much-discussed subject not only in the philosophical literature on causation (Harré and Madden 1975; Mackie 1980; Psillos 2002; Illari and Russo 2014), but also in philosophically informed works in IR (Wight 2006a; Kurki 2008). As a cursory glance at these two literatures will reveal, this so-called 'regularity view' is said to have originated in the writings of the well-known eighteenth-century philosopher, David Hume (1711–76), whose pronouncements about causation have shaped subsequent debates about this subject to an extraordinary degree. We therefore consider it fruitful to begin by interrogating what Hume himself actually argued. We can agree or disagree that he has provided a persuasive account of the relationship between causation and regularities and move on from there.

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0002

This is basically what we will do in this chapter. But it is first worth spelling out more fully our motive for discussing Hume's writings. In what follows, we will spend some time developing a fairly detailed exegetical analysis of Hume's thinking. We do so not because of any 'antiquarian' interest on our part in the history of philosophical thought, but because engaging closely with his texts reveals some subtle but significant limitations of prevailing interpretations of his work—limitations which have fed into discussions of causation in contemporary IR and shaped them in important ways. Let us explain.

There is a dominant view among recent IR writers on causation which claims that the Humean 'regularity view of causation' is challenged by the so-called 'realist view of causation' (Wight 2006a; Kurki 2008). The precise meanings of these two key terms will be clarified in Section II. For now, it is sufficient to know that the 'regularity view of causation' holds, in broad terms, that observable regularities have an intrinsic connection of some kind with causation and/or our idea of it; by contrast, the 'realist view of causation' claims that the essence of a causal relation is not the regularity with which cause-type events are followed by effect-type events, but rather the operation of causal powers which, according to realists, are present independently of human thought and existence. Those who juxtapose these two views consider Hume's position to be the paramount example of 'idealism'; for Hume, they suppose, causation is nothing more than regular conjunction and hence 'causal necessitation' is *only* an idea. But we find this dichotomy between a 'Humean', idealist, 'regularity view' of causation, on the one hand, and 'causal realism', on the other, to be mistaken and misleading. The best way in which to appreciate this is to go a little deeper into Hume's writings.[1]

This will reveal three things: (i) there is no such thing as *the* 'regularity view'—this label in fact subsumes three quite separate doctrines, all of which we will come to reject; (ii) the oft-invoked contrast between 'causal idealism' and 'causal realism' fails to take into account the possibility of a third position—'causal agnosticism'—to which, we will suggest, Hume himself subscribed; and (iii) concealed by the stereotyping label, 'the regularity view',

[1] To clarify the nature of causation, Milja Kurki claims, it is not necessary to study what Hume had himself written about it. Although, in writing her path-breaking book *Causation in International Relations* (2008), she pays a good deal of attention to what she defines as 'Humeanism', she claims that this is a construct, not intended as an interpretation of what Hume's texts tell us about his thought on causation. However, what she sets up as 'Humeanism' is basically what certain critics of Hume have interpreted him to be saying about causation (see Bhaskar 2008) and she (2008: 33–40) frequently refers to Hume's texts to support her claims. What this means is that Kurki is dealing with Hume's thought indirectly (and, to some extent, also directly), and that she is doing so within the presupposed framework of a 'Humean'–'anti-Humean' contrast, based on the juxtaposition between a regularity/idealist view of causation and a realist view. A close reading of Hume's text in this chapter leads us to note, however, that this dichotomous framework of analysis is too restrictive.

and its trade mark, the idea of 'constant conjunction', there are some quite nuanced ideas in Hume's writings concerning the ways in which observed regularities may or may not be relevant to causation, especially as evidence for causal relations. In short, this chapter will discard the so-called 'regularity view of causation' in all its main forms, show how a careful engagement with Hume can help us move beyond the realist–idealist dichotomy which has been so influential in IR in recent years, and open up the question of when, or under what conditions, regularities will constitute persuasive evidence for causal claims.

The chapter is structured as follows. In Section I, we present our exegetical analysis of Hume's writings on causation. Readers who are not yet accustomed to a critical analysis of a philosophical text may find this somewhat daunting. However, Table 1.1, which summarizes our findings, should help remove any uncertainties. We proceed, in Section II, to consider the so-called 'regularity view of causation' in light of what our analysis has revealed to be Hume's argument about the relationship between causation and regularities. As summarized in Table 1.2, we contend that the version of the 'regularity view' to which Hume in fact subscribed is misguided, and irrelevant to this book's concerns, but that he did not subscribe to either of the two, more prominent versions, which we join him in rejecting. We therefore find ourselves in a position to discard the 'regularity view' in all three forms. In Section III, we take a brief look at an important subject on which Hume did *not* develop his thought fully, but on which he nonetheless made some highly significant, though largely neglected, observations, viz. the evidentiary basis of causal judgements. This will pave the way for further consideration of that topic in subsequent chapters.

I. Making sense of Hume's writings on causation

As we begin our analysis of Hume's writings on causation, it may help to explain our approach to making sense of a text and how, accordingly, we structure our discussion in this section. When someone asks a question—especially when it doesn't seem to have an easy answer—it is important to be clear about what they are really asking and to understand why they are asking it. This involves exploring how best the question may be formulated, which may turn out to be somewhat different from how it was initially expressed; what the problem is that has prompted this question; and why this problem appears as 'problematical' at all. This way of approaching a question is perhaps too obvious to require spelling out in so many words. However, it has an important corollary, which is useful to keep in mind when reading a text

on a complex-seeming subject matter, such as 'causation and its relation to regularities', especially when the text, as with Hume's, comes from a historical period or intellectual culture some distance away from our own.

The corollary is this: when, after much complicated argument, an author finally delivers a conclusion and presents it as the answer they have been looking for, in order to understand this answer we, the reader, must be clear about what their question really was, which may not be exactly as they had initially expressed it; what the problem was that had motivated them to raise the question; and why they felt this problem needed to be addressed in the first place.

In the rest of this section, we therefore try to give precise formulations to the questions Hume was asking, identify what his fundamental puzzle was, and briefly outline the background against which his puzzlement arose. We then detail how he responded to this puzzle and enter a few brief remarks about the significance of our interpretation of what Hume was arguing.

Hume's two questions

Hume's 'conclusion' (or what is often mistaken for his conclusion!) is well known, especially among philosophers. It reads: 'we may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*' (1962: 76; emphasis original). This oft-cited statement is the source of a surprisingly common take on Hume that *the concept of causation* was, for him, that of regular conjunction.[2] But such an interpretation of Hume's view is only plausible if we focus solely on that oft-cited sentence and pay no attention to the complex passage in which it was embedded. This reads:

> Our thoughts and enquiries are … every moment, employed about this [causal] relation: Yet so imperfect are the ideas which we form concerning it, that it is impossible to give any just definition of cause, except what is drawn from something extraneous and foreign to it. Similar objects are always conjoined with similar. Of

[2] Hume's thought on causation is more popularly associated with his term 'constant conjunction'. However, he also spoke of '*regular* conjunction' (1962: 92, 96; emphasis added) and '*frequent* conjunction' (1969: 189; 1962: 70, 159; emphasis added) without differentiating the three terms. And, even though he was interested in 'the idea of *necessary* connexion' as the key feature of the concept of causation, he consistently referred to the consequent of a cause as its '*usual* attendant' (1962: 75, emphasis added). Clearly, he was aware that a conjunction observed may not be *truly constant* because there may well be some exceptions (1962: 86–7; 1969: 182–83); his analysis was intended to encompass correlations, as well as strict regularities. Still, he accorded a central place to 'regular conjunctions' and not to the fact that there may be exceptions, which is a feature shared by many contemporary writers (e.g., King, Keohane, and Verba 1994). In explaining Hume's thought on causation here, we will treat 'regular conjunction' as a convenient blanket term to cover 'constant' and 'frequent' conjunction unless there is a need to be more specific.

> this we have experience. Suitably to this experience, therefore, we may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second.* Or in other words *where, if the first object had not been, the second never had existed.* The appearance of a cause always conveys the mind, by a customary transition, to the idea of the effect. Of this also we have experience. We may, therefore, suitably to this experience, form another definition of cause, and call it, *an object followed by another, and whose appearance always conveys the thought to the other.* But though both these definitions be drawn from circumstances foreign to the cause, we cannot remedy this inconvenience, or attain any more perfect definition, which may point out that circumstance in the cause, which gives it a connexion with its effect.[3] (1962: 76–7; emphasis original)

These remarks are somewhat bewildering at first glance; they are a mixture of claims and caveats expressed in unfamiliar language. To understand the ideas Hume was conveying, we will need to study this passage in the context of his overall project. However, even a cursory glance is enough to reveal that, having given one definition of 'a cause', and then offered an apparent reformulation of it, Hume went on immediately to present another quite different definition. This is very odd.[4] What was he doing? To consider this, we must first identify, as best we can, the questions which Hume was seeking to answer.

We believe that the best way to make sense of Hume's two definitions is to treat them as his answers to two parallel questions, which he did not himself present explicitly:

Q1: What, to the best of our knowledge, can we claim goes on *in the world* when we say what we are witnessing is a case of one event causing another?

Q2: What, to the best of our knowledge, can we claim goes on *in the mind* when we say what we are witnessing is a case of one event causing another?[5]

[3] In his earlier work, *A Treatise on Human Nature*, first published in 1739/40, Hume had included 'precedence' of cause over effect and their 'contiguity' in his *definitions* (1969: 220). In the *Enquiry*, published in 1777, which Hume announced should be seen to supersede his earlier work (1962: 2), the idea of temporal priority of cause over effect is retained, as indicated in Hume's reference to 'an object, *followed by* another', but not that of spatial contiguity.

[4] Another oddity in this passage is the fact that, in presenting what we just referred to as an 'apparent reformulation' of the first definition, Hume wrote: 'Or in other words *where, if the first object had not been, the second never had existed*' (1962: 76; emphasis original). However, as some philosophers have pointed out, this does not appear to be a restatement of the first definition 'in other words', but 'proposes something altogether different: a counterfactual analysis of causation' (Lewis 1975: 181). Hume seems oblivious to this issue. Although this 'apparent reformulation' is often pointed to as being significant in discussions of Hume in IR (see, for example, Goertz and Mahoney 2012: 75–83), it is tangential to our analysis of his thought on causation.

[5] We use the expression 'in the world' (or, more fully, 'in the relevant segment of the world'), contrasted to 'in the mind', interchangeably with 'in the objects'.

Rather troublesomely, however, it is immediately obvious that neither Q1 nor Q2, as we have formulated them, is a request for a 'definition' of 'a cause' as this is usually understood—that is, as a definition of *the concept of causation*. It may rightly be asked, therefore, how it is even possible to suggest that Q1 and Q2 were the questions to which Hume offered his two *definitions* of 'a cause' as a response.

The *Oxford English Dictionary* points to an answer: it reveals that one of the key meanings of 'to define' is '[t]o state exactly what (a thing) is; to set forth or explain the essential nature of'; this predates what has come to be a more common meaning, namely '[t]o set forth or explain what (a word or expression) means; to declare the significance of (a word)'. In Hume's time, the *OED* reveals, the verb was used in both ways. Galen Strawson amplifies this point: '[a] definition of a natural phenomenon [as opposed to a definition of a concept] … records *human understanding's best take on that phenomenon*' (2007: 47; emphasis added). On this view, Hume was not *defining the concept* of 'cause', or stating its meaning, but rather presenting what he thought could be known about the nature of the phenomenon of 'causing' (see Stroud 1977: 89). Unsurprisingly, he turned to the physics of his day for answers.

In Hume's time, 'human understanding's best take' on natural phenomena was epitomized in Newton's physics, to which the idea of gravitational force was central. However, while Newton identified the well-known law of universal gravitation which is named after him, and which enables calculation of the magnitude of the 'pull' between two physical objects placed at a given distance, he could not explain by his empirical method what this thing called a 'gravitational force' really was, or how it produced this apparent 'pull', because neither of these could be observed by our senses (Newton 1995: 442–3).

Hume's treatment of causation in the objects was broadly in line with Newton's empiricist approach to gravity. When one billiard ball hits another and the latter moves on, Hume famously claimed, all we *see* is one billiard ball hitting another and the latter moving on; we do not see the first event (the hitting) *causing* the second event (the moving) (1962: 63). Or, as Newton might have said, when we see an apple fall off a tree, we *see* just that, and neither the gravitational force pulling the apple towards the ground nor how it is doing the pulling (see Hume 1962: 30). As illustrated by these examples, Hume maintained that the intrinsic nature of 'a cause', or what goes on in the world when we say causing goes on, is concealed from our senses and hence cannot be known (1962: 63–4). Hume is quite firm on this point; even with respect to those cases where, *according to what we are accustomed to thinking*,

our mind exerts a causal power over parts of our body and makes our limbs move, he argued that we have no experience of the actual operation of such a causal power itself (1962: 66–7).

So, according to Hume, a perfect 'definition' of 'a cause'—one that identifies 'the circumstance in the cause, which gives it a connexion with its effect' (1962: 77)—cannot be provided: it is outside the range of secure, empirically grounded knowledge. This does not imply, however, that inquiries into the nature of causation must come to a stop. Hume noted that we can and do have the experience of observing the regularity with which one type of event—be it the movement of a billiard ball or the motion of our limbs—follows another type of event which we take to be its cause. 'Suitably to this experience', Hume was saying, we can define 'a cause' as '*an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*' (1962: 76; emphasis original).

This helps to clarify what Hume meant in the somewhat opaque passage quoted earlier when, intriguingly, he said of his first definition of 'a cause' that it is formulated in terms of features 'extraneous and foreign to it' (1962: 76). Why should he enter such a caveat? It was because, according to him, the *intrinsic* nature of what we take to be 'a cause'—that 'which gives it a connexion with its effect' (1962: 77)—could not be revealed; it was therefore necessary to 'define', or elucidate, what happens in the world when causing takes place solely in terms of its external manifestations. Unsatisfactory though he apparently felt this resulting 'definition' was, he was resigned to accepting this epistemic predicament because, not having sensory access to the intrinsic nature of causation, it was not possible, he thought, to know or say what causation *really is* in the world; 'we cannot remedy this inconvenience', he remarked (1962: 77). All we can know about causation in the world is its outward manifestation as regular conjunction, he was claiming. This, we believe, is the true import of Hume's first 'definition'. It was not offered as a definition of the concept of causation; rather, he was trying to explain what causation in the world was, as understood at the forefront of scientific knowledge in his time.

In short, when he offered his first 'definition' of 'a cause', Hume was addressing the question we have formulated as Q1: 'What, to the best of our knowledge, can we claim goes on *in the world* when we say what we are witnessing is a case of one event causing another?' His response was along the following lines: 'Actually, we can't say anything at all beyond offering a comment that we regularly observe a similar sequence'!

In a parallel fashion, we can now see that when Hume offered his second 'definition' of 'a cause' as '*an object followed by another, and whose appearance*

always conveys the thought to the other' (1962: 76–7; emphasis original), he was summing up the outcome of his exploration of Q2: 'What, to the best of our knowledge, can we claim goes on *in the mind* when we say what we are witnessing is a case of one event causing another?'

Note that whereas Q1 concerns what we can really know about what goes on *in the world* when we say that what we are witnessing is a case of one event causing another, Q2 is about what we can reveal goes on *in the mind*. Unlike the former, which Hume was able to address by borrowing an insight from Newton's physics, in dealing with the latter there was no comparable science to speak of. Hume therefore relied on *introspection* in search of 'human understanding's best take'. Having engaged in a process of self-reflection, he revealed what he thought he had discovered: what we can say is that, at the appearance of the first type of event, the mind is always led to think of the second type of event in the sequence—this unfailing association of one type of event with the other is what goes on in the mind. This was what he was asserting when he formulated his less well-known second definition of a cause as '*an object followed by another, and whose appearance always conveys the thought to the other*'.

Hume's fundamental puzzle

As we delve further into what Hume wrote, it becomes clear that his two definitions of 'a cause' are not actually final 'conclusions' that complete his story, although some philosophers, focused on analysing the concept of causation, have used them as though they were detachable conclusions concerning how causation should be defined (see, for example, Lewis 1975). In fact, Hume's two 'definitions'—or, better, his two questions-and-answers—are interlinked and form part of a larger set of moves he implemented to address what he found to be really puzzling about causation.

The fundamental puzzle for Hume was how to account for the fact that we have the idea of causal necessitation at all. Hume's basic aim, in other words, was to give an account of the origins of our idea that a cause is not merely followed by, but actually *necessitates, or brings about*, its effect.[6] But why did he find the origins of this idea so intriguing? To cut a long story short, his puzzlement stemmed from: (i) his inherited assumption that all our ideas are derived from experience; (ii) his considered judgement that

[6] The title of the section in which Hume offered his definitions of a cause, first in *A Treatise of Human Nature* (1969) and later in *An Enquiry Concerning Human Understanding* (1962), is revealing: 'Of the Idea of Necessary Connexion'.

there is no experience corresponding to our idea of causal necessitation; and (iii) his acknowledgement that, nonetheless, we undoubtedly do possess the idea of causal necessitation. Given this background, we can appreciate Hume's puzzlement: if *all* our ideas are derived from our experience and if we have *no* experience of causal necessitation, yet we *do* possess the idea of causal necessitation, we will want to ask 'How can we square that triangle?' That was precisely Hume's fundamental problem, his underlying puzzle.

Let us amplify this a little further. Hume's belief that all our ideas are derived from our experience took the form of a doctrine known as 'the copy theory of ideas', which he inherited from his seventeenth-century predecessor, John Locke (1996). This old and now-defunct doctrine claims that *all* our ideas are faint copies, left behind in the mind, of the more vivid original perceptions (which Hume called 'impressions'). For example, a billiard ball hitting another will generate a visual perception of that event and leave behind a 'copy' of it in the observer's mind. According to the 'copy theory of ideas', this 'copy' which is left behind is none other than the 'idea' of 'a billiard ball hitting another'—an idea which is less vivid than the original perception but is capable of being called to the mind independently.[7]

Taking this doctrine for granted, Hume believed that it was important to identify from which 'impressions' our existing ideas have been copied so as to double-check that these ideas are not just empty words without grounding in experience (1969: 122). Unfortunately, Hume believed, there is no '*impression*' to which the idea of causal necessitation corresponds. Yet, we have the idea! From what experience, then, is the idea of causal necessitation formed, and how, through what kind of process? This was Hume's fundamental puzzle (FP):

> FP: Given that, as Hume claimed in responding to Q1, all we see when (what we take to be) causing occurs is a regular conjunction of events, and that we do not perceive such a thing as causal necessitation at all, from what experience and through what process is the idea of causal necessitation formed?

Philosophers differ over how Hume sought to resolve this puzzle and how successful he was.[8] This debate need not detain us, for it is not our aim either

[7] For Hume, 'impressions' are what we would nowadays, in common-sense terms, call 'perceptions', which the mind experiences, and include not only 'sensations', such as those of colour and taste, but also 'sentiments', or inner feelings, such as desire, aversion, hope, and fear (1969: 49–55).

[8] In addition to Hume's two books studied here (1962: esp., 75; 1969: esp., 217), readers wishing to pursue the matter further may find it useful to compare some of the following secondary sources: Kneale 1949:

to establish the origins of the idea of causal necessitation or to present a definitive interpretation of Hume's account of its origins. However, let us very briefly summarize what we take to be Hume's key moves.

According to Hume, we cannot make an inference from observing just one event (e.g., one billiard ball hitting another) to its consequence (e.g., the second ball moving on) by studying that first event alone and contemplating its necessary consequence in an a priori manner (1962: 27–30). But observing the movements of the two balls in sequence is not enough either (1962: 63), whether we observe the sequence just once or even a number of times (1962: 75); what was invisible in the first instance (that is, causal necessitation) remains incorrigibly so however many times we try to observe it. Nevertheless, Hume famously claimed, as we observe the similar sequence of events repeatedly, our mind comes to make an *inference* from one type of event (e.g., the first ball hitting the second) to the other (e.g., the second ball moving on).

This inference, or the unfailing movement of thought from one type of event to the other, which Hume was pointing to in his answer to Q2, is not, according to him, based on any kind of intellectual reasoning. This, he explained, is because no reasoning could possibly validate such an inference; on his account, this inference is grounded not in reason, but rather in *natural instincts* (1962: 32–9, 46–7, 55): the mind—which has observed a regular conjunction of two kinds of events and is now witnessing an event of the first kind—is, by 'custom' (1962: 43; 1969: 152) or 'habit' (1962: 43; 1969: 152), 'induced to expect' (1962: 36) its usual attendant.[9]

This line of thinking enabled Hume to formulate what we take to be his answer to his fundamental puzzle:

- as a consequence of observing a regular conjunction between events of two types, our mind, while not 'impelled by reason', is nonetheless 'determined by custom' (1969: 205) to expect what Hume, in addressing Q2, characterized as an unfailing association between them;
- the inner feeling, when witnessing an event of one kind, of being made, by what we would nowadays call 'conditioning', to expect the occurrence of its usual attendant is the experience from which our idea of causal necessitation is formed;

54; Basson 1958: 76–8; Stroud 1977: 85–6; Noonan 1999: 142–3; and an extensive analysis by Beebee 2006: 82–91, 110–119, 146–47; as well as Read and Richman 2007.

[9] On his choice of the word 'Custom' here, Hume explained: '[W]herever the repetition of any particular act or operation produces a propensity to renew the same act or operation, without being impelled by any reasoning or process of the understanding, we always say, that this propensity is the effect of Custom' (1962: 43). He used 'custom' and 'habit' interchangeably.

- this inner feeling is projected onto the world to generate the idea of causal necessitation in the objects (1969: 217).[10]

Hume's thinking here is complex and difficult to comprehend fully. However, what is important for our purposes is to have established the nature of Hume's fundamental puzzle: it was a puzzle in which he took for granted that our *concept of causation* is one of necessitation, or bringing about, but found it hard to reconcile that with his assumption that all ideas are derived from experience, given that he was unable to identify any experience which corresponds one-to-one to our idea of causal necessitation, the only experience we seem to have being that of observing the regularity with which one kind of event follows another.

In concluding this section, let us summarize the three key findings we have made so far.

First, Hume was not providing definitions of 'a cause' in the sense of defining what the term 'cause' should be understood to mean. A failure to grasp this point leads to an unnecessary debate about which of the two definitions was the more important for Hume, how the two definitions are related, and so forth. As we have explained, he was reporting to us, to the best of his knowledge, what we can claim goes on in the world and in the mind when we say causing goes on. His two 'definitions', therefore, were his responses to what we have presented as Hume's two parallel questions: Q1 and Q2.

Second, addressing these two questions was *not* Hume's ultimate aim. His deliberations and his answers to these two questions were interlinked as elements of his broader attempt to resolve his fundamental puzzle. In this sense, there is no order of priority in Hume's thinking between the two 'definitions' of 'a cause' which he provides.

Third, Hume was driven to investigate this puzzle because he accepted the copy theory of ideas, judged that we do not have any sense experience of causal necessitation as such, yet acknowledged that our idea of causation does involve the notion of causal necessitation. This account of his thinking is summarized in Table 1.1.

We believe that Hume was right to say that our *idea* of causation involves the idea of one event necessitating, or bringing about, the other. He was also

[10] Hume's answer, outlined here, combines two theories about the workings of the mind: the theory of 'conditioning', presaging Pavlov's famous discovery, and a theory some philosophers label 'projectivism'. Hume's fundamental puzzle did stem from the old copy theory of ideas, but his response to it, as outlined here, betrays his innovative intent to go beyond that theory. Indeed, his treatment of the subject of causation was a key element of his ambitious project to inaugurate an empirical science of human nature and understanding, which we would now call 'psychology', inspired by what Newton had done in his science of nature (see, for example, Hume 1969: 14).

Table 1.1 Questions and answers structuring Hume's thought on causation

	Hume's questions identified	Hume's answers explained
Q1	What, to the best of our knowledge, can we claim goes on *in the world* when we say what we are witnessing is a case of one event causing another?	Actually, we can't say anything at all beyond offering a comment that we regularly observe a similar sequence. (This is his first 'definition.')
Q2	What, to the best of our knowledge, can we claim goes on *in the mind* when we say what we are witnessing is a case of one event causing another?	What we can say is that, at the appearance of the first type of event, the mind is always led to think of the second type of event in the sequence and that this unfailing association is what goes on in the mind. (This is his second 'definition').
FP	Given that, as Hume claimed in responding to Q1, all we see when (what we take to be) causing occurs is a regular conjunction of events, and that we do not perceive such a thing as causal necessitation at all, from what experience and through what process is the idea of causal necessitation formed?	As a consequence of observing a regular conjunction between events of two types, our mind is 'determined by custom' to see what Hume, in addressing Q2, characterized as an unfailing association between them; the inner feeling, when witnessing an event of one kind, of being made to expect the occurrence of its usual attendant is the experience from which our idea of causal necessitation is formed; this inner feeling is projected onto the world to generate the idea of causal necessitation in the objects.

probably right to insist that we have no experience of causal necessitation as such. And the question of from where and how our idea of causal necessitation arose is not a trivial one, though the reason why Hume found this so fundamentally puzzling has to do with his acceptance of the now-defunct copy theory of ideas, which we reject as utterly implausible. However, what is vital for us is the understanding, gained through the exposition so far, that Hume did not define a causal relation as a regular conjunction of events, and that regularities played a key role in Hume's thinking about causation mainly because he found, in the experience of observing such regularities, a possible clue as to how he might answer his fundamental puzzle. This understanding lays the ground for evaluating the so-called 'regularity view of causation', which is what we will do in the next section.

II. Against the 'regularity view of causation'

It is often supposed that there is a—single—'regularity view' of causation, and, moreover, that Hume subscribed to this view. In this section we will

show how our exposition of Hume's thought in the previous section helps to distinguish three distinct versions of the 'regularity view': what we will call its 'psychological', 'semantic', and 'ontological' versions. Our analysis suggests that it was the psychological version that was central to Hume's thought, but, as we will explain, we have no reason to accept this version of the regularity view. Meanwhile, although the semantic and ontological versions of the regularity view are often associated with Hume, our analysis shows that they should not be attributed to him at all. Indeed, we find that Hume's arguments, analysed in the previous section, highlight very good reasons for rejecting both of these versions of the regularity view. With Hume's help, therefore, we find ourselves in a position to reject the so-called 'regularity view of causation' in all its three main versions. Our argument is summarized in Table 1.2.

The psychological version of the regularity view

The idea of regularity was undeniably a key element of Hume's argument about causation. It is not inappropriate, therefore, to speak of '*Hume's* regularity view'. However, it is important to be clear about the sense in which Hume was, and was not, a regularity theorist of causation, given the prevailing tendency to categorize him in a blunt way as 'the progenitor of the regularity view'. In this respect, it is vital to appreciate what our exposition in Section I has shown—namely, that in Hume's thought the role of regularity was restricted almost exclusively to the part it played in his attempt to formulate a psychological explanation of the origins of the idea of causal necessitation. It is important to have established this point; and we shall call Hume's own 'regularity view of causation' its 'psychological version'. It asserts that the observation of a regular conjunction of events plays an intrinsic role in generating the idea of causal necessitation because it produces the feeling, in the observer's mind, of being compelled to expect an event of one kind when witnessing an event of another kind.

However, we have no reservation in rejecting this 'psychological' version of the regularity view and moving on without further ado. Importantly, whereas Hume's aim was to account for the origins of the idea of causal necessitation, this is not our concern. Given that Hume's psychological thesis about the significance of our experience of observing regularities is tied to this aim, it is largely irrelevant to our interests in this book, regardless of the quality of his argument. In addition, Hume's argument is flawed in a number of ways, and we believe that these flaws are sufficient to preclude acceptance of his 'psychological' version of the regularity view of causation. Let us briefly point to just three areas of concern.

First, Hume's assertion that repeated observation of the same sequence of events gives rise to the feeling of being compelled to expect an event of one kind, when witnessing an event of the other kind, is no more than a semi-plausible speculation. One obvious problem with it is that it is not clear that we do all have this 'feeling' (Beebee 2006: 87).

Second, it is an important feature of contemporary causal thinking that a distinction is drawn between two types of regular conjunction: (i) where the prevailing conditions in a study sample are such that an *A*-type event is usually followed by a *B*-type event because the former causes the latter; and (ii) where the prevailing conditions in a study sample are such that an *A*-type event is usually followed by a *B*-type event because they are successive manifestations of an underlying common cause. The latter kind of situation is usually called a case of '*spurious* correlation' to acknowledge that there is no direct causal connection in the study sample between *A*-type and *B*-type events; they are collateral effects of an underlying common cause. But Hume's 'psychological version of the regularity view of causation' does not differentiate these two types of regular conjunction.[11]

Third, in advancing the 'psychological version of the regularity view of causation', Hume focused entirely on what goes on in one's—in fact, his own!—individuated mind. If we (the authors) were interested in explaining the origins of the idea of causal necessitation, we would not work within this individualistic, empiricist paradigm. Hume's approach neglected the important social and historical dimensions of the sources of ideas and, in particular, their embeddedness in socially and historically formed language use (see Collingwood 1937–8; Durkheim 1976: 362–9; Elias 1991: 9; Scheler 2017).

The semantic version of the regularity view

It is a troublesome feature of Hume's first 'definition' of 'a cause' as '*an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*' (1962: 76) that, if taken in isolation and at its face value, it can easily mislead the reader into thinking that he subscribed to a version of the regularity view according to which the verb 'to cause' *means the same as* 'to be regularly followed by'. We call this the 'semantic version' of the regularity view because it is a thesis about the *meaning* of the verb 'to cause'.

[11] Here, we are reminded of Thomas Reid's well-known jibe that it follows from what Hume says in his first definition 'that *night is the cause of day*, and day the cause of night' (1827: 603). Of course, 'night' (or absence of daylight) and 'day' (the presence of daylight) are collateral effects of the rotation of the Earth around its own axis, which causes the night–day alternation.

This version is echoed by what John Mackie has called 'a regularity theory of the *meaning* of causal statements'. This holds that causal statements are simply statements about the regular succession of events of different types. However, just as Mackie is emphatic that Hume was 'very far from holding a regularity theory of the *meaning* of causal statements' (1980: 20), so we hold very firmly that Hume did not subscribe to the semantic version of the regularity view; he does not believe that our *idea* of causation is merely an idea of regular succession. We have shown, and agree with Mackie, that 'Hume is quite confident that our idea of causation includes an idea of necessary connection' (1980: 19).

It simply could not be otherwise. Hume's fundamental puzzle would not have arisen had he not accepted that the idea of necessitation, or bringing about, is *integral to the concept of causation*; it was the origin of this very idea that he was trying to explain! The surprisingly common supposition that Hume equated the *concept* of causation with that of mere regular conjunction, rather than necessitation, must be rejected conclusively. It appears tenable only if we focus solely on one brief statement, his first definition, taken out of context, as if this very brief statement were enough to represent Hume's quite complex argument about causation.[12]

Importantly, Hume's concept of cause, which includes the idea of necessitation as its indispensable component, accords well with what is commonly meant by the verb 'to cause' in contemporary everyday usage and philosophical writings. As Jaegwon Kim remarks, '[m]ost philosophers will now agree that an idea of causation devoid of some notion of necessitation is not *our* idea of causation—perhaps not an idea of causation at all' (1993: 234; emphasis original; see also Anscombe 1975: 65). Similarly, Richard Taylor observes that the idea of 'making happen' appears ineradicable from the concept of 'causing' (1975). For Hume and most others, philosophers or not, 'to cause' does not mean 'to be always or usually followed by' but 'to necessitate, to bring about'. We see no reason to challenge or depart from this common understanding and usage. We therefore reject the semantic version of the regularity view. Throughout this volume we use the verb 'to cause' interchangeably with 'to bring about', 'to make happen', 'to produce', 'to necessitate', and the like. But it is important to understand that our point here is a *semantic* one: we are taking a view about what the concept of cause *means*; we are not advancing an *ontological* claim about what really, mind-independently, happens in the world when we say 'causing' goes on.

[12] We should also take note of Hume's own clear statements regarding 'the idea of causation': 'Shall we then rest contented with these two relations of contiguity and succession, as affording a compleat idea of causation? By no means. An object may be contiguous and prior to another, without being consider'd as its cause. There is NECESSARY CONNEXION to be taken into consideration; and that relation is of much greater importance, than any of the other two above-mention'd' (1969: 125; emphasis original).

The ontological version of the regularity view

Hume's first definition, again taken in isolation, could be interpreted, however, to mean that *what there is to causation in the world* is nothing more than a regular conjunction of events. To distinguish this from the semantic version, we call this the 'ontological version' of the regularity view. Because it holds that there is nothing more to causation in the world than the regular conjunction of events, this version implies that the idea of causal necessitation is merely an idea in the imagination. This position is therefore also known as 'causal idealism'.[13]

The 'regularity view of causation' is often equated with this 'ontological version' and it is fairly common to suppose that Hume was an 'idealist' (as noted, inter alia, in Beebee 2006; Wight 2006a: 29; Lebow 2014). However, the analysis of Hume's thought in Section I enables us to appreciate that this, too, misconstrues Hume's fundamental stance on causation. Demonstrating this point requires more work, however, than showing that Hume did not subscribe to the 'semantic version' of the regularity view. Let us elaborate.

Hume's focus, as we now know, was on how the idea of causal necessitation arises *in the mind*. This is not about what happens—in particular, whether 'necessitation' happens—*in the world* when causing occurs. As for what happens in the world, his concern was solely with *what we can know about it from an empirical point of view*, and he insisted that 'that circumstance in the cause, which gives it a connexion with its effect' (1962: 77), is 'wholly unknown to us' (1962: 54) and that we cannot have 'any comprehension' of it (1962: 96). This, it will be recalled, was the rationale behind Hume's first definition, which he felt compelled to formulate in terms of its *only knowable*, though 'extraneous and foreign', feature (1962: 76), viz. the regular conjunction of events of certain types.

Clearly, however, if one's claim is that we cannot know or understand what really happens in the world when causing occurs, because, as Hume reasons, we have no empirical means of accessing it, one cannot consistently claim to

[13] It may be wondered why it is appropriate to characterize 'causal *idealism*' as an '*ontological* position' at all, given that it *denies* that causal necessitation is part of the world existing independently of human thought. However, this would be a mistake. An 'ontological' position is a *metaphysical* stance we take about what exists in the world independently of human thought or existence, and thus the view that X does *not* exist in the world is just as much an ontological position as the contrary view that it does. In short, causal idealism and its pair-opposite, causal realism, claiming causal necessitation to be a mind-independent feature of the world, are *both* ontological positions. This is worth stressing because some realist critics of causal idealism claim that their position, *unlike* '*causal idealism*', is 'ontologically grounded'. This has the *rhetorical* impact of privileging causal realism over idealism by attributing the positive-sounding quality of 'ontological groundedness' to the former but not to the latter (see Kurki 2008: 10, 11, 26, 33, 155–61, 173, 294, 296).

know that *nothing* happens other than the mere regular succession of certain kinds of events. To present Hume as a causal idealist is to fail to notice the important difference between (i) *pleading our inescapable ignorance of what, if anything, happens*, which he did, and (ii) *claiming to know that nothing happens*, which he rightly did not.

Admittedly, in the *Treatise*, Hume did write that 'upon the whole, necessity is something, that exists in the mind, *not in objects*' (1969: 216; emphasis added), which suggests that he was advancing a causal idealist line. However, a few pages later he characterizes what is 'not in the objects' (or the world) as an '*intelligible* connexion' (1969: 219). This is significant; he was pointing out that there is no connection in the objects that is intelligible to us humans because we lack sensory access to any such connection. Correspondingly, in the *Enquiry*, we find him stating: 'upon the whole, there appears not, throughout all nature, any one instance of connexion *which is conceivable by us*' (1962: 74; emphasis added). One meaning of the verb 'to conceive', according to the *OED*, is 'to form a mental representation or idea of'. Hume's claim seems to be that there appears to be no instance of connection anywhere in the world that we can comprehend by forming an idea of it; we simply have *no idea* about what it really is (because we just can't see it or sense it in any other way)!

However, if all Hume was denying was the existence of any *comprehensible* connection between cause and effect in the world, it may be that he did believe that there *was* a real connection between them, his point being purely that we could not comprehend the nature of that really existing connection. Interestingly, in the *Enquiry*, Hume wrote in various places as if he took it for granted that there *are* causal powers in the world which really bring about or necessitate their effects. These include statements such as the following: 'since the particular powers, by which all natural operations are performed, never appear to the senses'; 'though we are ignorant of those powers and forces, on which this regular course and succession of objects totally depends'; 'many secret powers lurk in it [a human body], which are altogether beyond our comprehension' (1962: 42, 55, 87).[14]

So, might it be that Hume was a causal realist at heart? There is a very complex and intense debate among philosophers on this issue. However, unlike a realist such as Roy Bhaskar, whose ideas we will examine in chapter 2, Hume did not formulate any explicit *philosophical defence* of causal realism. The interpretation of Hume as, *philosophically*, a causal realist therefore

[14] But, in one such instance (1962: 33 n. 1), he explained that he was only talking loosely, in a popular fashion, and referred the reader to his main argument, which he developed later in his book, concerning the origins of *the idea of causal necessitation*.

goes too far in the opposite direction from the other, but equally mistaken, view that he was a causal idealist (Read and Richman 2007; compare Beebee 2006).

In our view, Hume was neither a causal idealist nor a causal realist; he is best understood as a causal *agnostic*, whose main point was that we should refrain from talking *knowingly* about the 'operation of causal powers' and the like because such things are *beyond our capacity to know*. We believe this interpretation to be more plausible than either the causal idealist or causal realist take on Hume's thought.[15] Moreover, our interpretation of Hume as a causal agnostic is suggestive of the limitations of the realist–idealist dichotomy, which, as we noted at the beginning of this chapter, has dominated recent debates in IR about the nature of causation, under the influence of causal realists such as Bhaskar.

In Chapter 2, we will come to outline and assess Bhaskar's defence of causal realism; we will explore causal agnosticism further in Chapter 3. For now, the key point is that, like Hume, we in this book do *not* subscribe to causal idealism. We have two main reasons for this. First, and most fundamentally, it is unclear how one could come to *know* that all there is to causation in the world is regular conjunction. Second, there is an uncomfortable tension in, on the one hand, accepting, as most of us do, that the concept of causation is that of necessitation and characterizing one's inquiry into the world as a 'causal' one and, on the other hand, claiming to *know* that such necessitation is not a feature of the world one is investigating. 'Causal idealism', in other words, does not provide an intellectually coherent basis on which to engage in causal inquiry.[16] The idea of causal necessitation is not, in any case, *just* an idea in the way in which, for example, the idea of a unicorn is. The idea of causal necessitation, or bringing about, is a central feature of both scientific and lay discourse; it helps make sense of the world, and it is hard to do without. It is, indeed, difficult to imagine a discipline of IR from which the idea of causal necessitation in the world has been excised.

In light of these considerations, we find ourselves in a position categorically to reject all three versions of the regularity view of causation which we have distinguished. As we have explained, the psychological version of the regularity view makes a number of dubious assumptions and claims and, although it is a key component of Hume's explanation of the origins of the

[15] Interestingly, whereas Ned Lebow (2014) had previously identified Hume unreservedly as a causal idealist, he has, more recently, come in part to acknowledge that Hume might be interpreted as a 'causal agnostic' (see Lebow 2020: 136, 137, 144; see, however, 2020: 195, 207, 286).

[16] By contrast, we believe that causal realism *does* provide an intellectually coherent basis on which to engage in causal inquiry. However, as we will make clear in Chapter 2, we do not find the arguments *for* causal realism which are most often cited in IR to be convincing.

idea of causation necessitation, this does not concern us; we can therefore discard this version of the regularity view. Meanwhile, Hume himself did not subscribe to either the semantic or the ontological version of the regularity view, although both are quite commonly ascribed to him. We join with Hume in rejecting these versions: in our view, it is quite clear that our concept of causation involves the idea of bringing about, and not just the idea of regular succession, while we agree with Hume that our inability to observe causal necessitation does not imply that it is nothing more than an idea. In any case, we find 'causal idealism' incoherent as a basis for causal inquiry and we find Hume's apparent agnosticism suggestive of the limitations of the realist–idealist dichotomy that is often invoked in discussions of these issues in IR. This analysis is summarized in Table 1.2.

Table 1.2 Three versions of the regularity view, Hume's position, and our own

Versions of the regularity view	The subject matter	Contention	Hume's position	Our position in this book
The psychological version	How the mind comes to be made to expect the regular attendant of a given type of event.	The mind, exposed to a regular sequence of events, is induced by habit to expect the regular attendant of a given type of event through the process of psychological conditioning.	This is Hume's own regularity view, which is a key element of his attempt to explain the origin of the idea of causal necessitation.	This view relies on assumptions which can be rejected and is not directly relevant to our aims.
The semantic version; the regularity theory of the meaning of causal statements	The meaning of the word 'cause'; the meaning of causal statements.	The concept of causation is that of regular conjunction or succession. 'To cause' *means the same as* 'to be regularly followed by'.	Hume does not subscribe to this and accepts that 'to cause' means the same as 'to necessitate, to bring about'.	We do not subscribe to it, either. With Hume, we accept that 'to cause' means the same as 'to necessitate, to bring about'.
The ontological version; causal idealism	What goes on in the world when causing goes on.	All there is to causation in the world is a regular conjunction of events, the idea of causal necessitation being just an idea in our imagination.	Hume does not subscribe to this, nor does he subscribe to its pair-opposite, causal realism; he is a causal agnostic.	We do not subscribe to it either; we also find Hume's stance to be suggestive of the limitations of the realist–idealist dichotomy.

Our rejection, collectively, of all three of these versions of the so-called 'regularity view of causation' suggests, quite strongly, that the basic idea of the regularity view is suspect. On the basis of our investigations so far, we have discovered no grounds for believing that there is any kind of *intrinsic connection* between causation and regularities. Yet, while we have made considerable progress, we have not yet exhausted the topic of causation and its relationship to regularities. There is one important aspect of this relationship which we have not yet discussed: the fact that observed regularities are often cited *as evidence for* causal judgements. This is the topic we take up in Section III.

III. Regularities as evidence for causal relations

It is well known that 'correlation is not causation'. To put this in a different way: observing a correlation, or regularity, is not sufficient to establish a causal claim. The presence of a causal relation can be inferred from a correlation only if it can be shown that the correlation is non-spurious, and we will consider, in later chapters, how this might be done. Our focus, right now, is on a related, but different question: is evidence of a correlation *necessary* to establish a causal claim? If we were to accept either the semantic or ontological versions of the regularity view, then the answer would be 'yes'. If causal statements were just statements about regularities, or if causal relations in the world were nothing but regular conjunctions, then regularities would constitute *necessary evidence* for causal claims—in either case, a causal statement could only be backed up by evidence about empirically observable regularities.

We have rejected both these versions of the regularity theory of causation. We have also shown that they should not be attributed to Hume. Some of his statements appear to suggest, nonetheless, that evidence of regularities is *required* to support a causal claim. In assessing these statements, it is important to note that Hume did not develop a full argument on this subject. The short section he wrote initially in the *Treatise* (1969: 223–5) on 'Rules by which to judge of causes and effects' finds no corresponding section, nor is there any further discussion of that subject, in his later work, *Enquiry* (1962). Moreover, though they often go unnoticed, Hume did in fact make some highly pertinent observations on this subject which appear to contradict the idea that evidence of regularities is required for causal claims. This section aims to lay the ground for the fuller analysis which will follow in Chapter 2 by drawing out the valuable, though much neglected, insights Hume offers on regularities as evidence for causal relations.

A likely reason for Hume's failure to develop a full analysis of this subject is easy to identify. His fundamental puzzle did not concern the empirical grounds for particular causal judgements, but rather the origins of our *idea* of causal necessitation. Moreover, Hume believed that the inference we make from a given type of event to its usual attendant does not involve any intellectual reasoning or argument (1962: 34, 39, 47); rather, it is based on our 'very extraordinary, and inexplicable' (1962: 108) instinct.

However, we believe Hume went too far when he negated the role of what he termed 'ratiocination' (1962: 39) in causal inquiry. Except in the case of some common-sense practical understandings—for example, that a flame causes the skin to burn (1962: 39)—causal inferences, especially of the sort drawn in a subject such as IR, require us to engage in evidence-based deliberation beyond what our instinct might (though most probably would not!) tell us (see Chapter 8).

Despite his relative reticence in commenting on the question of what counts as evidence for causal judgements, Hume does make the important observation, noted in Section I, that looking at what turns out to be a regular sequence of events *only once* is not enough to make one think of causation in that instance (1962: 63). This appears to suggest that *repeated* observations of the same sequence of events might provide a more significant clue that a causal relation is present. In the *Treatise*, Hume in fact made an even stronger claim: that one of the key rules by which we may know when two objects are related as cause and effect is that '[t]here *must be* a constant union betwixt' them (1969: 223; emphasis added). This appears to suggest that, for Hume, a regular conjunction between events of two types constitutes *necessary evidence* for the judgement that they are related as cause and effect.

There is, however, a significant difference between the claim (i) that observing the same sequence of events repeatedly might give an observer *a clear reason to suspect* the presence of a causal relation and the claim (ii) that observing the same sequence of events repeatedly constitutes *necessary evidence* for the presence of a causal relation. And, despite the passage we have just quoted, it is doubtful that Hume fully subscribed to the latter view. Not only did he refrain from repeating this claim in his later work, *Enquiry*, he also made a very significant qualification to this line of thinking, which should not go unnoticed. Quite astutely, he observed:

> 'Tis certain, that not only in philosophy [i.e., in science], but even in common life, we may attain the knowledge of a particular cause *merely by one experiment*, provided it be made with judgement, and after a careful removal of all foreign and superfluous circumstances. (1969: 154; emphasis added; see also 1962: 107)

This observation is extremely brief but strikingly insightful. In it, Hume is making two important points. First, he is indicating that an observed regularity between *A*-type and *B*-type events is not a *necessary* condition for making a justifiable claim that *A* causes *B*, for a justifiable claim can be made on the basis of a single well-conducted experiment. Second, he is explaining how it is that controlled experiments can yield evidence to support a causal claim: they can do so if 'all foreign and superfluous circumstances' (which are now often described as 'confounders') have been carefully removed. (We will revisit this issue in later chapters.)

Hume also recognized a further important point: that a causal statement, such as '*A* causes *B*', does not imply that '*A*-type events are *always* followed by *B*-type events'. In a noteworthy passage, he wrote:

> The vulgar, who take things according to their first appearance, attribute the uncertainty of events to such an uncertainty in the causes as makes the latter often fail of their usual influence; though they meet with no impediment in their operation. But philosophers, observing that, almost in every part of nature, there is contained a vast variety of springs and principles, which are hid, by reason of their minuteness or remoteness, find, that it is at least possible the contrariety of events may not proceed from any contingency in the cause, but from the secret operation of contrary causes. This possibility is converted into certainty by farther observation, when they remark that, upon an exact scrutiny, a contrariety of effects always betrays a contrariety of causes, and proceeds from their mutual opposition. A peasant can give no better reason for the stopping of any clock or watch than to say that it does not commonly go right: But an artist easily perceives that the same force in the spring or pendulum has always the same influence on the wheels; but fails of its usual effect, perhaps by reason of a grain of dust, which puts a stop to the whole movement. From the observation of several parallel instances, philosophers form a maxim that the connexion between all cases and effects is equally necessary, and that its seeming uncertainty in some instances proceeds from the secret opposition of contrary causes. (1962: 86–7; see also 1969: 182–3).

This is a crucial insight, and one which is closely interlinked with Hume's recognition of the power of controlled experiments: it is precisely because there is 'a vast variety' of causal processes hidden 'in every part of nature' that controlled experiments, which are carefully designed to remove 'all foreign and superfluous circumstances', can generate such persuasive evidence in support of a causal claim. However, Hume did not say that controlled experiments are 'necessary' either; his claim was rather that a single well-controlled experiment would be 'sufficient' to generate requisite evidence.

We agree with his judgement here. If it is possible to conduct a controlled experiment, this will often be desirable, but it is not *necessary* to conduct such an experiment in order to advance a causal claim with confidence and, indeed, it will often not be *possible* to conduct such experiments when studying world politics and other complex subject matters. This is a topic we will pursue further in Chapter 8.

Before we conclude, it is worth making one final observation about Hume's comments on regularities as evidence for causal judgements. Hume rightly acknowledged that if '*A* causes *B*', this does not imply that '*A*-type events are *always* followed by *B*-type events'. However, he seems to have supposed that if '*A* causes *B*', then '*A*-type events are *usually* followed by *B*-type events'. This is clear from the fact that, in his writings, he consistently characterized an effect of a cause as its '*usual* attendant' (see, e.g., 1962: 75; emphasis added). Such a view is problematic because, as we shall come to stress in later chapters, whether *B*-type events are a *usual* attendant of *A*-type events will depend on the nature of the conditions under which *A*-type events occur. For now, the key point is that whereas many of Hume's supposed intellectual descendants have assumed that in order to establish a causal relation between *A* and *B*, it is necessary to establish that A-type events are usually conjoined with *B*-type events, this assumption is mistaken.[17] Very much to Hume's credit, such a claim is flatly contradicted by his important observation that a single well-controlled experiment can be sufficient to generate compelling evidence in support of a causal inference.

Conclusion

The aim of this chapter has been to explore the relationship between causation and regularities and to remove various confusions that have sedimented around this subject. We have focused, in particular, on the so-called 'regularity view of causation', which is closely associated with Hume and which asserts, in broad terms, that there is an intrinsic connection of some kind between causation and regularities. In order to unpack this 'regularity view' and thereby to reveal the true relationship between causation and regularities it seemed natural to study what Hume, widely regarded as the progenitor of the 'regularity view', had written on this subject. A careful reading of Hume's texts has not disappointed us!

[17] Although not necessarily coming down directly from Hume, this line of thinking appears to have had a formative influence on IR in the mostly North American positivist tradition. See Rummel 1972: 32; Haas 1974: 59; Singer 1979: 181; see also Jackson 2016: 58–82.

Through our exposition of Hume's arguments, we have developed three important insights which we will carry forward into subsequent chapters. First, the so-called 'regularity view of causation' is not a unified doctrine and Hume did not in fact endorse the versions of this view with which he is commonly linked. In fact, our analysis of Hume's argument has given us the confidence to reject all the three forms of the 'regularity view of causation' associated with Hume. Second, our analysis has brought home the inadequacy of the dichotomous view, which predominates in IR and other disciplines, according to which there are just two ontological positions which it is possible to adopt on causation: realism and idealism. Our discovery that Hume appeared to remain agnostic about the reality of causal powers suggests the need to broaden the parameters of this debate. Third, we have noted that despite some occasional comments that appear to point in this direction, Hume did not consider regularities to constitute necessary evidence for causation, but rather recognized the potential power of controlled experiments. By way of concluding this chapter and indicating the direction we will take in subsequent chapters, we further expound these points below.

Against the 'regularity view of causation'

Although reference is often made to *the* 'regularity view of causation' (see, for example, Psillos 2002), it is a mistake to suppose that there must be one unified doctrine designated by that name. Rather, we have shown that 'the regularity view' is best understood as a blanket term, covering what we have called its 'psychological', 'semantic', and 'ontological' versions.

As shown in Table 1.2, these 'versions' have their own distinct subject matters with respect to which a relationship is said to be found between causation and regularity. Although Hume is often associated with the 'semantic' and/or 'ontological' versions, we have shown that he in fact advanced a 'psychological' thesis concerning the relationship between causation and regularities. Contrary to Hume, however, we do not find this 'psychological' version of the regularity view to be persuasive; in any case, the origins of our idea of causal necessitation, which constituted Hume's fundamental puzzle, are not relevant to this book's concerns. As with Hume, we do not subscribe to the 'semantic version' of the regularity view; both for Hume and for us, 'to cause' means 'to bring about', and not 'to be regularly followed by'. As for the 'ontological' version—which is 'the idealist view of causation', or 'causal idealism' by another name—we have shown that, common suppositions to the contrary notwithstanding, Hume did *not* subscribe to this position; neither do we.

In short, there is no such thing as *the* 'regularity view of causation', and we find no good reason to subscribe to *any* of its three versions identified here: 'psychological', 'semantic', or 'ontological'. If this casts considerable doubt on the idea that there is an intrinsic connection of some kind between causation and regularities, it does not mean that 'regularities' have *nothing* at all to do with 'causal relations'. As noted in Section III, an observed regularity may give us a clue or reason for suspecting that there is a causal relation underpinning it. More on that in a moment.

Significant implications of finding causal agnosticism in Hume

We do not subscribe to 'causal idealism': the thesis that all there is to causation in the world is the regular conjunction of certain kinds of events, causal necessitation being no more than an idea. As we have explained, we do not find this a coherent basis on which to engage in causal inquiry. If causal necessitation is no more than an idea, why would we engage in *causal* inquiry at all? However, contrary to what readers may gather from recent IR works on causation (see, for example, Wight 2006a; Kurki 2008), the incoherence of causal idealism does not mean that we must all accept 'causal realism' or be 'causal realists' at heart. This is because there is a third position—'causal agnosticism'—which we have interpreted Hume as favouring.

To appreciate the significant implications of locating agnosticism in Hume, especially in the context of IR, it is important to recall that recent IR literature on causation has been dominated by the causal realist–idealist dichotomy; and that, in this dichotomous frame of analysis, Hume has come to be treated as a figurehead of the causal–idealist camp.[18] Given this, the realization that Hume cannot be shown to have subscribed to causal idealism takes the wind out of the sails of those who have positioned themselves in the debates about causation by waving the banner of 'anti-Humeanism'. Moreover, the realization that Hume was a causal agnostic challenges the dichotomous frame of analysis itself by revealing the possibility of a third position.

This has an important implication that encourages us to follow a new avenue of exploration. In a dichotomous frame, it will be natural to ask which of the two metaphysical positions is philosophically the more persuasive. Indeed, in the recent IR literature on causation, much energy has been expended on discussing the metaphysical question of whether causal

[18] It is important to recognize that Hume has been treated in this way both by advocates of causal realism (e.g., Harré and Madden 1975; Patomäki and Wight 2000; Wight 2006a; Bhaskar 2008; Kurki 2008) *and* by advocates of causal idealism (e.g., Lebow 2014).

necessitation is really a feature of the world, independently of human thought. When the possibility of an agnostic position is acknowledged, the urgency to defend causal realism or idealism as philosophically the more creditable of the two doctrines recedes, for, according to causal agnosticism, it is not possible to *know* which of the other two positions is right. When the significance of this metaphysical debate is reduced, this frees up space for methodological questions about what is involved in conducting causal inquiry and how such inquiries should be carried out. In short, our finding that Hume was not a causal idealist, but rather an agnostic, encourages a move away from metaphysics towards methodology, a move we make in Part II of this book.

Regularities as evidence for causal relations

Finally, we have shown that Hume offers some quite nuanced and insightful ideas about the evidential bases of causal judgements. These go far beyond what the stereotypical labelling of Hume as the progenitor of the 'regularity view' tends to convey. Three points are worth briefly reiterating here.

First, observing the regularity with which *A*-type events are followed by *B*-type events is not a necessary condition for making a claim that '*A* causes *B*'. Despite some inconsistencies, we believe that Hume clearly recognized this important point.

Second, an observation of a regularity may, nevertheless, provide a valuable clue that there could be a causal relation underlying this observed regularity. We take this to be implicit in Hume's argument.

Third, in a controlled experiment, in which confounding factors can safely be assumed to have been blocked out, observing a single sequence of events can provide sufficient evidence for the presence of a causal relation. This is a particularly important observation which Hume made and which is discussed further by Roy Bhaskar, whose thought on causation we shall consider in the next chapter. Although we will raise significant questions about Bhaskar's defence of causal realism—which his 'critical realist' followers in IR regard as a way to enhance causal inquiries—we will find, in his writings, important insights concerning the *conduct* of causal inquiry, a topic which remained frustratingly underdeveloped in Hume's writings.

2

Bhaskar, Causal Realism, and the Conduct of Causal Inquiry

Having discussed Hume's thoughts on causation in some detail, we now turn to those of Bhaskar. He has attained something of a guru-like status in contemporary International Relations (IR) debates about causation, in which the so-called 'Humean regularity view of causation' is presented as being challenged by causal realism (see, inter alia, Patomäki 2002; Wight 2006a; Kurki 2008; Joseph and Wight 2010; Jackson 2011, 2016). In IR, this position—according to which causal powers exist as components of reality independently of human thought—has been advocated most notably by the so-called 'critical realists', led by Bhaskar.[1] In view of his widely acknowledged influence on recent IR debates about the nature of causation (see Jackson 2011, 2016), this chapter will take a close look at Bhaskar's defence of causal realism, developed in his seminal work, *A Realist Theory of Science* (2008), first published in 1975.

The key aim of Bhaskar-inspired critical realists in IR is to overcome what they see as some significant deficiencies in predominant approaches to causal inquiry in the discipline. In particular, they criticize what they see as an excessive focus on the identification of observable regularities as evidence for the presence of causal relations (see, for example, Kurki 2006, 2008); they point out that, under everyday conditions, causal powers will not usually give rise to observable regularities in the world around us, other than, as Bhaskar noted, 'rough-and-ready' ones (2008: 68). They advocate, as a solution, a rejection of what they characterize as 'Humeanism', which they equate with 'causal idealism' (a version of the regularity view of causation), and a turn to 'ontology' or metaphysics. From such a standpoint, critical realists contend that causal inquiries can be enhanced through a greater focus on questions of 'ontology'

[1] When Bhaskar entered the philosophical scene, Rom Harré was already advocating an 'anti-Humean' realist theory of science and causation (see Harré 1972; Harré and Madden 1975); the influence of Harré's philosophy on the study of society can be seen, for example, in Benton (1977: see especially 219, note 45) and Sayer (1992).

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0003

and, specifically, through an acceptance of the reality of causal powers—that is, the acceptance of causal realism as the correct ontological doctrine.[2]

In Chapter 1, we have provided some reasons for scepticism about this agenda, and especially about the juxtaposition of Bhaskar's causal realist view against Hume's view of causation. As we showed, Hume was not a causal idealist. So, Hume's view should not be thought of as the pair-opposite of Bhaskar's causal realism. Moreover, our analysis suggested that it is unhelpful to think in terms of a realist–idealist dichotomy; as Hume's own thinking suggests, other, more nuanced positions may be available. We nonetheless believe that it is important for those conducting and thinking about causal inquiry in IR to engage with Bhaskar's work, as we propose to do in this chapter. There are two reasons for this.

First, we consider Bhaskar's argument for causal realism to be flawed. Given the considerable influence his argument has exerted on debates about causation in IR, this needs to be spelled out. Indeed, our systematic critique of Bhaskar's argument for causal realism, which we develop below, provides a further reason for scepticism about the opposition between Humeanism and Bhaskar-inspired causal realism which has shaped so much of the recent philosophically informed discussion of causation in IR. If Bhaskar's defence of causal realism does not stand up to scrutiny, it is in no position to stand in opposition to *any* other views of causation, regardless of how one might interpret Hume's view. Second, although we consider Bhaskar's argument for causal realism to be flawed, our critical engagement with his work reveals a number of insightful ideas which we will carry forward into our analysis of the conduct of causal inquiry. Importantly, we will contend that these ideas are *methodologically significant* even if the *ontological connotations* that Bhaskar reads into them are, as we suggest, rather suspect. 'Beware of gurus', Hollis and Smith famously warned (1991b). We take heed of their advice in relation to the influence Bhaskar's realist philosophy has had in IR. Nevertheless, we consider Bhaskar to have made significant contributions to clarifying methodological aspects of causal inquiry, taking us considerably beyond where Hume had left us.

Our aim in this chapter is therefore twofold. First, we examine, and seek to expose some serious flaws in, Bhaskar's argument for causal realism. In doing so, we cast considerable doubt on claims that Bhaskar-inspired causal realism might constitute an improved basis on which to conduct causal inquiries in IR. Second, we highlight a series of conceptual and methodological insights

[2] Recall that, as we noted in Chapter 1, 'causal idealism', which *denies* the real existence of causal powers, is also properly an 'ontological' doctrine, concerning what can and *cannot* be said to exist.

which, we believe, can be extracted from Bhaskar's argument and which can, with appropriate modifications, be usefully applied to causal inquiries in IR, regardless of whether Bhaskar's (or any other) argument for causal realism is accepted. In doing so, we prepare the ground for the more detailed analysis of the logic of causal inquiry which we will provide in Part II of this book.

Bhaskar's argument for causal realism, which is a key element of his overall defence of philosophical realism, is dense and complicated. His writing is often repetitious and at times quite opaque; it also relies heavily on a conceptual vocabulary which many readers, not already acquainted with 'Bhaskar's world', will find unfamiliar. In very broad terms, however, we believe that it can be articulated as proceeding in three stages, whereby the claims made in the first two prepare the ground for his conclusion that causal 'powers of things' are real—that they are part of the world existing independently of human thought.

First, Bhaskar explores why scientists engage in the practice of controlled experiments and what this practice reveals about the nature of the world as it exists independently of human thought—or, to use the language of his 'transcendental' reasoning, what it *presupposes* about the world. He argues that what this practice presupposes, and reveals, is that the world must be what he terms an 'open system'—a system in which the regular conjunctions which might provide a valuable clue as to the presence of causal relations cannot be routinely observed.

This leads him to explore, second, what the outcomes of controlled experiments are evidence for. He argues that controlled experiments are intended to generate observable sequences of events which provide evidence for causal laws, and that while these laws are not themselves observable, they must, nevertheless, really exist, independently of human thought. Moreover, he claims, experiments are worth conducting precisely because the causal laws they reveal also hold outside of experimental conditions—that is, in open systems. In short, causal laws are not only real, they also embody what Bhaskar characterizes as 'tendencies', this term pointing, among other things, to what he calls the 'transfactual' quality of causal laws—that is, the fact that they hold in open systems as well as in the artificially closed systems created for the purposes of scientific experiments.

Third, and finally, Bhaskar asks whether these 'tendencies' are the most basic components of the world. He argues that they are not: they are generated by underlying 'powers of things'—that is, causal powers which are not directly observable, but which are nonetheless present in the world independently of human thought. This is the heart of his causal realism; his aim is to show that causal powers are a real feature of the world and to make the case

that the goal of causal inquiry should be to reveal and further investigate these causal powers and the tendencies to which they give rise.[3]

Our analysis of Bhaskar's argument for causal realism unfolds in three main sections, which correspond to the three stages of his overall argument. The first section deals with his analysis of the scientific practice of controlled experiments and his contention that the world must be an open system. In discussing this aspect of Bhaskar's argument, we find it striking that he, like Hume, highlights the utility of controlled experiments, conducted in closed systems, as a source of evidence for causal relations, and we endorse this. In addition, we consider Bhaskar's distinction between 'open' and 'closed' systems to be extremely valuable: as we will explain, we believe it methodologically prudent to assume that the environments studied in IR are, in most cases, 'open' with respect to whatever causal process is being investigated. However, we consider the realist inflection which Bhaskar brings to his argument to be misleading: we do not find Bhaskar's *ontological* claim about the real nature of the world convincing, nor do we believe that it is necessary for those conducting causal inquiries in IR to concern themselves with such issues.

The second section examines Bhaskar's view that controlled experiments provide empirical evidence for real 'causal laws' and that these laws hold not only under experimental conditions, but also in the world at large—in his parlance, that they exist 'transfactually', as 'tendencies'. Bhaskar's contention that 'causal laws' embody 'tendencies' (or what we prefer to call 'propensities') contains an important insight; we will argue, in Part II of this book, that causal theories are properly understood as propensity statements, and we will show that this has important implications for the logic of causal inquiry. As before, however, we will find Bhaskar's claim to have established an ontological thesis about the nature of the world as it exists independently of human thought to be unconvincing. We will also find certain other aspects of his argumentation unpersuasive.

The final section examines the heart of Bhaskar's realist ontology: his claim that causal laws, which exist as 'tendencies' in the world, must, in turn, be generated by underlying 'causal powers of things' and that these, too, can

[3] It is worth noting here that Bhaskar's causal realism, and his realist ontology more generally, extends to the social sphere. In the study of society, he, and those in the 'critical realist' movement he initiated, have stressed the importance of deep structural causes of social phenomena and the role of what they call 'explanatory critique'—an attempt to expose and supersede socially dominant explanations as often neglectful of structural causes and hence as ideologically loaded (see Bhaskar 1998; see also Archer et al. 1998: ix–xxiv; Patomäki 2002). In this chapter, however, our focus is on Bhaskar's argument for causal realism, and his realist ontology more broadly, based on his transcendental analysis of natural scientific practice. In Part II of this book, we will develop our own account of the logic of causal inquiry, an account which provides a tool for the critique of causal explanations in *any* field of study, natural or social.

be known to exist independently of human thought. Here, again, we find Bhaskar's argument unpersuasive. Although he paints a picture which may correspond to the way many people think, he goes too far in claiming to have derived secure ontological conclusions about the deep underlying reality of the world. As we also point out, this deficiency has important implications for the oft-heard claim in IR and elsewhere that the proper aim of causal inquiry is to uncover *underlying causal mechanisms*.

Our overall verdict on Bhaskar's thoughts on causation is hence that there are significant weaknesses in his argumentation, such that we remain sceptical of his ontological claims and his purported demonstration, in particular, of the reality of causal powers. What we find most useful in Bhaskar's work is not his defence of causal realism, but rather some of the key conceptual and methodological ideas to which he alerts us in the course of his argument. Key among these are (i) the importance of the distinction between open and closed systems, (ii) the importance of controlled experiments in generating evidence for causal claims, and (iii) the importance of recognizing that causal theories describe propensities. As our argument unfolds in later chapters, we will come to show why these ideas are important for the conduct of causal inquiries in IR.

I. Bhaskar on controlled experiments and 'open' and 'closed' systems

Bhaskar's argument for causal realism forms the key part of his overall realist ontology, which is based centrally on what he presents as a 'transcendental analysis' of scientific practice. This philosophical method comprises, first, analysing how natural scientists in fact set about discovering causal relations, and, second, investigating what this practice presupposes and reveals—or, more specifically, asking what must be the case with the world for this practice of scientific inquiry into the world to be intelligible and possible (2008: 8, 23, 108).[4] Following this approach, Bhaskar observes that, when they can, natural scientists use controlled experiments—a feature which, he claims, was not properly highlighted in the philosophy of science before him (1998: 9–11; 2008: 33). He argues that, for such a practice to be intelligible and possible, the world must be an 'open system'—a system in which a multitude of causal powers coexist and interact, such that scientists wishing to find out about one

[4] On Bhaskar's distinctive reliance on a *transcendental* argument for causal realism see, inter alia, Clarke 2010.

particular causal relation have a need to isolate it from interference—but one which is susceptible to spatio-temporally delimited regional closures, such as scientists devise for the purpose of isolating and studying particular causal relations of interest.[5]

In this section we will consider how Bhaskar's inquiry leads him to this judgement, which we shall designate as his thesis (A):

> (A): The world is, or exists as, an open system, which is susceptible to spatio-temporally delimited regional closures. This is why scientists engage in controlled experiments.

To expound a little further on the two contrasting terms Bhaskar employs, a 'closed' system is one in which, given a causal relation, '*A* causes *B*', events of type *A* are invariably followed by events of type *B*, whereas an 'open' system is one in which this invariable pattern is not observed (Bhaskar 2008: 33). In a 'closed' system, this regular pattern is observed because interference with the unfolding of a given causal process is prevented; in an 'open' system, such interference is *not* prevented, which is why an event of type *A* is *not* invariably followed by an event of type *B* (Bhaskar 2008: 46, 53, 98). Bhaskar introduces this important distinction primarily in order to make his case that the world exists as an 'open' system but that it contains the possibility of spatio-temporally delimited regional closures, which scientists exploit when they conduct controlled experiments designed to uncover causal relations (2008: 33, 46, 65, 66, 67, 103, 221).

Before we assess Bhaskar's argument in defence of (A), we should enter one important qualification. Bhaskar frequently repeats his assertion that the possibility of science presupposes that 'the world is an open system', though susceptible to regional closures (2008: 66, 78, 116, 126, 144). However, he occasionally refers also to the 'ubiquity of open systems in nature' (2008: 91, 104). There is no indication that he is drawing any distinction between 'the world' and 'nature', so we can assume that he is using the two terms interchangeably. There is also no indication that, when he characterizes the world, or nature, as 'open' he is especially conscious of his use of the singular form ('an open system') for the most part and a plural form ('open systems') at other times. However, his reference to 'the ubiquity of open systems in nature', in addition to the claim that 'the world' is a—single—open system, is noteworthy. This is because, whereas his argument, summarized in (A),

[5] Bhaskar's remarks which support this summary are scattered in many parts of his book in closely interlinked and partially overlapping ways. See Bhaskar (2008: 30, 33, 38, 46, 47, 53, 65, 66, 67, 78, 91, 98, 103, 104, 116, 126, 144, 221).

that 'the world' is an open system is a *philosophical* argument, based on an analysis of what is presupposed by scientific practice, his observation about what is 'ubiquitous' in nature is clearly an *empirical* assertion. It concerns the frequency of open systems encountered in nature, derived no doubt from Bhaskar's appreciation of the extent to which experiments are employed in studying it—though, again, there is no indication that he is conscious that he is making an empirical, rather than a philosophical, claim here. At this point, we will not press this distinction any further, but we will return to it later in this chapter. For now, our focus is on clarifying Bhaskar's philosophical argument in defence of (A) and assessing the quality of his reasoning.

Exposition

To understand how Bhaskar defends thesis (A), it is useful to consider how scientists themselves use controlled experiments to identify the laws governing phenomena that interest them. A classic illustration is found in the case of what Rom Harré (1981: 78–81) calls 'Galileo's experimental discovery of the form of a kinematic law'. According to the formula Galileo confirmed by carefully analysing his experimental outcomes, the distance traversed by an object freely descending towards the Earth is invariantly proportionate to the square of the time it takes to travel that distance (Harré 1981: 81).

To discover this law, Galileo is said to have set up a contraption in his study. He placed on a table a long piece of wooden beam with a gentle incline. On this beam was cut a narrow channel, lined with parchment, made as well-polished and smooth as possible. Along this groove, Galileo let a hard, smooth, and very round bronze ball roll down freely. This experimental setup was designed, among other things, to minimize friction between the descending object and the surface on which it moved.[6]

What this story tells us is this: to discover a law pertaining to a process of interest—in this case, the free fall of an object—scientists need to create conditions under which the process in question will unfold with as little interference as possible. It is from this basic understanding about scientific practice that Bhaskar develops his contention that the world is an 'open' system but susceptible to regional closures. How does he do this?

[6] See http://galileo.rice.edu/lib/student_work/experiment95/inclined_plane.html (accessed 28 Sep. 2023). See also 'Galileo's Measure of Gravity Explained by Jim Al-Khalili' (https://www.youtube.com/watch?v=ZBr8Q2ROX9s&t=10s) (accessed 28 Sep. 2023). Rather than observe an object fall freely from a given height directly to the ground, Galileo let a ball roll down a slope with a gentle incline to make it possible, with the technology available then, to measure the time it took for the ball to traverse a given distance.

Bhaskar points out that scientists conduct controlled experiments in order to generate observable sequences of events which constitute compelling evidence for causal laws that interest them. He contends, however, that this practice would have no intelligible rationale if investigators could readily observe such sequences simply by studying the world around them (2008: 33, 91). The very fact that scientists find it necessary to engage in controlled experiments reveals, for Bhaskar, that the world must be an open system; otherwise, the practice of controlled experiments would be redundant. Scientists, being rational, could not possibly engage in it, he seems to be thinking. Meanwhile, the world, as an open system, must be susceptible to regional closures; otherwise, controlled experiments could not possibly be conducted. Bhaskar therefore reasons that if the world were not an open system susceptible to regional closures, then science, centred on the practice of controlled experiments, would not be possible at all; and, he opines, since science *is* possible, inasmuch as we know it occurs (2008: 106, see also 29, 116, 220), we can conclude that the world is, or exists as, an open system with a possibility of regional closures.

Assessment

At first glance, Bhaskar's argument may appear persuasive, containing some quite smart moves. However, we are not persuaded. It is significant that Bhaskar draws attention to controlled experiments as the principal means by which natural scientists seek to discover causal relations; we endorse the thought that, if they can be conducted, such experiments are likely to be the most secure basis on which causal knowledge can be developed. We also consider the terminology of 'open' and 'closed' systems, and the idea that it is vital to distinguish between them to be a key ingredient of any discussion of the conduct of causal inquiry. That said, we believe that the superficial plausibility of Bhaskar's argument betrays some significant problems which it is important to understand, and which undermine Bhaskar's claim to have demonstrated that the world is or exists as an open system with the possibility of regional closures.

If the world were indeed as Bhaskar describes it to be, then it would be rational for scientists to resort to controlled experiments when seeking to discover causal relations. Similarly, if scientists *assumed* the world to be such (rather than *knew* it to be so), then it would also be rational for them to conduct controlled experiments. Crucially, however, the reverse does not hold. The fact that scientists continue to engage in the practice of controlled

experiments, thereby giving rise to science as we know it, does not imply that scientists know the world to be an open system, with the possibility of regional closures, or even that they suppose that it is. Such a precondition is not required by the practice of controlled experiments. Accordingly, we contend, Bhaskar's argument is at best inconclusive. Let us explain.

To begin with, while stressing the open-systemic nature of the world, Bhaskar is far from suggesting that the world is so chaotic that nothing is predictable. He acknowledges that 'many processes are *effectively* isolated and many systems *more or less* closed', and that 'rough-and-ready regularities are everywhere at hand' (2008: 68; emphasis added). This enables us to avoid an unhelpful supposition that 'openness' and 'closedness' are absolute categories, such that a given system must be either (100%) open or (100%) closed. On the contrary, in Bhaskar's own thinking, openness and closedness allow for gradations (compare Patomäki 2010, 2017).

It is also important to note the following remark that Bhaskar makes in passing (2008: 73): 'A closure is of course always relative to a particular set of events and a particular region of space and period of time'. We take Bhaskar to mean that it is only possible to speak of a given environment as 'closed', to whatever extent, *if we specify in advance the causal process of interest* in relation to which it is 'closed'.

We believe Bhaskar is right on this issue; it is only in relation to a given causal process that an environment under study can be said to be 'closed', to whatever extent. To elaborate further, an environment under study (a spatio-temporally delimited region) is 'closed', with respect to a given causal process, if other causal processes which have the potential to interfere with the full unfolding of the process of interest, or to bring about similar outcomes even in its absence, are excluded. Conversely, an environment under study is 'not closed', with respect to a given causal process, if any of these confounding factors *are* present in it. In short, what Bhaskar means by a 'closed system' is a more or less 'controlled environment' with respect to a given causal process; and what keeps interferences under some degree of control is, for the most part, imposed artificially, as in scientific experiments (2008: 33, 65, 221), or, under exceptional circumstances, present naturally, thereby providing an opportunity for so-called 'natural experiments' (see Bhaskar 2008: 33, 58, 65, 104, 261; see also Cartwright 2007: 31; Diamond and Robinson 2011; Dunning 2012; and discussions in Chapter 8).

So far, so good, it seems. There is one point that cannot go unnoticed, however: Bhaskar appears not to fully appreciate that whether, or to what extent, a given environment under study is deemed to be 'closed' with respect to a given causal process is *a matter of scientific judgement*, based on

relevant causal knowledge and informed causal guesses and, where available, supported by evidence derived from interpreting the results of experiments. Indeed, we cannot even form a preliminary judgement as to whether or how far an environment under study is 'closed' with respect to a given causal process without having some knowledge about how other relevant causal processes will work there; and we cannot be sure that our preliminary judgement was adequately correct unless we have empirical evidence to support it. This seems to have escaped Bhaskar's attention, perhaps because he was focused on what can be learned from the transcendental analysis of the conditions of the possibility of controlled experimentation and did not look closely enough into the methodological aspect of that scientific practice. Let us explain why this is significant.

Clearly, when scientists devise a controlled experiment, they do so with a particular causal process in mind. Based on their existing causal knowledge and informed guesses, they judge what other causal processes might also be operating in the environment under which the process of interest is to be studied and which of these might have the potential to interfere with the unfolding of this process. They then take appropriate measures to reduce anticipated interferences. To put this differently: they judge—or, if they do not do so *consciously*, they certainly act as if they are judging—that a given environment in which a causal process of interest unfolds is unlikely to be a closed system unless experimental controls are introduced, for certain interferences can reasonably be anticipated; hence, they seek to devise controls that will prevent these interferences, thereby allowing the causal process of interest to unfold unhindered.[7]

If this line of thinking is correct, then scientists must consider, *case by case*, whether a *given environment* in which they wish to study a causal process of interest is closed with respect to this process, but they need not make any universalized judgement at all about whether the world—the entire realm in principle subject to scientific inquiry—is open or closed. The practice of scientific experimentation, contrary to what Bhaskar seems to suggest, does not presuppose that 'the world' is an open system; what it presupposes is only that scientists judge that some environments with respect to which some causal processes are studied are not naturally closed.

At this point, we are ready to revisit the issue we noted earlier in this section: Bhaskar's view about the *ubiquity* of open systems in the world,

[7] Much trial and error is likely to be involved in devising an appropriately controlled test environment for the purposes of demonstrating the presence of a given causal process.

which he offers in addition to his main claim that the world is a—single—open system.

As we have already noted, Bhaskar's claim that open systems are ubiquitous in the world must be an empirical judgement: it amounts to a rough, quantitative assessment of how commonly open systems are to be found in the world. This realization is significant, for in defending this particular claim—that open systems are ubiquitous in the world—Bhaskar could not possibly employ his preferred philosophical method of reasoning. Clearly, no form of philosophical argumentation, even a sound one, could defend an empirical claim—and a rough quantitative assessment at that.

Could Bhaskar not argue, though, that the scientific practice of introducing experimental controls into a particular environment under study would only be intelligible and possible if that environment were really 'open'? Again, clearly not: for this scientific practice to be intelligible and possible, it is sufficient that the investigator *judges*, on the basis of their causal knowledge and informed causal guesses, that the environment under study is likely to be open with respect to their causal process of interest; it is not necessary that this environment is *in fact* an open system (let alone that the world as a whole is). As to whether the investigator is correct in their judgement that a given environment is open or closed with respect to a given causal process, it is important to note that this is a substantive scientific question; this kind of question cannot even be addressed, let alone resolved, by any kind of philosophical reasoning, including Bhaskar's transcendental approach.

These considerations lead us to the conclusion that Bhaskar's case for his thesis (A) is at best inconclusive. Whereas Bhaskar sought to establish an ontological thesis concerning the nature of the world, we do not find this aspect of his argument convincing. This indicates a pattern which will repeat itself in Sections II and III: it is Bhaskar's philosophical attempt to establish the real nature of the world as it exists independently of human thought that we find unconvincing.

Despite this significant reservation about Bhaskar's attempt to defend an ontological claim about the nature of the world existing independently of human thought, we do find some important insights in his arguments. We believe he is right to have drawn attention to the importance of controlled experiments in providing evidence for laws and causal relations. We also believe it *methodologically prudent* to assume, unless a persuasive case can be made to the contrary, that any given environment in which a causal process of interest might be studied is 'open' with respect to this process.

The need for such methodological cautiousness is particularly pertinent in a discipline such as IR. In any field of study, we can treat a sequence of events—one which conforms to the pattern paradigmatically associated with a causal process of interest—as persuasive evidence of a causal relation if and only if we can be sure that it has been observed under relevantly controlled conditions. By the same token, any constant or regular conjunction which may be observed, and which may make us *suspect* the presence of a corresponding causal relation, will not constitute persuasive evidence of such a relation unless we can make a good case that it was observed under relevantly controlled conditions. Many of the specific subject fields of IR (e.g., interstate war) are fields in which we might reasonably expect to find many causal processes interacting in an uncontrolled manner. It is therefore especially important, in the study of world politics, to remain vigilant about whether it is reasonable to assume that the specific environments in which causal processes of interest are thought to unfold are closed with respect to those processes.[8]

By way of concluding this section, Table 2.1 summarizes our findings so far, comprising Bhaskar's realist thesis about the nature of the world, our critique

Table 2.1 A summary of Section I: Bhaskar on controlled experiments and 'open' and 'closed' systems

Bhaskar's realist thesis	Our critique of his philosophical reasoning	Methodological takeaways from our engagement with Bhaskar's thinking
(A): The world is, or exists as, an open system, which is susceptible to spatio-temporally delimited regional closures. This is why scientists engage in controlled experiments.	Contrary to Bhaskar's claim, the scientific practice of controlled experiments does not presuppose the world to be an open system, though with the possibility of regional closures; his transcendental analysis of the practice of controlled experiments therefore fails to support his thesis (A).	However, Bhaskar's arguments highlight the importance of controlled experiments for generating evidence for causal laws and the prudence, in the absence of a persuasive case to the contrary, of assuming that an environment under study is 'open' with respect to any particular causal process of interest.

Bhaskar's question addressed by (A): Why do scientists engage in controlled experiments at all and what does this prevailing scientific practice presuppose and reveal about the nature of the world as it exists independently of human thought?

[8] Our methodological advice here is consonant with Bhaskar's 'ubiquity of open systems' view, insofar as this is understood as an empirical judgement concerning specific environments in which specific causal processes are being studied.

of his philosophical reasoning, and important methodological ideas we take from our engagement with Bhaskar's thinking.

II. Bhaskar's view of what the outcomes of controlled experiments are evidence for

The second part of Bhaskar's argument for causal realism focuses on what the outcomes of controlled experiments constitute evidence for. He makes two assertions in this regard which we believe are worth closer scrutiny. First, he claims that the outcomes of controlled experiments are sequences of events which provide evidence for what he terms 'causal laws', and that these laws and the sequences of events observed, while distinct, are both real, existing independently of human thought (2008: 12, 18, 33, 54).[9] We will designate this thesis as (B)(i) and examine it in the first subsection ('The outcomes of controlled experiments').

Second, Bhaskar argues that although causal laws are, for the most part, discovered by means of controlled experiments, it is a clear error to suppose that these laws exist and hold only in the controlled environments in which they are discovered. On the contrary, he claims, causal laws exist and hold not only in these closed environments, but also in the world at large—that is, in open systems (2008: 46). He therefore contends that causal laws embody what he terms 'tendencies', which hold 'transfactually'—that is, outside of the controlled environments devised to discover them, as well as inside (2008: 50, 97, 229, 252–3), and that these causal tendencies are part of the world existing independently of human thought. We will designate this thesis as (B)(ii) and examine it in the second subsection ('Causal laws').

Putting these two theses together, Bhaskar is arguing that the outcomes of controlled experiments are evidence for causal laws which are part of the real world, and that these laws embody tendencies which hold in open systems as well as in closed systems.

As in Section I, we will find Bhaskar's claim to have established an ontological thesis about the nature of the world as it exists independently of human thought unconvincing. We will also find some other aspects of his argumentation unpersuasive. Nevertheless, we believe Bhaskar is right to identify the outcomes of controlled experiments as evidence of causal laws; we will also

[9] What Bhaskar terms 'causal laws' appear broadly to correspond to what we will, in Chapter 4, term 'abstract' statements of causal relations—that is, statements of causal relations which have the paradigmatic form '*A* causes *B*'. Instead of repeating 'what Bhaskar terms "causal laws"' and the like, we will from now on refer simply to 'causal laws' (but *without* the inverted commas), when discussing Bhaskar's view on them.

endorse the idea that statements of causal laws (or relations) are statements of 'tendencies', or, as we would put it, 'propensity' statements. These are methodologically significant ideas which we take from Bhaskar and apply in our own account of causal inquiry in Part II of this book.[10]

The outcomes of controlled experiments as evidence for 'causal laws' which exist independently of human thought

In a controlled experiment, Bhaskar points out, the investigator tries to create an appropriately closed environment and triggers a causal process which, if it unfolds without interference, will generate a sequence of events having a 'one-to-one relationship' with the 'characteristic pattern' of events described in the statement of the causal law in question (2008: 46). In other words, the investigator tries to create an environment in which, when an *A*-type event is introduced, with all potentially confounding factors kept under control, a *B*-type event follows, this sequence of events conforming to the characteristic pattern described in the law '*A* causes *B*' and thereby constituting evidence for this law. For Bhaskar, therefore, the sequences of events which constitute the outcomes of successful controlled experiments are the observable manifestations of, and hence serve as empirical evidence for, causal laws. Importantly, however, he stresses that the causal laws in question and the experimentally generated sequences of events are *ontologically* distinct (see 2008: 12, 18, 33, 54, 65). First, we need to clarify what Bhaskar may mean by this. We will then present and evaluate the reasoning that supports this assertion.

[10] We should enter here one important observation on the expression, 'existing independently of human thought', which we have already used several times. Throughout this book, we employ this phrase interchangeably with 'really existing', to mean 'existing not just as an idea in our minds but as part of objective reality', 'existing mind-independently', 'existing regardless of what humans think, know, or do', and the like. However, as already noted briefly, Bhaskar and critical realists extend their realist philosophy to the study of society. This necessitates some modification to the equation of 'really existing' with 'existing independently of human thought'. In the study of society, social institutions, such as the institution of money, are often presented as 'real' or 'really existing' although, clearly, such institutions do not exist 'independently of human thought (or actions)'. Indeed, they are sustained by the thought and actions of those human beings who live inside them. When such institutions are characterized as 'real', what is meant is roughly that they confront individual members of the society concerned as a 'reality' that none of them, individually, can alter or wish away. When, therefore, they are said to 'really exist', the adverb 'really' is not equivalent to 'mind-independently' in the full sense. Nonetheless, it is interchangeable with the term 'independently of human thought' insofar as this is understood, *in this context*, to mean 'independently of thought or actions on the part of any particular human': any particular person thinking and acting in a manner that defies the institution in question would not undo it. Importantly, however, with respect to the central subject of this chapter—the reality or otherwise of *causal powers*—no such qualification is required; to say that they 'really exist' is to say that they do so 'as part of objective reality'.

Exposition

In claiming that experimental outcomes and causal laws are *ontologically* distinct, Bhaskar is not suggesting that they are different in the sense that whereas the former exist, the latter do not. His point is rather that they *both* exist—notwithstanding the fact that the sequences of events constituting experimental outcomes are observable whereas causal laws for which they provide evidence are not. This reflects a core contention of Bhaskar's realism: that what can be said to really exist is not narrowly confined to empirically observable objects. For him, observable objects only constitute a portion of the world of existing things (2008: 58). Therefore, when he says that the two things are *ontologically* distinct, what he seems to mean, in part, is that 'sequences of events produced as outcomes of controlled experiments' and 'causal laws' name different kinds of objects—the former observable, the latter not—but that, nevertheless, these objects both exist independently of human thought.[11]

Thus, Bhaskar's thesis (B)(i) can be stated as follows:

> (B)(i): The outcomes of controlled experiments are sequences of events which constitute evidence for causal laws; these and the sequences of events produced are distinct, but they both exist independently of human thought.[12]

How does Bhaskar demonstrate (B)(i)? His argument is extremely compressed. He wrote:

> It is a condition of the intelligibility of experimental activity that in an experiment the experimenter is a causal agent of a sequence of events but not of the causal law

[11] Bhaskar is thinking that they belong to different *levels* of the world of existing things (see, for example, 2008: 102), but we are inclined to interpret such language as a metaphor and hence set it to one side.

[12] Although we believe (B)(i) accurately captures Bhaskar's thought, we should enter one qualification: he seems to slip into thinking that it is *empirical regularities* (rather than *single sequences of events*) that constitute the outcomes of controlled experiments, and that causal laws are ontologically distinct from *them* (2008: 33, 256, 259); he aims this line of thinking directly against what he treats as the Humean (mis)conception of causal laws as mere empirical regularities (1998: 9; 2008: 14–5, 33, 104–5, 149 fn6). But this twist in Bhaskar's thinking, while playing a key role in his attack on Humeanism, need not detain us here. Suffice it to note that Hume did not have a conception of causal laws as mere empirical regularities and, further, that what scientists present as evidence for a causal law resulting from a controlled experiment is, as Hume himself noted in passing, *a single sequence of events* that conforms to the causally expected pattern: this should not be confused with a regular pattern identified in repeatedly observed sequences of events. It is a misunderstanding of scientific practice to imagine otherwise. True, as Bhaskar points out, regular patterns are, for the most part, only observable under artificially controlled conditions, and scientists create controlled conditions in the hope of generating evidence for causal laws. But it does not follow, and we must not get trapped into supposing, that what scientists observe as outcomes of controlled experiments and present as evidence for causal laws are themselves *empirical regularities*—a point that appears to have escaped Bhaskar's attention.

> which the sequence of events enables him to identify. This suggests that there is a [*sic*] ontological distinction between scientific laws and patterns of events. (2008: 12; see also 45)

Assessment

In our judgement, the above-quoted paragraph, in which Bhaskar sums up his argument for (B)(i), conceals a flaw in his reasoning. He seems right, on the one hand, to contend (a) that it is a condition of the intelligibility of experimental activity that in an experiment the experimenter is a causal agent of a sequence of events but not of the causal law which the sequence of events enables them to identify. However, we believe he is wrong, on the other hand, to claim that (a) suggests (b) that there is an 'ontological distinction', *in his intended sense as clarified above*, between scientific laws and patterns of events. Contrary to his belief, (a) does not suggest (b)—that 'scientific laws' and 'patterns of events', while being distinct, *both really exist*. Let us amplify.

To begin with, in order to demonstrate the existence of an object, it is necessary to present the grounds on which it is possible to rationally agree that it exists. Scientists usually do this by presenting some persuasive empirical evidence to support the idea that it exists; but, in the case of causal laws we lack such evidence, for causal laws are unobservable. Bhaskar accepts this: it is because causal laws, if they exist, are not observable, that he seeks to develop an argument, demonstrating that they really exist, through an analysis of prevailing scientific practice.

Bhaskar is right to believe that the practice of controlled experiments, aiming to provide evidence for causal laws, would be unintelligible if causal laws and the sequences of events cited as evidence for them were not thought of as distinct: 'X' and 'evidence for X' could not be one and the same thing! He may also be right to contend that the practice would be unintelligible if causal laws and the sequences of events providing evidence for them were not thought of as distinct *kinds of things*, and, in particular, if causal laws, like sequences of events, were also thought of as human-made—as produced, rather than discovered, by the experimenters. Contrary to what he seems to believe, however, this does not show that causal laws necessarily exist independently of human thought. This is a simple but possibly elusive point, and so let us elaborate a little further.

Pointing out that an object named by the term 'X' cannot be thought of as being human made is clearly not enough to show that such an object really

exists. For example, our idea of a dragon is that of a naturally occurring, not of a human-made, thing. But that is not enough to show that the objects named by the term 'dragon' really exist as naturally occurring objects. What we can say in this connection is only that dragons, *if they were to exist*, would be naturally occurring, not human made. By the same token, Bhaskar's contention that, unlike the sequences of events made observable as experimental outcomes, causal laws—or the objects named by the term 'causal laws'—could not be thought of as generated by the experimenters is not sufficient to show that they are components of the world existing independently of human thought. Bhaskar's argument, stated in the short paragraph quoted above, does not therefore adequately support (B)(i).

Now, of course, Bhaskar may retort that, for the scientific practice of controlled experiment to be intelligible, scientists must *assume* that causal laws, for which they are trying to generate evidence, really exist; why, otherwise, would they bother to seek empirical evidence for causal laws by devising controlled experiments, which is often a complex task (2008: 53)? To this, we can reply as follows.

There is no doubt that some individual scientists, in engaging in controlled experiments, may suppose that causal laws exist as pregiven elements of nature. Bhaskar himself seems to think this way: throughout his book, he equates 'causal laws' and 'laws of nature'—laws of nature being, on his supposition, built into the makeup of the world, dictating how nature ought to behave by pregiven natural necessity and incapable of being produced or altered by human agents (2008: 35, 97, 112–14, 221, 259).[13]

Significantly, however, Bhaskar does not spend any time considering whether science might perhaps be possible even without making this metaphysical leap and assuming causal laws to be pregiven elements of nature. In particular, he fails to consider the possibility that some scientists might appreciate the important role that knowledge of causal laws appears to play in human understanding and hence seek evidence for such laws, even while remaining agnostic about whether these laws *really* exist, independently of human thought.[14] We will explore the potential of this kind of agnosticism in more detail in Chapter 3.

[13] Bhaskar could not, of course, argue that *as laws of nature*, causal laws are pregiven elements of reality; that would be to assume the existence of causal laws in advance by a definitional fiat!

[14] One reason why Bhaskar has neglected this possibility may be that he has framed his argument for causal realism in part as a repudiation of the 'Humean' view of causation. In doing so, however, he has failed to recognize the important point we made in Chapter 1: that Hume is not a causal idealist, but rather an agnostic. Bhaskar has thereby failed to consider the possibility of an agnostic alternative to both causal realism and causal idealism.

For now, it is sufficient to have noted that, whereas Bhaskar seems right to say that the practice of controlled experiments assumes that the experimentally generated sequences of events and the causal laws in question are distinct kinds of objects, he does not explicitly defend the idea that, while distinct, they *both* exist independently of human thought. Instead, he simply assumes that the distinctiveness he points to is enough to show that they must both exist, though as different kinds of things.

We agree with Bhaskar that causal laws are different from the sequences of events which are generated in controlled experiments, and which constitute evidence for those laws. Bhaskar is also right to stress the importance of the outcomes of controlled experiments as evidence of causal laws. However, we find Bhaskar's demonstration of his thesis (B)(i) to be inadequate. We are not, of course, suggesting that causal laws (and causal relations) are just an idea; this much should be clear from our discussion in Chapter 1. Our point is rather that Bhaskar fails to establish a compelling case for the *ontological* thesis that causal laws really exist as a component of the world, independently of human thought. His argument does not establish for sure whether causal laws are real or whether scientists merely talk and act as if they are.

'Causal laws' exist as 'tendencies' in the world at large

As we have just noted, when Bhaskar claims that the sequences of events produced as outcomes of controlled experiments and the causal laws for which these sequences serve as evidence are 'ontologically distinct', he means, in part, that, though distinct, these two things both really exist. However, as the phrase 'in part' indicates, there is a second dimension to his argument that these two things are 'ontologically distinct'. This is the idea that the existence of causal laws, unlike that of experimentally produced sequences of events, is not confined to the controlled environments in which they are discovered, but that these laws 'endure and continue to operate' (2008: 13, 33, 46; see also 65, 221, 236) also in open systems in the world at large. We shall designate this Bhaskar's thesis (B)(ii), which can be stated as follows:

> B(ii): Causal laws hold not only inside the controlled environments in which they are discovered, but also in any open systems outside; causal laws therefore embody what we can aptly call 'tendencies'; and these tendencies exist as part of the world, independently of human thought.

Below, we will first outline Bhaskar's argument backing this thesis, and then examine his reasoning to see if he has defended it adequately.

Exposition

Bhaskar's argument for (B)(ii) is that it would be a clear error to think otherwise: to think that causal laws, the evidence for which is made observable by scientists under controlled conditions, hold only under those conditions, but not under any other conditions—that the causal laws discovered under controlled conditions do *not* hold in open systems in the world at large. If this were the case, he opines, the practice of controlled experiments would be unintelligible, having no satisfactory rationale. This is because, in his view, the point of conducting such experiments is not just to discover how things work inside scientists' laboratories. The point is rather to reveal the causal laws that hold in the world at large—that is, to reveal the causal laws that scientists take to hold in open systems, even though they recognize that, in these open systems, the same sequences of events which are observable under controlled conditions (and which, under controlled conditions, constitute evidence for causal laws) are not routinely observed (Bhaskar 2008: 13, 33).

In any case, Bhaskar contends, the idea that causal laws identified through controlled experiments only hold inside closed systems, and do not 'endure and continue to operate in their normal way' outside of them (2008: 33; see also 46), leaves unanswered the question of what, if anything, governs phenomena in open systems. It also fails to provide a rationale for the practical application, in open systems, of the knowledge obtained under experimental conditions—that is, 'the explanation, prediction, construction and diagnosis of the phenomena of ordinary life' (2008: 65, 92).[15]

The solution to all these problems, Bhaskar thinks, would be to accept that the laws identified in closed systems also 'persisted and operated' (2008: 65) in any environment outside; they exist, and govern phenomena, in open systems as well as in closed systems. He therefore argues that a causal law embodies what he terms a 'tendency', a word which conveys for him the idea that a causal law operates 'transfactually'—in open as well as in closed systems—while at the same time acknowledging that it only shows up as a sequence of events directly conforming to the law under controlled

[15] Regarding the question of what governs phenomena in open systems, he clearly believes that something, rather than nothing, does (2008: 65; 259–60).

conditions—that is, when interference is prevented.[16] In short, he argues that causal laws, which controlled experiments provide evidence for, are best conceptualized as embodying 'tendencies' existing as part of the world independently of human thought.

Assessment

We do not reject Bhaskar's thesis (B)(ii) in its entirety. We in fact broadly agree with him that the word 'tendency' captures what a statement of a causal law points to. But because the term 'tendency' is commonly used to mean 'frequency', we will, in subsequent chapters, use the term 'propensity' instead; we will argue that a statement of a causal law (or what we will characterize as an 'abstract' statement of a causal relation) is a 'propensity statement'. We use this terminology to highlight the 'normic' feature of a causal law or relation, namely that the relation '*A* causes *B*' will manifest as an observable pattern 'if an *A*-type event, then a *B*-type event' only under the right conditions—that is, only when interference with the full unfolding of this causal relation is prevented. We will discuss this further in Chapter 4. Having said that, while we endorse this aspect of Bhaskar's argument, we are sceptical of his claim to have established an ontological proposition. We have two significant objections to his argument, and one substantial reservation about it.

First, as we argued with respect to (B)(i), Bhaskar has not successfully demonstrated that causal laws really exist as components of the world independently of human thought. Adding an observation, as a key component of (B)(ii), that causal laws are aptly characterized as embodying 'tendencies' does nothing to demonstrate that such tendencies exist as part of the world independently of human thought.

Second, Bhaskar's contention that the whole point of finding evidence for causal laws under controlled conditions must be to identify those laws which hold transfactually is insufficient to justify his realist interpretation of causal tendencies. Clarifying the rationale or aim of scientists conducting controlled experiments is not enough to demonstrate that any causal laws which scientists manage to identify are *in fact* transfactual. Some individual scientists may of course think and act as if they suppose that causal laws for which they produce empirical evidence embody 'tendencies' that really exist transfactually. However, the same criticism as we raised in connection with

[16] Wight and Joseph (2010: 11) use the term 'transcontextual' in place of Bhaskar's term 'transfactual': this conveys Bhaskar's intended sense more clearly.

(B)(i) applies here: to render his argument effective, Bhaskar would need to demonstrate that it is impossible to provide a coherent account of science which does not rely upon a metaphysical supposition that 'tendencies' are real, but he does not do this. In particular, he fails to rule out the possibility that scientists might talk and act as if the causal laws they discover probably hold transfactually and yet be content to remain agnostic about whether the 'tendencies' which these laws appear to embody really are part of the world existing independently of human thought.

Third, Bhaskar seems right to suggest that routine scientific practice indicates that scientists typically suppose there to be some degree of transfactuality in the knowledge obtained through controlled experiments. However, because scientific practice is quite limited and imprecise with respect to open systems (see Cartwright 1999), the scope of this supposition is not as clear as Bhaskar appears to think. Some scientists may indeed suppose that exactly the same laws as found in closed systems all endure and operate in open systems. But it is not entirely unreasonable to hold that science has yet to uncover the laws that operate outside controlled environments, at least if we mean by this that science has yet fully to discover *how* laws identified through controlled experiments manifest outside the controlled conditions. Let us give a brief illustration.

Suppose that two items are dropped from the same height outdoors: one an aerodynamically designed heavy object and the other a banknote, which is highly non-aerodynamic. Whereas the first object falls in a straight line, accelerating quickly downwards, the banknote does not fall in the same way: it takes much longer to reach the ground and does not fall in a straight line, landing some distance away from where the first object landed. Concerning such an episode, scientists can explain the fact that the banknote did not fall in a straight line, that it took longer to reach the ground than the aerodynamically designed heavy object, and that it landed some distance away from where the heavier object landed, by pointing to the effects of air resistance and turbulence caused by the wind, both of which can be demonstrated, experimentally, to interfere with the effects of gravity, especially in the case of relatively light and non-aerodynamic objects. But what they cannot do is explain why the banknote landed exactly when or where it did in relation to the other object.

The problem scientists face is that although controlled experiments may furnish them with secure knowledge of the laws of gravity, and of many other relevant laws, what controlled experiments cannot reveal is how exactly these laws combine with one another in open systems, where their interaction is also affected by the presence of other causal processes which may interfere

with this interaction. In the circumstances, what they do is to indicate some of the causal processes which they presume are in play, and whose presence seems to make observed outcomes intelligible, but they fail, generally, to produce precise descriptions of the interaction of these causal processes—in this case, to explain exactly when and where the banknote fell (see Clarke 2010: 301–2; Cartwright 1999: esp., 27). In short, science has only a limited grasp of *how* different laws combine under different circumstances, and this fact is obscured somewhat by Bhaskar's rather strident assertion concerning the transfactuality of causal laws.

Despite these critical remarks regarding Bhaskar's argument in relation to (B)(ii), we believe, as we acknowledged earlier, that his claim that causal laws should be understood as embodying 'tendencies' is quite insightful. It draws attention to a crucial feature of causal *statements*: they describe processes which are not thought to be restricted solely to closed environments, but which nonetheless unfold *fully* only when interference is prevented, as it is in controlled experiments. This is an insight which we will draw on when we come to present our own account of causal inquiry in Part II of this book. However, when we do so, as we have also explained, we will use the term 'propensity', rather than 'tendency'. We will argue that an abstract statement of a causal relation, paradigmatically '*A* causes *B*', is properly understood as a propensity statement; we will argue, moreover, that this insight is crucial for the logic of causal inquiry.

It is important to note that in developing Bhaskar's insight, we will only be analysing what is asserted in *statements* of causal relations, rather than, as he seeks to do, commenting on the nature of causal relations themselves. We remain unpersuaded that Bhaskar has successfully demonstrated his realist contention that 'causal laws', and hence also the 'tendencies' they are considered to embody, are really components of the world existing independently of human thought. In any case, we will go on to show, in Chapter 3, that a realist commitment of this kind is not necessary as a basis for causal inquiry.[17]

By way of concluding this section, Table 2.2 summarizes its key findings.

[17] Bhaskar analyses statements of causal laws as statements of tendencies and also argues for the mind-independent existence of causal laws as tendencies in the world. In his Postscript to the Second Edition, however, he acknowledges that he should have been clearer in distinguishing these two aspects of his work but insists that it should be apparent to the reader that his focus is on the existence in the world of tendencies referred to in causal statements (2008: 251–2). By contrast, our focus will be on the content of causal statements; we are content to remain agnostic about whether laws and tendencies are really part of the world existing independently of human thought. We will explore these issues further in subsequent chapters.

Table 2.2 A summary of Section II: Bhaskar's view of what the outcomes of controlled experiments are evidence for

Bhaskar's realist thesis	Our critique of his philosophical reasoning	Methodological takeaways from our engagement with Bhaskar's thinking
(B)(i): The outcomes of controlled experiments are sequences of events which constitute evidence for causal laws; these and the sequences of events produced are distinct, but they both exist independently of human thought.	Bhaskar's analysis of the scientific practice of controlled experiments is inadequate to demonstrate that causal laws exist independently of human thought.	Bhaskar is right to stress the importance of the outcomes of controlled experiments as evidence of causal laws.
(B)(ii): Causal laws hold not only inside the controlled environments in which they are discovered, but also in any open systems outside; causal laws therefore embody what we can aptly call 'tendencies'; and these tendencies exist as part of the world, independently of human thought.	Bhaskar's contention that causal laws embody 'tendencies' does nothing to demonstrate that such tendencies really exist. Although he is right that scientists typically suppose, and act is if, causal laws hold in open systems as well as in the closed systems in which they are discovered, he does not sufficiently acknowledge the limitations of scientific knowledge regarding open-system phenomena.	We extract from our engagement with Bhaskar's thinking the idea that *statements* of causal laws are *propensity* statements; this will play a key role in our analysis of the logic of causal inquiry in Part II of this book.

Bhaskar's question addressed by (B)(i) and (B)(ii): Given that it is the open-systemic nature of the world that makes it necessary for scientists to conduct controlled experiments, what do the outcomes of controlled experiments constitute evidence for?

III. Bhaskar's argument for the reality of 'causal powers'

The final stage of Bhaskar's argument for causal realism consists in a contention that the 'tendencies' which are embodied in the 'causal laws' for which controlled experiments provide evidence are not the most basic elements of the world; rather, these 'tendencies' are generated by the 'causal powers of things'—real causal powers which exist independently of human thought,

and which underpin all causal relations. We shall designate this as Bhaskar's thesis (C), which states:

> (C): For there to be causal laws, existing as 'tendencies' in the world, there must be 'things with causal powers' in the world; these 'causal powers' are not just an idea, as causal idealists believe, but really exist in the world.

Exposition

In order to defend this thesis, Bhaskar engages in an analysis of the concept of tendency and argues that tendencies, which he considers causal laws to embody, cannot be thought of as the most basic element of the world. He reasons that for there to be a tendency, there must be a causal agent, a 'real something', endowed with a power (or possessing a liability) (2008: 49).[18] He considers it to be the aim of science, by combining theoretical and experimental work, to identify what kinds of things the world contains and the 'ways of acting of these independently existing and active things' (2008: 24)—that is, to identify what powers (or liabilities) things in the world possess by their nature.[19]

Bhaskar's argument about the relationship between 'power' and 'tendency' involves two moves. First, he contends that these are distinct items. Most humans have the *power* to steal but do not have a *tendency* to steal; kleptomaniacs do have this tendency—they are actively and continuously looking out for an opportunity to steal and will do so unless stopped in some way. Bhaskar considers that this idea of a given attribute of a thing being in play and thereby continuing to be active is captured well by the concept of 'tendency', in contrast to that of 'power' as he differentiates them in his thinking.[20]

[18] By a 'liability', Bhaskar (2008: 87, n. 32) means 'what Hobbes called a "passive power"': a propensity to suffer an effect caused.

[19] Bhaskar at various points in his argument equates 'ways of acting' with 'laws', 'powers', and 'mechanisms' (2008: 51, 52, 109, 184, 202, 219) and, accordingly, he expresses his interpretation of the aim of science in various interrelated, but not identical, ways. Notwithstanding his quite frequent reformulations and qualifications, however, Bhaskar's view adheres fundamentally to a consistent, realist line: that the aim of science is to identify how the world really is and what it contains (beyond what it is possible to observe directly), including causal powers.

[20] Bhaskar uses the term 'tendency' primarily to express the 'transfactual' and 'normic' characters of causal laws. However, he adds that it also points to the nature of causal laws as 'enduring orientations' (2008: 229), continuing to be active or 'in play' (2008: 50). It is this last feature, Bhaskar is saying, that distinguishes 'tendencies' from 'powers'. He remarks: 'In the concept of tendency, the concept of power is … literally dynamized or set in motion' (2008: 50). However, such an idea of a potential of a given object 'waiting' or 'pushing' for its fuller realization is an animistic metaphor and does not form part of the notion of 'propensity' employed in this book.

Second, Bhaskar argues that in order for X to have a tendency to do Ø—such that X will do Ø in the absence of intervening (or countervailing) causes—X must have the power (or liability) to do (or suffer) Ø in the first place. To use the same example again, for a kleptomaniac to have the tendency or propensity to steal, they must, qua humans, have a capacity, or power, to steal. It is in this sense that Bhaskar considers powers to be more basic than tendencies: he argues that powers underpin tendencies. Hence, he contends that the powers of things are more basic elements of the world than its laws (2008: 98–9, 229–31).[21]

Assessment

In arguing for his ontological theses, culminating in (C), Bhaskar is trying to present a plausible depiction of the 'real' world, as it exists independently of human thought. He, of course, is a philosopher and not, for instance, a visual artist, and therefore, rather than shapes and colours, he uses words, expressing concepts, as a means by which to depict this 'real' world. But readers may have noted that the concepts he uses are everyday ones, such as 'power' and 'tendency', and his method of arriving at what he presents as a plausible depiction of the 'real' world is based on investigating the *logical relationship between these everyday concepts*. He is arguing that, *according to our everyday thinking*, 'powers' are logically prior to 'tendencies' and that *therefore* it must be the case that the 'powers' of things, not their 'tendencies', are the more basic elements of the 'real' world.

This may sound good, but is it really so? Why should we believe that it is possible to arrive at knowledge of the 'real' world, existing independently of human thought, by analysing the everyday concepts with which we make sense of the phenomena that the world reveals to us? What warrant is there that we are thereby revealing the nature of the 'real' world as it is, independently of human thought about it, rather than simply analysing our everyday

[21] Bhaskar does not necessarily believe 'powers of things' to be the *ultimate* elements of the word. He wrote: 'A power-statement then acts as an imperative for scientists to find, and as a temporary place-holder for, that explanation which, by capturing *the essence of the thing*, will allow the most stringent possible criterion for our knowledge' (2008: 238; emphasis added; see also 177). However, according to Bhaskar, '[t]he class of "things" [that science tries to capture] is far wider than that of "material objects": it includes *fluids, gases, electronic structures, fields of potentials, genetic codes, etc.*' (2008: 98; emphasis added); they 'may be *powers, forces, fields or just complex structures or sets of relationships*' (2008: 226; emphasis added). But if 'things' can also be 'powers', it hardly makes sense to speak of 'powers of things'. At this point in his argument, Bhaskar does not appear to be engaged in a transcendental analysis of scientific practice, but is rather sketching scientific thinking concerning what the world comprises. It is sufficient for our purpose here to have established that, in philosophical terms, Bhaskar's conclusion is that 'causal powers of things' are more basic than 'causal laws' in the structure of the world.

conceptual scheme? We see none. One of the reasons why Bhaskar's realist view may appear convincing is that it coheres with the way we ordinarily tend to think, and find it hard not to think, about the world—partly because such a way of thinking is firmly embedded in our everyday language. But the way we tend to think and talk about the world may or may not yield a plausible, let alone necessary, conception of the world independent of our knowledge.

We share Bhaskar's scepticism towards old-style metaphysics but, contrary to his proclamation that metaphysics should be superseded by 'conceptual science' (2008: 36), we find it difficult to accept that analysis of everyday concepts can help answer inescapably metaphysical questions. We therefore have serious reservations about the claim that Bhaskar's thesis (C) is grounded in sound, or even relevant, philosophical reasoning.

In expressing these reservations, we are not claiming that Bhaskar is wrong in his conclusion; we are not denying that 'tendencies' *might* be rooted in 'causal powers'. Rather, we are accepting the broad thrust of Hume's observation that there is no empirical evidence which could confirm the existence of causal powers underpinning causal laws; and we are also expressing scepticism about the idea that Bhaskar has managed to get round this problem through a purely logical argument about our everyday concepts.

Before we bring this chapter to a close, there is one subject that needs to be touched on in connection with Bhaskar's claim that 'causal powers of things' underpin causal laws, because this has been taken to have an important implication for the practice of causal explanation in IR and elsewhere. To understand this, it is necessary to note that Bhaskar found a close association between 'powers of things' and 'mechanisms'. Indeed, he at times equates the two. He maintains, for instance, that mechanisms 'are *nothing but* the powers of things' (2008: 221; emphasis added) and 'exist *as* powers of things' (2008: 50, 186, 219; emphasis added). However, since, according to Bhaskar, 'powers of things' are ontologically more basic than laws and the tendencies these embody, it follows that 'mechanisms' have an ontologically more basic existence than laws. It is consonant with this way of thinking that 'mechanisms' *ground* 'laws', as Bhaskar also asserts (2008: 46, 95).[22]

[22] Confusingly, however, Bhaskar also characterizes *both* 'mechanisms' and 'laws' as the 'ways of acting of things' (2008, 51, 52, 109, 184). This suggests that the two things—mechanisms and laws—are the same; but if they are the same, one cannot be characterized as 'grounding' the other. To add to this confusion, Bhaskar even states that 'mechanisms' are 'nothing other than the powers *or* ways of acting of things' (2008, 202, 219; emphasis added), thereby reconfirming the equivalence of 'mechanisms' and 'powers of things' but, at the same time, apparently equating 'powers of things' with 'ways of acting of things'; however, he has used the latter term to talk of both 'mechanisms' and 'laws'. It would be difficult to sort out this muddle, but, even if we were to succeed, it would remain unclear how the analysis of the logical relationships between these concepts could help reveal the supposedly deep underlying structure of mind-independent reality.

This point is especially noteworthy because there are some widely read IR methods texts which, referencing Bhaskar's work to support their claim, insist that causal explanations of events are only adequate if they uncover *underlying* generative or causal mechanisms—that is, mechanisms found at a deeper level of reality (see Harré 1972; Benton 1977; Dessler 1991; George and Bennett 2005: 135–6; see also Wight 2006a, 32; Bennett 2013).[23]

As we have shown, however, Bhaskar has not provided a philosophically adequate defence of the claim that 'powers of things', and hence also 'mechanisms' (according to Bhaskar's equation of the two), lie at a deeper level of reality. Indeed, we doubt that such a defence can be provided. This undermines the claim that an adequate causal explanation must identify *underlying* mechanisms. Indeed, we do not think it makes sense to say that 'mechanisms' exist *at a deeper level of reality* at all, at least if 'deeper' is understood in metaphysical terms, rather than as a metaphor. Rather, 'mechanisms' are what we describe in response to an explanation request when we are asked to give *a*

Table 2.3 A summary of Section III: Bhaskar's argument for the reality of 'causal powers'

Bhaskar's realist thesis	Our critique of his philosophical reasoning	Methodological takeaways from our engagement with Bhaskar's thinking
(C): For there to be causal laws, existing as 'tendencies' in the world, there must be 'things with causal powers' in the world; these 'causal powers' are not just an idea, as causal idealists believe, but really exist in the world.	Bhaskar claims to have identified the deep structure of mind-independent reality by analysing the logical relationship between our everyday concepts of 'tendencies' and 'powers', but there is no warrant to the notion that this type of conceptual analysis yields a relevant, let alone a correct, answer to his question; at best, this merely reiterates what may perhaps be a conventional image of the world.	The view that causal mechanisms lie at a deeper level of reality is often used to lend philosophical support to the notion, occasionally encountered in IR and other subjects, that an adequate causal explanation of an event of interest must identify causal mechanisms lying at a deeper level of reality; but our critique of Bhaskar's realist thesis (C) removes a supposed philosophical underpinning from this received wisdom.

Bhaskar's question addressed by (C): Given that the outcomes of controlled experiments are evidence for causal laws which exist as 'tendencies' in the world, are these the most basic components of the world?

[23] It must be noted, however, that Bhaskar's point is that what he calls 'mechanisms' ground 'causal laws', not that the former grounds empirically observed 'regular conjunctions of events', which he was determined to distinguish from 'causal laws'.

more detailed causal account than has thus far been provided. (This point will be developed further in Chapter 6.) By way of concluding this section, Table 2.3 summarizes its key findings.

Conclusion

As we explained at the beginning of this chapter, we have aimed to do two things: on the one hand, to examine, and expose serious flaws in, Bhaskar's argument in support of causal realism, and, on the other, to extract from what he wrote some conceptual and methodological insights which may be relevant to the conduct of causal inquiry in IR.

We found that Bhaskar's argument for causal realism builds stage by stage, at each of which he defends a realist position, as summarized below in a progressive series of questions and theses presented in response:

Question A:	Why do scientists engage in controlled experiments at all, and what does this prevailing scientific practice presuppose and reveal about the nature of the world as it exists independently of human thought?
Thesis (A):	The world is, or exists as, an open system, which is susceptible to spatio-temporally delimited regional closures. This is why scientists engage in controlled experiments.
Question B:	Given that it is the open-systemic nature of the world that makes it necessary for scientists to conduct controlled experiments, what do the outcomes of controlled experiments constitute evidence for?
Thesis (B)(i):	The outcomes of controlled experiments are sequences of events which constitute evidence for causal laws; these and the sequences of events produced are distinct, but they both exist independently of human thought.
Thesis (B)(ii):	Causal laws hold not only inside the appropriately closed environments in which they manifest fully in observable sequences of events, but also in any open systems outside; causal laws therefore embody what we can aptly call 'tendencies'; and these tendencies exist as part of the world, independently of human thought.
Question C:	Given that the outcomes of controlled experiments are evidence for causal laws which exist as 'tendencies' in the world, are these the most basic components of the world?
Thesis (C):	For there to be causal laws, existing as 'tendencies' in the world, there must be 'things with causal powers' in the world; these 'causal powers' are not just an idea, as causal idealists believe, but really exist in the world.

We believe our very detailed critical exegesis of Bhaskar's argument in these three stages has cast serious doubt on his defence of causal realism. Of course, causal realism might still be defended by means of a different argument. However, those exploring causation in IR will often witness the formula 'Hume: Bhaskar = causal idealism: causal realism' in recent debates. Our analysis, in this chapter and in Chapter 1, has demonstrated that this formula cannot be taken seriously. This is because Bhaskar's defence of causal realism does not stand up to detailed critical scrutiny and Hume did not espouse causal idealism. The claim, advanced by critical realists in IR, that Bhaskar's causal realism triumphs over the 'Humean regularity view of causation' needs to be discarded, as does the dichotomous perspective on causation which underpins this. In addition to causal realism and causal idealism, it is important to allow for other positions, such as the causal agnosticism which, on our reading, Hume espoused.

That said, we find that reading Bhaskar in conjunction with Hume is fruitful. As will be recalled, a close examination of Hume's texts in Chapter 1 provided us with some important insights on the relationship between causation and regularities and, especially, on the evidentiary basis of causal claims. According to him, to provide evidence for a causal relation, '*A* causes *B*', it is not necessary *repeatedly* to observe a sequence of events whereby an *A*-type event is followed by a *B*-type event; it is sufficient to observe the sequence once, in a single, well-controlled, experiment. However, Hume's interest in developing what we termed the 'psychological version' of the regularity view of causation stood in the way of him developing this promising idea further. This is where Bhaskar makes a significant contribution. Far from being 'anti-Humean', what Bhaskar does, in effect, is to develop and extend Hume's embryonic insight on the evidentiary basis of causal claims by explicitly incorporating important concepts and conceptual distinctions, such as the contrast between open and closed systems.

Bhaskar has confirmed that a controlled experiment, showing *in a single sequence* that an *A*-type event was followed by a *B*-type event in an environment where all other possible causes of the occurrence of *B*-type events are likely to have been excluded, provides key evidence for a causal relation, '*A* causes *B*'. Moreover, although observing a regular association between events of these two types, outside of a controlled environment, may give us an impetus to inquire whether there may be some causal connection between them, it does not constitute either necessary or sufficient evidence for a causal claim.[24]

[24] This indicates why so much causal inquiry in IR relies on statistical controls, as, for example, in multivariate regression analyses. Where it is not possible to conduct experiments, showing that our causal

Interestingly, however, and in parallel with the way Hume's focus on developing a psychological theory diverted him from expanding his astute remarks about the evidentiary significance of controlled experiments, Bhaskar's focus on developing a philosophical ontology of causation got in the way of him presenting a more comprehensive discussion on the methodology of causal inquiry. Nevertheless, through a constructive reading of Bhaskar's text, we were able to extract some valuable insights on this subject, which will inform our own analysis of causal inquiry in Part II of this book.

What we have extracted and adapted from Bhaskar include a number of closely interrelated ideas and principles: (i) the importance of considering whether the environment in which a causal process of interest operates is an open or a closed system; (ii) the importance of controlled experiments in providing evidence for laws of science and assertions of causal relations; and (iii) the importance of recognizing that causal statements describe propensities which unfold fully only when interference is prevented.

It is important to stress that, unlike Bhaskar, we are not attaching any realist connotation to these ideas. In particular, when we speak of open and closed systems, it is not our intention to convey an ontological point that 'the world is' an open system; for us, the significance of the contrast between open and closed systems is methodological. We emphasize the importance of considering whether the specific environment in which a causal process of interest is to be studied is likely to be open or closed with respect to that process. Given that closed systems occur principally as a result of scientists' deliberate efforts to introduce experimental controls, we have endorsed the methodological principle that it should be the investigators' default assumption that the environments they study are open systems—unless, that is, a persuasive case can be made to the contrary.

When we speak of 'propensities' in the context of discussing laws of science and causal relations, we are not suggesting that propensities, or what Bhaskar terms 'tendencies', really exist, independently of human thought; as will become apparent, we prefer to remain agnostic on that question. Rather, we will be drawing attention to the importance of the idea that *statements* of causal relations are 'propensity statements'.

Before we broach these issues more fully in Part II of this book, there is one further issue to consider: is there a viable alternative to causal realism?

claims have taken into account possible interference by confounders must be achieved in some other way; using statistical, in place of experimental, controls is one such method (see, for example, King, Keohane, and Verba 1994; Brady and Collier 2010). In recent times, however, this approach has been largely superseded by the design-based approach (see Dunning 2010, 2012).

In a noteworthy Postscript to the second (1978) edition of *A Realist Theory of Science*, Bhaskar hinted at the possibility that a different philosophical view of science might come to challenge his realist position:

> [I]t should perhaps be stressed that I have not demonstrated that transcendental realism [which is the name he gives to his realist philosophy of science] is the only possible theory of science consistent with these [scientific] activities; only that it is the only theory *at present* known to us that is consistent with them. (2008: 260; emphasis original)

Bhaskar's formulation here is characteristically self-confident—overly so, in the light of what we have exposed in this chapter—but he was right in sensing that the realist philosophy of science and causal realism might come to be challenged by a different theory of science and view of causation. We will consider this in Chapter 3 in the light of subsequent developments in the philosophy of science, instigated by the work of Bas van Fraassen.

3

Beyond Causal Realism and Idealism

Van Fraassen and the Possibility of Agnosticism

Although we are unconvinced by Bhaskar's argument for causal realism, our analysis in Chapter 2 furnished us with some important conceptual and methodological ideas which we will carry forward into our own analysis of causal inquiry in Part II of this book. Before we do that, we will wrap up our engagement with existing philosophically informed debates about causation in IR by returning to a question we floated in Chapter 1, but left hanging at that point: is there an alternative to embracing either causal idealism or causal realism and, if so, does this alternative offer a viable basis on which to conduct causal inquiries?

We argued in Chapter 1 that causal idealism is not coherent as a basis for causal inquiry: if causation is known to be no more than an idea, then why conduct *causal* inquiries at all? By contrast, we accept that causal realism *does* provide a coherent basis on which to conduct causal inquiries. However, our analysis in Chapter 2 cast serious doubt on the positive argument *for* causal realism that is most prominently cited in IR: Bhaskar's attempt to show that causal realism is presupposed by scientific practice and hence must be true. Of course, even if we have succeeded in showing that Bhaskar's argument is flawed, this does not constitute a compelling argument against causal realism per se, for we have only rejected one possible argument for it.[1] Yet, if causal idealism does not provide a coherent basis on which to conduct causal inquiries, and if the principal argument in favour of causal realism is flawed, then it is worth considering what alternative positions may be available. Our analysis of Hume's arguments about causation, in Chapter 1, pointed towards a likely candidate, viz. causal agnosticism: a position which neither asserts nor denies that causation is real, but rather emphasizes that its reality cannot be known and hence refrains from making a judgement. This chapter

[1] That said, we find it hard to imagine how it would be possible to develop a convincing argument for causal realism, given that it advances a claim about the nature of the world as it is independently of human thought; it is hard to see how human thought could offer any purchase on what the world is like independently of such thought.

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0004

will explore the viability of agnosticism as a basis on which to conduct causal inquiries, asking whether it hinders causal inquiry in a way that realism does not.

This chapter has two main aims. First, we will flesh out the substance of an agnostic position. We will do this by exploring the work of Bas van Fraassen. He is a largely neglected figure in IR (see Humphreys 2019), yet he is well known in the philosophy of science as the key progenitor of contemporary efforts to articulate an alternative to scientific and causal realism (see Ladyman 2000: 837). Our exploration of Van Fraassen's arguments will therefore not only elucidate an agnostic alternative to causal realism, but also reveal how discussions in IR connect to wider debates in the philosophy of science.

Second, we will ask in what way, if any, the choice between causal realism and agnosticism might matter for causal inquiry. We will argue that adopting an agnostic rather than a realist stance does not imply any change in how causal inquiries should be conducted or in what substantive conclusions can be reached: there is no theory that a realist can accept which an agnostic cannot also accept, and there is no substantive causal explanation that a realist can offer which an agnostic cannot also offer. We will also spell out how it is possible to continue meaningfully to use causal language and to engage in causal inquiries even while refraining from judgement about whether causation really is a part of the world existing independently of human thought. We will contend, in short, that agnosticism provides just as good a basis for causal inquiries in IR as realism does.

Following a brief introduction to Van Fraassen's intellectual project in Section I of this chapter, we will present our discussion of agnosticism in two main parts: Section II will explore Van Fraassen's defence of agnosticism about the reality of the unobservable entities referred to in many scientific theories; Section III will explore his defence of agnosticism about the reality of causation itself. In each of these sections we will proceed, first, by laying out Van Fraassen's argument and, second, by asking whether agnosticism hinders scientific and causal inquiry in a way that would be alleviated by an embrace of realism. We will argue that the choice between realism and agnosticism makes no practical difference for causal and scientific inquiry, and hence that it is not even necessary for those conducting causal inquiries in IR to make an explicit choice between them.

We will conclude the chapter by explaining why it is nevertheless important to have identified agnosticism as a viable alternative to realism. The leading arguments for causal realism in IR privilege metaphysics over methodology. By focusing narrowly on the question of whether those conducting causal inquiries accept the reality of causation (and of unobservable entities

such as social structures), these arguments treat methodological issues as a second-order concern. We will argue that this is mistaken. Precisely because the choice between realism and agnosticism makes no difference for causal inquiry, it can, without detriment, be set to one side, facilitating a more sustained focus on the conduct of causal inquiry itself. That is the topic to which we will turn in Part II of this book, drawing on the conceptual and methodological ideas we developed over the course of these first three chapters.

I. Van Fraassen's intellectual project

Van Fraassen describes himself as an empiricist, but he points out that debates between empiricists and their critics have evolved considerably over the centuries. Consequently, he argues, empiricism should not be thought of as a 'single, specific philosophical position': the most we can do is to identify a 'tradition of philosophical positions that we call empiricist' (2002: 31). What lies at the heart of empiricism, he contends, is not a specific proposition, for example, that 'experience is the sole legitimate source of information about the world' (1985: 258), but rather a broad disdain for metaphysics, allied to a 'positive attitude toward science', understood as a form of inquiry which subjects all factual beliefs to the test of empirical evidence (see 2002: 63).[2]

What Van Fraassen terms 'constructive empiricism' is his attempt to articulate an empiricist alternative to scientific and causal realism.[3] These positions endorse metaphysical commitments that an empiricist will wish to avoid: belief in the reality of the unobservable entities referred to in leading scientific theories and belief in the reality of causation itself. However, Van Fraassen contends that it is possible to make 'sense of science, and of scientific activity … without inflationary metaphysics' (1980: 73)—that is, without making what he considers to be unnecessary metaphysical commitments. He seeks to demonstrate that an empiricist can legitimately 'withhold belief in anything that goes beyond the actual, observable phenomena' (1980: 203) by remaining agnostic about it. He therefore endorses agnosticism about the

[2] Note that characterizations of 'empiricism' in IR typically *do* associate it with a proposition of just the kind Van Fraassen rejects (see, for example, Smith 1996: 16; Wight 2002: 43, n19).

[3] First elaborated in *The Scientific Image* (1980), 'constructive empiricism' was defended and further developed over subsequent decades (see especially Van Fraassen 1985, 1989, 1993, 1994, 2001; see also Churchland and Hooker 1985; Monton 2007; Monton and Mohler 2021). Van Fraassen explains that he describes his empiricism as 'constructive' in order to emphasize that, on his account, science consists in the 'construction of models' that are 'adequate to' observable phenomena and not 'discovery of truth' about what the world is really like independently of human thought (1980: 5).

reality both of causation and of unobservable entities. The relevance of this project to our central concern in this chapter—whether agnosticism can provide a viable alternative to causal realism as a basis on which to conduct causal inquiry—should, we hope, be obvious.

As we saw in Chapter 2, Bhaskar contends that natural science is inherently realist: what scientists discover, through their controlled experiments, are tendencies which also hold outside of experimental conditions and which are rooted in real but unobservable causal powers of things. An important lesson he draws is that causal realism is also required for a successful *social science*—that is, a social science capable of revealing how real but unobservable social structures can shape not only political behaviours, but also human understanding of what is politically possible (see Bhaskar 1998). This line of thinking is largely accepted by causal realists in IR: they contend that satisfactory causal inquiries in IR must take seriously the potential influence of unobservable entities such as social structures, and that this requires belief in the reality of those entities and of their causal powers. Importantly, this means that causal realists in IR advocate not only 'causal realism'—realism about causal powers—but also 'scientific realism'—that is, realism about the unobservable entities referred to in scientific theories (see, for example, Wendt 1987, 1999; Patomäki 2002; Joseph 2007; Wight 2007; Kurki 2008; Joseph and Wight 2010; Bennett 2013).

Van Fraassen does not engage directly with Bhaskar, whose work has proved more influential in the social sciences than in mainstream philosophy of science. Moreover, he confines his analysis to natural science; unlike Bhaskar, he does not derive lessons for the social sciences. There is, nonetheless, an important similarity in their approaches, for Van Fraassen, like Bhaskar, investigates the practice of natural science and asks what it presupposes. His conclusions are, however, very different from Bhaskar's. Van Fraassen argues that the practice of natural science requires belief *only* in the reality of what is empirically observable; he *denies* that science requires belief in the reality either of unobservable entities or of causation. Yet, it is important to note that he does not endorse causal idealism. As we pointed out in Chapter 1 in relation to Hume's views on causation, to insist that causation is *only* an idea, and that it is therefore *not* part of the world existing independently of human thought, would be to advance a metaphysical claim. This is precisely what Van Fraassen seeks to avoid. The position he defends is an agnostic one, according to which the reality of unobservable entities or of causation cannot be authoritatively asserted or denied.

Before we get into the details of Van Fraassen's arguments, it is worth making two brief preliminary observations about how they relate to existing debates in IR.

First, Van Fraassen's arguments are directed principally against what he terms 'scientific realism', but there is an important ambiguity in how this term is used. So far, we have used it quite narrowly, distinguishing 'scientific realism' from 'causal realism', but often it is used as an umbrella term, encompassing 'causal realism'. Let us briefly elaborate.

To be a 'realist' is to believe 'in the reality of something—an existence that does not depend on minds, human or otherwise' (Chakravartty 2007: 8). Heikki Patomäki and Colin Wight therefore suggest that *some* kind of realism is a 'condition of possibility for science', for science assumes that *something* exists to be investigated; the question, they contend, is hence 'not whether one should be a realist', but rather 'what kind' of realist one should be (2000: 218). From this perspective, Van Fraassen is an *empirical* (or *common-sense*) realist: he is committed to the reality of empirically observable things but seeks to avoid firm beliefs about unobservable things (see Joseph 2007: 346). In contrast, *scientific* realism involves 'realism about whatever is described by our best scientific theories' (Chakravartty 2017 §1.2), including putative unobservable parts of the world. However, scientific realists differ among themselves about *which* unobservable parts of the world they take to be real: these might include unobservable entities, causation, laws, modalities, structures, and more (see Chernoff 2002: 191–2; Chakravartty 2017: §1). Hence, one way of thinking about 'causal realism' is that it is a *kind* of 'scientific realism'.

In developing his 'constructive empiricism', Van Fraassen is especially concerned to resist the idea that science requires belief in the reality of the unobservable entities referred to in many accepted scientific theories. However, it is important to appreciate that his arguments are directed against 'scientific realism' understood in a broad sense: his defence of an agnostic alternative to realism about unobservable entities sets the scene for a broader opposition to all forms of scientific realism, including realism about causation.[4] In what follows, we will consider how Van Fraassen defends agnosticism about the reality of unobservable entities *and* about the reality of causation. However, in order to make the exposition easier to follow, we will use the term 'scientific realism' in a slightly different way from

[4] As Monton and Mohler note (2021), Van Fraassen is also opposed to realism about the laws of nature (see Van Fraassen 1989), objective modality (see Van Fraassen 1980: 196–203), and more. These aspects of his work are not directly relevant to our concerns in this book; we will therefore set them to one side.

Van Fraassen: rather than employing it as an umbrella term, we will reserve it solely to denote realism about unobservable entities and use the term 'causal realism' to denote realism about causation.[5] This is summarized in Table 3.1.

Our second preliminary observation is that while Van Fraassen is opposed to all forms of scientific realism, he acknowledges the success of scientific realist arguments against previous versions of anti-realism, including phenomenalism, conventionalism, fictionalism, logical empiricism, and logical positivism. As he puts it: 'no one can adhere to any of these philosophical positions to any large extent' (1980: 2; see also 1980: 41; 2002: xviii). He explicitly seeks to articulate a new form of anti-realism which avoids the mistakes of these earlier positions (see Chakravartty 2007: 11–12).

This is important because Van Fraassen's updating of empiricism is not seriously taken into account, let alone refuted, by scientific and causal realists in IR, who tend to associate 'empiricism' with 'logical positivism'.[6] Bhaskar's argument for causal realism, which we examined in Chapter 2, was developed in the 1970s, and Bhaskar explicitly contrasted his position with what he described variously as 'positivism', 'empiricism', and 'orthodox philosophy

Table 3.1 Usages of the term 'scientific realism'

'Scientific realism' as an umbrella term	**Our usage in this chapter**	
	'Scientific realism'	**'Causal realism'**
Realism about whatever is described by our best scientific theories: this may include unobservable entities, causation, laws, modalities, structures, and more	Realism specifically about unobservable *entities* referred to in scientific theories	Realism specifically about causation

[5] There is a further nuance to Van Fraassen's use of the term 'scientific realism'. It is common to describe a scientist who believes in the reality of the entities (laws, etc.) referred to in our best theories as a 'scientific realist'. However, Van Fraassen (1994: 182) considers it more precise to describe such a scientist as a 'gnostic'—one who knows or believes—and to contrast them with an 'agnostic' scientist—one who withholds belief on such matters. This enables him to reserve the term 'scientific realism' as a label for the *philosophical* position that endorses belief in the reality of the entities (laws, etc.) referred to in our best theories as *proper to science*. We will adopt a more conventional approach, using the term '(scientific) realist' *both* to refer to the philosopher who endorses belief in the reality of entities (laws, etc.) referred to in our best theories as proper to science *and* to the scientist who in fact believes in their reality.

[6] For example, Wight and Joseph (2010: 4) acknowledge 'Bass [*sic*] Van Fraassen's "constructive empiricism"' to be 'the only serious contender' to scientific realism in philosophy of science, but dismiss it, without providing any justification for this stance, as 'a lost cause'. There is no further discussion of Van Fraassen or constructive empiricism in their whole book (Joseph and Wight 2010).

of science' (2008: 12, 14, 15)—that is, with the orthodoxy of mid-twentieth century, logical positivist philosophy of science. However, he admitted that this was an 'orthodoxy' to which 'few, if any, modern philosophers' (and this was in the 1970s!) would unambiguously subscribe (2008: 26). This exemplifies a broader pattern in which bold claims about 'the difference that realism makes' in the social sciences (see, for example, Shapiro and Wendt 1992) contrast realism with outdated forms of anti-realism.[7] Thus, when advocates of causal and scientific realism in IR depict a binary choice between their preferred position and a supposedly 'Humean' causal idealism tied to 'empiricism' and 'positivism' (see, for example, Patomäki and Wight 2000), they are ignoring just the kind of alternative view that Van Fraassen defends.

II. Agnosticism about the reality of unobservable entities

The central component of Van Fraassen's constructive empiricism is a defence of agnosticism about the reality of the unobservable entities postulated in scientific theories. This is an important topic because, as Van Fraassen readily acknowledges, many of the most widely accepted theories in the natural sciences do in fact refer to unobservable entities such as electrons and quarks. It is also highly relevant to IR, for some of the most widely invoked theories in IR refer to unobservable entities, such as the structure of the international system (see Waltz 1979).[8]

The challenge Van Fraassen faces is to articulate how a theory that refers to unobservable entities can be accepted without this requiring belief in the reality of such entities. Before we consider how he meets this challenge, it is worth reiterating that what is at stake here is the possibility of agnosticism about the reality of unobservable theoretical entities and not (yet) about the reality of causation. This reflects the structure of Van Fraassen's own argumentation: in *The Scientific Image* (1980), he begins by asking how acceptance of a theory that refers to unobservable entities can be reconciled with agnosticism about the reality of such entities; having done this, he then goes on to consider a range of other issues, including how, from an empiricist perspective,

[7] It is also indicative of a problem with how the term 'empiricism' is used in IR: when it is defined at all, it is typically identified as the epistemological element of a methodological 'positivism' that is itself derived from 'logical positivism' (see, for example, Smith 2000: 383). Guilhot (2008) and Jackson (2011: 44–59) discuss some of the tensions implicit in how these ideas have been imported into IR.

[8] Note that 'unobservable' is not the same as 'unobserved'. As Fred Chernoff explains, unobservable theoretical entities 'are not just things we have not yet observed, like a postulated bird species that we think must be the link between two known species'; rather, they are entities, postulated by scientific theories, 'that cannot, even in principle, be observed' (2007a: 110).

to make sense of the fact that scientists often use causal language to describe the world. Our analysis will follow this broad structure. In this section, we consider how Van Fraassen's account of theory acceptance in science opens the door to agnosticism about the reality of unobservable theoretical entities; in the next section, we will assess his account of how scientists use causal language and his defence of agnosticism about the reality of causation.

Van Fraassen's account of theory acceptance in science

Van Fraassen characterizes the core of scientific realism as follows: '*Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true*' (1980: 8; emphasis original). There are two elements to this. First, Van Fraassen attributes to scientific realists a view about the aim of science—about 'what counts as success' in science (1980: 8): it aims, through its theories, to offer a true account of what the world is like.[9] Second, he attributes to scientific realists a further view about what accepting a scientific theory involves: it involves believing that this theory is true in what it says about what the world is like. The key implication of these two views, which may not be immediately obvious, is as follows: providing a literally true account of what the world is like involves describing 'what there really is' in the world existing independently of human thought (1980: 7); many of our best scientific theories describe not only observable parts of the world, but also unobservable parts; if accepting such theories involves believing them to be true in what they say about what the world is like, then it involves believing that the unobservable entities to which they refer really are components of the world existing independently of human thought.

Van Fraassen's purpose in developing this account of scientific realism is to clarify what a constructive empiricist who wishes to avoid metaphysical commitments concerning the reality of unobservable entities will need to reject: they will need to reject *both* the idea that the aim of science is to provide a literally true story of what the world is like *and also* the idea that accepting a scientific theory involves believing it to be true (see 1980: 9).

In order to flesh out how an anti-realist might do this, Van Fraassen starts with the idea that science aims to provide a literally true account of what the world is like. He observes that there are two aspects to this idea: that

[9] Note that Van Fraassen is *not* attributing to scientific realists the naïve belief 'that today's theories are correct'; this would make scientific realism subject to empirical refutation (1980: 7). Rather, he is attributing to scientific realists an understanding of what science aims to achieve.

the language employed 'is to be literally construed'; and that 'so construed, the account is true' (1980: 10). This indicates two ways in which an anti-realist might resist scientific realism: they can either (i) reject the idea that the statements in scientific theories should be construed literally or (ii) accept that these statements should be construed literally yet reject the idea that the aim of science is to provide a literally true account of what the world is like.

The first of these options is pursued by instrumentalists, who hold that theoretical terms which appear to refer to unobservable entities *have no literal meaning* and hence lack truth values: statements which involve such terms 'are not even capable of being true or false' (Chakravartty 2007: 11). Instrumentalists justify theories which contain such terms *not* on the grounds that these terms describe features of the world, but rather on the grounds that the theories are useful: they generate accurate predictions about observable phenomena (see Chakravartty 2017: §4.1). Probably the best-known defence of instrumentalism in the social sciences was provided by Milton Friedman. He argued that the aim of social scientific theories should be to generate valid and meaningful predictions about observable phenomena, that theories capable of generating such predictions will often contain assumptions which are unrealistic (that is, known not to be true), but that this does not give cause for concern: when evaluating a theory, the only thing that matters is whether it generates accurate and useful predictions. He added that when a theory appears to refer to an unobservable entity, such references need not be construed literally: the theory can be interpreted as asserting that observable phenomena behave *as if* the unobservable entity existed (Friedman 1994: 148–9, 152–8). Kenneth Waltz articulates a similar view in chapter 1 of *Theory of International Politics* (1979).[10]

We mention instrumentalism principally to emphasize that this option is *not* the one Van Fraassen takes.[11] He argues that 'the language of science should be literally construed' and hence that 'the apparent statements of science really are statements, *capable of* being true or false' (1980: 10; emphasis original). In other words, scientific statements mean what they say, even if they refer to unobservable entities: '[i]f a theory's statements include "There are electrons", then the theory says that there are electrons'—that entities called 'electrons' exist (1980: 11). Van Fraassen insists, however, that construing scientific theories literally need not lead to scientific realism. This is because construing theories literally relates not 'to the aim we pursue in

[10] Waltz's various statements are, however, inconsistent, permitting multiple interpretations of his implicit philosophy of science (see, inter alia, Onuf 2009; Waever 2009; Joseph 2010; Humphreys 2012).

[11] This has sometimes been misunderstood in IR (see, for example, Wendt 1987: 351; Jackson 2011: 80).

constructing theories, but only to the correct understanding of *what a theory says*' (1980: 11; emphasis original). Van Fraassen agrees with scientific realists that the statements of scientific theories should be construed literally, but he denies that the *aim* of science is to provide a literally true story about what the world is like. Rather, on his account, the aim of science is to provide theories that are 'empirically adequate' (1980: 12).

Van Fraassen explains that to characterize a theory as 'empirically adequate' is to assert that 'what it says about the observable things and events in this world, is true' (1980: 12). His contention that science aims to provide theories which are empirically adequate therefore implies that what counts as success in science is providing theories that are true in what they say about *observable* parts of the world.[12] His point is to deny the scientific realist contention that what counts as success in science is developing theories that are, *in addition to being empirically adequate, also true in what they say about unobservable entities*. As a corollary of this view of the aims of science, Van Fraassen contends that accepting a theory does *not* require belief that it is true in *all* it says—in what it says about unobservables *as well as* about observables—but *only* belief that it is 'empirically adequate'—that it is true in what it says about observable parts of the world. On his account, one could coherently accept a theory that asserts that electrons exist, construe it literally, yet withhold judgement on the question of whether electrons are really part of the world existing independently of human thought.

Let us flesh this out by means of an analogy that Van Fraassen himself employs. He points out that the 'theist, the agnostic, and the atheist presumably agree with each other (though not with liberal theologians) in their understanding of the statement that God, or gods, or angels exist' (1980: 11).[13] However, they disagree about whether this statement should be believed: the theist believes this statement to be true, the atheist regards it as untrue, the agnostic withholds judgement about its truth, and the liberal theologian declines to recognize it as the kind of statement which is capable of being true or false.

In this analogy, the counterpart, in philosophy of science, to the theist is the scientific realist who construes statements referring to unobservable entities literally and believes these statements to be true—believes that these entities really exist in the world independently of human thought. The counterpart to the atheist would be some kind of philosophical idealist. However,

[12] Clearly, it would be inconsistent with the aims of science to seek to develop theories that were *not* true in what they say about observable things and events; empirical adequacy, at least, is a requirement!

[13] By a 'liberal theologian', Van Fraassen appears to have in mind a theologian who denies that references to God, angels, and the like should be construed literally.

proponents of such a view are 'rarely encountered in contemporary philosophy of science' (Chakravartty 2017: §1.2) and, in any case, it appears inconsistent to *accept* theories that refer to unobservable entities, construe them literally, yet believe them to be false.[14] The counterpart to the liberal theologian is the instrumentalist who denies that statements referring to unobservable entities are capable of being true or false. The counterpart to the agnostic is the scientist who construes statements referring to unobservable entities literally yet withholds judgement about whether they are true. This agnostic position is the one that Van Fraassen defends as part of his constructive empiricism. 'After deciding that the language of science must be literally understood', he argues, 'we can still say that there is no need to believe good theories to be true, nor believe *ipso facto* that the entities they postulate are real' (1980: 11–12). These positions are summarized in Table 3.2.

Table 3.2 Positions in the philosophy of science and their theological analogues

Position in the philosophy of science	Attitude towards statements, in accepted scientific theories, that refer to unobservable entities	Theological analogue
Scientific realism	These statements are to be construed literally; accepting a theory which contains such statements involves believing them to be true	Theism
Constructive empiricism	These statements are to be construed literally; but accepting a theory which contains such statements does *not* involve believing them to be true; it is coherent to accept theories that contain such statements while withholding judgement about whether these statements are true	Agnosticism
Philosophical idealism	These statements are to be construed literally; they are known (or believed) to be false	Atheism
Instrumentalism	These statements should not be construed literally; they are not capable of being true or false	Liberal theology

[14] In relation to causation, the counterpart to 'atheism' is 'causal idealism'. As we explained in Chapter 1, the apparent inconsistency involved in conducting causal inquiries while believing (and claiming to know) that causation (understood as a relation of production) is *not* part of the world existing independently of human thought leads us to consider causal idealism incoherent as a basis on which to conduct causal inquiries.

As an alternative to scientific realism, Van Fraassen therefore proposes the following statement of constructive empiricism: '*Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate*' (1980: 12; emphasis original). He advances two key arguments in support of this view.

First, he notes that 'experience can give us information only about what is both observable and actual' (1985: 253): by definition, *empirical* evidence can speak only to a theory's empirical adequacy—to the truth of what it says about observables—and not to the truth of what it says about unobservables (see 1980: 71). Because claims about unobservables have no empirical content, there is no evidence that can support or refute such claims. Van Fraassen therefore characterizes the scientific realist claim that the aim of science is to provide theories that are true in *all* they say (and not *just* in what they say about observables), and that theory acceptance involves belief that theories are true in *all* they say (and not *just* in what they say about observables), as 'empty strutting and posturing' (1985: 255).

Second, Van Fraassen points out that accepting a scientific theory 'involves not only belief but a certain commitment': because 'we are never confronted with a complete theory', when a scientist accepts a theory, they commit themselves to 'a certain sort of research programme'. This commitment involves, among other things, confronting 'any future phenomena by means of the conceptual resources of this theory' and drawing on it when seeking to provide explanations (1980: 12). Hence, regardless of how much they believe of what is asserted in the theories they accept, scientists will 'immerse' themselves in the 'world-picture' provided by these theories, using their 'full theoretical language without qualm' (1980: 81), including terms which refer to unobservables.

What Van Fraassen is drawing attention to here is that the 'amount of belief involved' (1980: 13) in accepting a scientific theory cannot be inferred from the way in which scientists talk about the world; the fact that practising scientists routinely refer to unobservable entities reveals that they are immersed in the world-picture provided by their theories, but it does *not* demonstrate that accepting these theories must involve belief that such entities are real. As Van Fraassen puts it: immersion in a 'theoretical world-picture does not preclude 'bracketing' its ontological implications' (1980: 81); one can accept and thereby commit to working with a scientific theory while remaining agnostic about the reality of whatever unobservable entities it may refer to.

As this discussion has been quite technical and abstract, let us give a brief example. In 2012, physicists at CERN 'discovered' the Higgs boson, almost

fifty years after its existence was first postulated.[15] Suppose that, some years prior to this, a scientist had accepted the standard model of particle physics, in which the Higgs boson was postulated. It seems clearly plausible to suppose that, at that stage, they could have accepted this model as the best available, believed it to be empirically adequate (believed it to be true in what it says about observable parts of the world), yet remained agnostic about the existence of the Higgs boson.

Let us go one step further, though. Following the success of the experiments conducted at CERN, could a scientist *still* accept the standard model and yet remain agnostic about the reality of the Higgs boson? Van Fraassen thinks so. He points out that 'detection' by means of an experiment is not equivalent to 'observation': 'detecting' an unobservable entity, such as the Higgs boson, by observing empirical evidence that matches what a theory which refers to this entity says should be observable, is not the same as (directly) 'observing' this entity (1980: 16–17).[16] On Van Fraassen's account, therefore, one can applaud the experiments conducted at CERN for successfully generating new evidence of the empirical adequacy of the standard model (by corroborating previously uncorroborated predictions) and yet remain agnostic on the further question of whether the Higgs boson 'really exists' in the world independently of human thought.[17]

Some readers may wish to interject at this point that many scientists *do* believe the theories they accept are true in what they say about unobservable entities, including the Higgs boson. Yet, even if this is true, it is irrelevant to an evaluation of Van Fraassen's argument (see Van Fraassen 1980: 8; Rosen 1994; Van Fraassen 1994; Ladyman 2000: 839). Van Fraassen seeks to provide an account of science which shows how it is *possible* (i) to accept our best scientific theories, construed literally, yet (ii) to remain agnostic about the existence of unobservable entities. He does *not* contend that belief in the truth of what our best scientific theories say about unobservable entities is irrational (see 1985: 252; 1994: 182), and he certainly does not contend that it is incorrect—for how could he know? His claim is rather that such belief is not required by science: it 'is supererogatory as far as science is concerned' (Van Fraassen 2007: 343; see also 1985: 255). The question of what individual scientists believe can therefore be set to one side. Although he himself prefers to remain 'agnostic about the existence of the unobservable aspects

[15] See https://home.cern/science/physics/higgs-boson/how (accessed 19 Sep. 2023).

[16] For a fuller account of the distinction between 'observation' and 'detection' and the role it plays in constructive empiricism, see Van Fraassen 1985: 252–8; 2001; see also Churchland 1985: 36–41; Hacking 1985; Musgrave 1985: 204–9; Jackson 2011: 85–7.

[17] This possibility appears to have been overlooked by Wight (2013: 340).

of the world described by science' (1980: 72), Van Fraassen does not expect all scientists to make the same choice. Constructive empiricism claims that agnosticism is consistent with the aims of science, not that all scientists are or should be agnostics (see 1994: 182). It is not refuted just because some scientists *do* believe that some unobservable entities referred to in accepted theories are real.[18]

Although this qualification is an important one, it may lead some readers to wonder what all the fuss is about: why have we taken the time to lay out in such detail an argument that appears to carry no implications for scientific practice? As will become apparent, it is precisely this lack of any obvious methodological corollary to the choice between realism and agnosticism that we consider significant. We pointed out earlier that advocates of causal realism in IR tend also to advocate realism about unobservable entities; in each case, they think in binary terms. They argue that, in relation to the reality of unobservable entities, one can be either a scientific realist or an instrumentalist, and that, in relation to causation, one can either be a causal realist or a causal idealist: in each case, there is no other possibility.[19] It is in light of the prevalence of this kind of binary thinking that the power of Van Fraassen's argument emerges: if he is successful in defending the possibility of agnosticism in relation to unobservable entities, then this demonstrates that there *is* a viable alternative to scientific realism and instrumentalism, and hence that the binary choice between these two options that is often presented in IR is misleading.

We will explore this thought in more detail once we have considered Van Fraassen's defence of agnosticism about causation. Before we turn to that topic, however, let us consider more fully whether remaining agnostic hinders the conduct of scientific inquiry in a way that would be alleviated by scientific realism.

Does agnosticism about the reality of unobservable entities hinder scientific inquiry?

It is crucial to note, in the context of this question, that Van Fraassen intends his constructive empiricism to be 'non-revisionary' (Rosen 1994: 156): his

[18] Similarly, scientific realism is not refuted if some scientists do *not* believe that some unobservable entities referred to in accepted theories are real (see Rosen 1994; Van Fraassen 1994).

[19] It is worth recalling here that whereas causal idealism is an ontological position, instrumentalism is a view about the meaning of scientific statements: in relation to unobservable entities, the equivalent of causal idealism is not instrumentalism but rather some kind of philosophical idealism. The fact that such a position is not widely defended in contemporary philosophy of science is presumably why advocates of scientific realism in IR have ignored it, focusing on instrumentalism instead.

intention is that it should not demand any revision to the practice of science. To flesh this out, consider the following question, posed by Gideon Rosen: if you adhered to constructive empiricism, and hence defended the possibility of agnosticism in relation to unobservable entities, would there be 'anything in actual scientific practice that you could not in good conscience *do*—anything that depends for its rationality or intelligibility on believing in the unobservable objects your theory describes, or in having as your aim the discovery of truths about unobservables'? (1994: 155). His answer is that '*nothing in the actual practice of science … is incompatible with the view that science aims to produce empirically adequate theories and that acceptance involves as belief only the belief that our best theories are empirically adequate*' (Rosen 1994: 153; see also Van Fraassen 1994: 184).[20]

Moreover, we have already encountered a substantive reason to doubt that agnosticism has any negative consequences for scientific inquiry: the fact that empirical evidence can count for or against a theory's empirical adequacy but cannot speak to the truth of what it says about unobservables. This is important because it indicates that there is no evidence that a realist might cite in support of their claim that a theory is true—true in all it says, including what it says about unobservables—which would not also support the agnostic's claim that this same theory is empirically adequate—that is, true in what it says about observable parts of the world. If a realist considers that the evidence is such as to justify acceptance of a particular theory as true, then this same evidence must also justify acceptance of the theory as empirically adequate. Consequently, there can be no theory that is acceptable to a realist which is not also acceptable to an agnostic.

One specific claim that has, nevertheless, been advanced in IR about how philosophies of science impact upon scientific practice is Colin Wight's contention that 'the realism principle ('is it really the case?') provides the motor which keeps science running'; it is what prevents science from prematurely coming 'to an end' (2007: 383). In relation to Van Fraassen's constructive empiricism, the concern might be that if the aim of science is *merely* to achieve empirical adequacy, then once empirical adequacy has been achieved, further opportunities to explore how the world 'really works' (Wendt 1999: 61) may be foregone. Let us set aside the question of how an empirical science could establish how the world 'really works' except by

[20] Wendt is therefore quite wrong to say that whereas 'realists assume that scientists, not philosophers, are the final arbiters of what is "scientific"', empiricism is 'explicitly an artificial reconstruction of what scientists are or should be doing' (1987: 351). Van Fraassen's aim is *not* to tell scientists what they should be doing, but to provide an account of the aims of science that is adequate to actual scientific practice. He has this in common with Bhaskar (2008).

providing more and more empirical evidence (see Chernoff 2002: 195). The important point to appreciate is that on Van Fraassen's account science could never come to an end.

Van Fraassen points out that 'there have always been reasons to doubt the empirical adequacy of extant theories' (1980: 93), from which it follows that there have always been reasons to continue with scientific inquiries. Recall, in this context, that a theory is empirically adequate if what it says about the *observable* world is true. Because what is *observable* is considerably more extensive than what has so far been (or will ever be) *observed* (see Van Fraassen 1980: 12; Rosen 1994: 161–2), it is not clear how a science which aims to provide theories that are true in what they say about *all that is observable* (now and in the future), and which is also committed to holding all factual beliefs hostage 'to the fortunes of future empirical evidence' (Van Fraassen 2002: 63), could *ever* come to an end (see Monton and Van Fraassen 2003: 407–8)!

With these considerations in mind, it is hard to see how agnosticism might hinder scientific inquiry. At one point, Wendt, who is a leading advocate of scientific realism in IR, appears to acknowledge this: he observes that 'realist and anti-realist physicists disagree about the ontological status of quarks, but this does not affect their research' (1999: 48; see also 41). Why, then, is scientific realism so strongly advocated in IR?

Sometimes, positive arguments *for* scientific realism are presented. For example, Alex Wendt contends that the 'most convincing argument for realism is what is known as the 'Ultimate' or 'Miracle' argument' (1999: 64; see also 64—7): the argument that realism 'is the only philosophy that doesn't make the success of science a miracle' (Putnam 1975: 73).[21] More often, however, scientific realism is simply presented as the default option. Let us give two brief examples.

[21] The basic idea behind this argument is as follows: if scientific theories latch on to the deep structure of the world as it exists independently of human thought, then this would appear to 'explain' the success of science in enabling us to predict and to manipulate the world; to anti-realists, this success is inexplicable; the 'best explanation' for the success of science is hence that scientific realist claims that our best scientific theories *do* latch on to the deep structure of the world are true; scientific realism should therefore be accepted (see Wendt 1999: 64–7; Chakravartty 2017: §2.1). However, anti-realists, such as Van Fraassen, can resist this argument in several ways. First, they can resist the idea that the success of science *requires* explanation (see Van Fraassen 1980: 23–5). Second, they can provide non-realist explanations for the success of science. For example, Van Fraassen argues that modern scientific theories succeed in helping us to control and manipulate the world because only theories which are successful in this way survive (1980: 40). A third way for anti-realists to resist the 'miracle' argument is to point out that it relies on 'inference to the best explanation': the idea is that *if* the truth of scientific realism constitutes the best explanation for the success of science, *then* scientific realism should be accepted as true. Van Fraassen denies that this form of argument is rationally compelling: although he recognizes that we do often argue in this way, he insists that it is always reasonable to decline to accept a conclusion reached in this way (for a detailed discussion, see Van Fraassen 1989: 142–9; Psillos 1996; Ladyman, Douven, Horsten, and Van Fraassen 1997).

First, advocates of scientific realism in IR tend to envisage a binary choice between scientific realism and instrumentalism: identifying significant problems with instrumentalism, they therefore come to endorse scientific realism (see, for example, Wendt 1999: 60–4; Wight 2007). As we have pointed out, however, Van Fraassen is not an instrumentalist: the agnostic position he defends constitutes a neglected alternative to both scientific realism and instrumentalism.

Second, advocates of scientific realism in IR point out that simplistic ideas associated with *outdated* forms of empiricism have, historically, shaped causal inquiry in the social sciences in potentially problematic ways: for example, by steering mainstream research away from a more systematic theorization of unobservable social structures (Kurki 2008; see also Wendt 1987; Wendt 1999; Patomäki and Wight 2000). However, contrary to what these advocates of scientific realism appear to suppose, the fact that previous forms of empiricism are now recognized to be outdated does not constitute an argument *for* scientific realism unless scientific realism is the *only* alternative. As Van Fraassen demonstrates, it is not!

The broad point here is that the case that is typically made for scientific realism in IR is that it is superior to the available alternatives. Yet, when the range of possible alternatives is outlined, the agnostic position defended by Van Fraassen is systematically neglected. Once the viability of agnosticism as a basis for scientific inquiries is recognized, the case for scientific realism looks much weaker.

III. Agnosticism about the reality of causation

So far, we have examined Van Fraassen's defence of agnosticism about the reality of unobservable entities. We will now consider how he defends agnosticism about the reality of causation. We will proceed in the way we did in Section II, first presenting Van Fraassen's position and then asking whether agnosticism hinders causal inquiry in a way that realism does not. That will put us in a position, in the final section of this chapter, to present our own stance and to indicate what we will take forward from this discussion into our analysis of the logic and conduct of causal inquiry in Part II of this book.

Van Fraassen's defence of causal agnosticism

In discussing the possibility of agnosticism about the reality of unobservable entities, Van Fraassen started by noting that many of our best scientific theories do in fact refer to such entities. As we have seen, he insists that these

theories must be construed literally—that is, as meaning what they say when they assert the existence of unobservable entities—yet he argues that one can accept a theory, and construe it literally, while bracketing its ontological implications, remaining agnostic about whether the unobservable entities to which it refers really exist in the world independently of human thought. One might suppose that he pursues an analogous line of argument in relation to causation, holding that what our best scientific theories say about causal relations must be construed literally, but that one can accept these theories, construing them literally, and yet bracket their ontological implications, remaining agnostic about whether the causal relations they identify really exist in the world independently of human thought. However, this is *not* the line he in fact takes. The reason why his defence of agnosticism about causation differs from his defence of agnosticism about unobservable entities is important.

Whereas Van Fraassen acknowledges that many of our best scientific theories refer to unobservable entities, he denies that these theories actually make references to causation and causal relations. As he puts it: 'To say that a sequence [of events] is a causal sequence' is to say 'that it is a sequence of a certain kind of events: causings, so to speak' (1993: 435). But whereas our best theories (in physics) undoubtedly refer to unobservable entities such as photons and quarks, Van Fraassen argues that 'the models which scientists offer us contain no structure which we can describe as putatively representing causings, or as distinguishing between causings and similar events which are not causings' (1993: 437–8).[22]

Van Fraassen explicitly accepts that 'when scientists describe the world they do so in causal discourse' (1993: 438). He also accepts that 'causal discourse is irreducible': if we attempted to translate many of the ideas which are ordinarily expressed using causal language into non-causal language, 'the loss would be severe' (1993: 440–1). He is therefore *not* suggesting that scientists should refrain from using causal language. His point is that the role that causal discourse plays in science is quite different from the role played by terms which refer to unobservable entities. When scientists use terms which refer to unobservable entities, they are using terms which are taken directly from the most precise statements of their best theories. By contrast, when scientists employ causal discourse, they are not using terminology which is taken directly from the most precise statements of their best theories. Rather, they are employing

[22] The term 'causings' is Nancy Cartwright's (1993): Van Fraassen is objecting to her contention that scientists observe 'causings'.

a language that is *not* given in their theories to render more intelligible what these theories are telling us.

Consider Newton's laws of motion, which Van Fraassen summarizes, in rough and ready form, as stating, first, that 'the velocity of a body to which no force … is applied remains constant', second, that the force acting on a body 'equals the mass times the acceleration', and third, that 'action equals reaction'. Van Fraassen points out that in Newton's own thinking, these laws probably served to define 'the causes of acceleration'. Yet, as the theory was subsequently developed, 'it became equally usable without the acceptance of any idea of causation: taken in textbook form it only says that for each acceleration there exists a force incident on the body, equal to the acceleration times the mass' (1989: 282; see also 1989: 7, 350).

It is worth taking care to appreciate accurately what Van Fraassen is saying here. He acknowledges that, in practice, Newton's laws of motion are often *described*, or *characterized*, using causal language, but insists that it is possible both to state them and to understand them fully *without* using causal language. (By contrast, it is *not* possible to understand what a theory says by excising its references to unobservable entities.) Moreover, he points out that the most *precise* statement of such laws—that is, as a formal series of equations—does not contain *any* terms which refer to causes (see 1989: 282–3)! He argues, in short, that while causal language is 'prevalent in the informal presentation of scientific theory' (1989: 213), this language is a 'verbal gloss' (1989: 6) which does not correspond to anything found in the most precise statements of those theories.[23]

This observation about the use of causal language can also be applied to theories in IR. For example, it might seem natural to characterize neorealism by using causal language—to say that it describes the *effects* of systems of different polarities, identifies *causes* of state behaviour, or *explains* outcomes in the international political system. It is interesting to note, however, that while Waltz himself talked, sporadically, of causes (see, for example, 1979: 18, 74, 78, 129, 161), his own authoritative statements of what his theory reveals are typically *not* framed in causal terms. For example, he says that it generates the following 'expectations about behaviours and outcomes': that 'states will engage in balancing behaviour', that there will be 'a strong tendency toward balance in the system', and that 'states will display characteristics common to competitors: namely, that they will imitate each other and become socialized

[23] If, hypothetically, the most precise statement of a scientific theory *were* to contain the assertion that a causal relation exists, it would presumably remain possible to follow the approach Van Fraassen takes in relation to unobservable entities—that is, to accept the theory but bracket its ontological implications, remaining agnostic about whether what it asserts about causal relations is true.

to their system' (1979: 128). Similarly, his most famous finding, that bipolar systems are more stable than multipolar systems (see 1979: 161–93), is most naturally stated without using causal language.

If Van Fraassen is right that scientific theories do not refer to causal relations in the same way that they refer to unobservable entities, what is implied by his acknowledgement that scientists do, nevertheless, often describe the world using causal discourse and that this language is irreducible—that we cannot do without it? The key to understanding his position is to appreciate that he draws a categorical distinction between the language appropriate to 'everyday life and applied science', on the one hand, and the language appropriate to 'theoretical science', on the other hand (1993: 441). As we have seen, he argues that causal language is *not* part of *theoretical* science: 'causation is found nowhere in the most fundamental descriptions of nature by modern physics'. By contrast, he argues that causal discourse *is* a central part of *applied* science, for when we draw on theoretical knowledge to manipulate the world around us, we conceive of ourselves as 'agents with goals and intentions, engaged in effective action' (1993: 442). His point is that it would make no sense to seek to manipulate the world (or even to offer explanations or predictions) unless we conceived of ourselves as agents capable of making a difference in the world; hence, when we engage in 'applied science' it is natural that we should employ causal discourse, even though such discourse is not, strictly speaking, part of theoretical science.[24]

Van Fraassen fully recognizes that the language appropriate to applied science is sometimes *also* used in relation to theoretical science. As he puts it: '[w]hen physicists start describing the world of physics to laymen, they will use language which is a metaphorical and analytical extension of the discourse of applied science and everyday life'—that is, causal language (1993: 442). Thus, for example, when describing what Newton's laws of motion reveal, scientists will often introduce causal terminology. Van Fraassen insists, however, that we will not 'do justice' to this feature of scientific discourse 'by reifying its terms, or devising an ontology of causes' (1993: 438). In other words, we should not presume that because scientists sometimes employ causal language to describe what their theories tell us, these theories therefore refer to causes, let alone that causation must, therefore, be a part of the world existing independently of human thought.

[24] Van Fraassen appears to be thinking that in offering explanations and predictions we view the world as susceptible of manipulation (by us) and hence suppose that observed outcomes might have been made different; this supposition is part of what makes explanation and prediction appear valuable. However, he insists that this way of thinking is part of applied science—it stems from how we conceive of ourselves in relation to the world—but is not (strictly speaking) part of theoretical science.

We now have the ingredients necessary to lay out Van Fraassen's defence of agnosticism about the reality of causation. Put simply: he believes that agnosticism in relation to causation is justified because he believes that 'theoretical science' provides no reason to believe that causation really is part of the world existing independently of human thought! Whereas our best scientific theories *do* unambiguously assert the existence of certain unobservable entities, they do *not* unambiguously assert the existence of causal relations. Meanwhile, the fact that we, as humans and scientists, think of ourselves as agents engaged in effective action, and that we therefore find it difficult to function in the world without resorting to causal language, does not constitute *any* kind of reason for asserting that causation is part of the world existing *independently of human thought*. Of course, Van Fraassen recognizes that remaining agnostic about causation will not mean giving up causal language, for the way in which we conceive of ourselves as agents capable of making a difference in the world involves implicit causal thinking. What agnosticism about causation requires is a clear awareness of the gap between the causal language we commonly use to orient ourselves in the world and what science strictly tells us about the world existing independently of human thought.

Despite this clarification, some readers may wonder if Van Fraassen is not in fact a causal idealist. This is not the case. Milja Kurki describes the 'conception of causation' associated with empiricism as entailing that 'causal relations are [just] regularity relations of patterns of *observables*' (2008: 6; emphasis original) and associates empiricism with the idea that causal necessitation is 'simply an "imagined" relation between successively observed events' (2008: 36; see also Patomäki and Wight 2000: 220).[25] Yet, Van Fraassen neither defines the concept of causation merely as a regularity relation nor contends that all there is to causation in the world is regularities, causal necessitation being nothing but an idea.[26] While he warns against supposing that our causal discourse discloses the hidden structure of the world, he also denies 'the contrary metaphysical view that we [merely] project this [causal] structure onto the perceived world' (1993: 435).[27] His reason for rejecting causal idealism is hence the same as that which, in Chapter 1, we

[25] Kurki also associates both of these views with Hume, or, at least, 'Humeanism'. As we showed in Chapter 1, however, Hume did not hold such views.

[26] To express this in the terms we used in Chapter 1: Van Fraassen subscribes neither to the semantic nor to the ontological version of the regularity view of causation.

[27] Admittedly, some of Van Fraassen's formulations might, if read in isolation, sound quite 'idealist'. He argues, for example, that causal discourse describes 'features of our models, not features of the world' (1989: 214; see also 1980: 80). But his point here is not to *deny* that causation is part of the world existing independently of human thought. His point is rather that whereas scientists often employ causal language to *characterize* what their models say, these models do not assert the existence of causal relations in the way that they assert the existence of unobservable entities.

attributed to Hume: in order to know that our causal discourse is *merely* projected onto the world, we would have to know that the unobservable part of the world does *not* include causation, but this is precisely what we cannot know! Hence, when Van Fraassen argues that our causal discourse expresses our sense of ourselves as agents capable of acting effectively in the world, his point is not that this is *all there is* to causation, but rather that we *cannot know* whether there is more to causation than this: we cannot know whether causation is (also) part of the world existing independently of human thought.[28]

Does agnosticism about the reality of causal powers hinder causal inquiry?

As with agnosticism about the reality of unobservable entities, it is important to consider whether agnosticism about causation has any negative consequences which would be alleviated by an embrace of realism and which would therefore undermine the viability of agnosticism as a basis on which to conduct causal inquiries.

Given that an agnostic will refrain from judging whether causation is part of the world existing independently of human thought, let us start by considering whether it makes sense for an agnostic to conduct *causal* inquiries at all. As we have seen, it appears incoherent to do so from an idealist perspective: if causation is believed to be nothing but an idea, then why characterize an inquiry into the relationship between events observed in the world as 'causal'? This issue does not arise for agnostics, for they do not hold such a belief: rather than claiming to know that causation is nothing but an idea, they deny that it is possible to know such a thing: either to know that causation is *nothing but an idea* and that it is hence *not* part of the world existing independently of human thought (as idealists claim), or to know that it *is* part of the world existing independently of human thought (as realists claim). From an agnostic perspective, engaging in causal inquiries means investigating whether there is evidence for putative relationships of a kind which are often characterized as 'causal', but refraining from judgement about

[28] A similar thought might apply to Collingwood, who is often described as a 'causal idealist' (see, for example, Bernstein 2017). Collingwood argues that the original, and hence in his view the 'proper', sense of the word 'cause' is one in which what 'is caused is the free and deliberate act of a conscious and responsible agent and 'causing' him [*sic*] to do it means affording him a motive for doing it' (1937–8: 86). But to draw attention to this sense of the word 'cause', and even to describe it as its 'proper' sense, is not to assert that causation—a relation of production—is *not* a feature of the world existing independently of human thought.

whether this characterization latches on to anything in the world existing independently of human thought. Let us elaborate.

Consider the proposition that an increase in the energy entering a system of a certain kind (say, a car's engine) produces an increase in the energy it puts out (say, as measured by the speed of rotation of the crankshaft). If presented with this proposition, one question we might ask is whether it is supported by suitable evidence. As this is a purely empirical question, the causal realist, the causal idealist, and the agnostic can approach it in the same fashion. They can and should all design and conduct experimental tests of this proposition and, if it is experimentally confirmed, they can and should all report: 'yes, when the energy entering a system of this kind was increased, the energy it put out also increased'.

Another question we might ask is whether, if it is empirically confirmed, this relationship is appropriately characterized as 'causal'—whether it is appropriate to say that an increase in the energy entering a system of this kind *produces* an increase in the energy it puts out. To this question, the realist, the idealist, and the agnostic are likely to give different answers.

The idealist will presumably regard this characterization as inaccurate and hence inappropriate: if causation is believed *not* to be part of the world existing independently of human thought, then it seems misleading to characterize an empirical relationship using causal language. By contrast, the realist will consider it both appropriate and accurate to characterize this relationship as 'causal'. On a realist account, science is capable of identifying causal relations; moreover, the paradigmatic way of doing this is through controlled experiments (see, for example, Cartwright 1983: 6). Realists will therefore consider it both appropriate and accurate to say that what a controlled experiment shows is that an increase in the energy entering a system *produces* an increase in the energy it puts out; they will believe that this language latches on to an aspect of the world existing independently of human thought. The agnostic will side with the realist in considering it appropriate to use causal language in this way, for such language reflects our sense of ourselves as agents capable of intervening in the world and is hard to do without; however, the agnostic will withhold judgement about whether such language is 'accurate' if what this means is that it latches on to an aspect of the world existing independently of human thought, for they will deny that this is something it is possible to know.

What this illustrates is that the difference between realism and agnosticism does *not* relate *either* to how scientists should proceed in designing and conducting causal inquiries *or* to how and when we should use causal language: realism and agnosticism differ only in *the significance they attach to causal*

language when it is used to characterize what science reveals about the world. Whereas realists regard causal language as indicating something important about the world as it exists independently of human thought, agnostics deflate this claim, acknowledging that causal language is intimately bound up with how we, as human beings, relate to and make sense of the world and our place within it, but denying that it is possible to know whether it latches on to a part of the world existing independently of human thought.

Given what we have said so far, it is hard to see why causal agnosticism should hinder causal inquiries. However, we should also consider the ability of the agnostic to offer causal explanations.

Recall that while Van Fraassen wishes to remain agnostic about whether causation is part of the world existing independently of human thought, he acknowledges that causal language is central to the application of science, including explanation. A key challenge he faces is hence to provide an account of causal explanation that, on the one hand, does justice to the established scientific practice of offering such explanations while, on the other hand, reassuring us that this practice does not implicitly presuppose that causation is in fact part of the world existing independently of human thought. We will provide our own account of causal explanation in Chapter 6. At this point, therefore, we will simply present the outlines of Van Fraassen's account, showing how he meets this challenge.

According to Van Fraassen, a causal explanation provides information, drawn from the picture of the world provided by theoretical science, which serves to answer a contextually specific 'why-question' (1980: 126) about why an event of interest occurred. As he puts it, theoretical science provides 'a picture of the world as a net of interconnected events, related to each other in a complex but orderly way' (1980: 123). For example, it indicates that if the energy entering into a system of a certain kind increases, so does the energy it puts out. Van Fraassen argues that what a causal explanation does is to pick out, from this picture of the world, salient pieces of information about how the event of interest stands in relation to events of certain other kinds; this information is selected in order to fill gaps in the understanding of the intended recipient of the explanation—gaps which led them to pose the 'why-question' in the first place (1980: 124–5; cf. Lewis 1993). For example, if someone were to ask why the energy put out by a car's engine increased, one might explain that it was because the energy entering the engine increased; the idea is that information about the relationship between events of these kinds will serve to explain the event of interest (the increase in the energy put out by the car's engine) if it is information which the recipient of the explanation had previously lacked.

Causal language may enter into such explanations in two ways. First, although Van Fraassen himself has serious doubts about 'the adequacy of the terminology of cause and effect to describe' the picture of the world provided by theoretical science, he acknowledges that this picture *is sometimes* described in causal terms (1980: 123–4). This reflects the fact, which we noted above, that scientists do sometimes turn to causal language when describing, in non-technical terms, what their theories tell us. Second, whether or not the picture of the world provided by theoretical science is described using causal language, it is common to describe an event which is said to explain another event as its 'cause'. On Van Fraassen's account, however, this is not because science itself gives us a picture of the world that is hierarchically organized into 'causes' and 'non-causes': in describing an event as a 'cause' of another event, we are *not* following a hierarchy given in science itself (see Monton and Mohler 2021: §2.5). Rather, an event of kind *A* will be described as a 'cause' of an event of kind *B* (the event being explained) when two conditions are fulfilled: (i) when theoretical science indicates that there is a systematic relationship between events of these kinds, such that, under controlled conditions, if an event of kind *A* occurs, so does an event of kind *B*, *and* (ii) when this information is identified as being information that the intended recipient of the information previously lacked and which will help to resolve their puzzle about why the event of interest occurred.

The key point in all this is that, according to Van Fraassen, the use of causal language in the context of explanation does *not* imply that causation is part of the world existing independently of human thought. He recognizes that it would be natural, if asked why the energy put out by a car's engine increased, to say that it was *caused* by an increase in the energy entering the engine, but on his account the characterization of this relationship *as a causal relationship* is not found in theoretical science, but is rather a feature of how we naturally talk in the context of explanation. In short, the practice of causal explanation presupposes nothing about the nature of the world as it exists independently of human thought: whether the characterization of the relationship between events of certain kinds as 'causal' latches on to a feature of the world existing independently of human thought cannot be known and is therefore a matter on which those offering causal explanations can remain agnostic. Moreover, because the agnostic can engage in the same inquiries as the realist and accept the same theories, there is no reason to suppose that they will draw on a different picture of the world when providing explanations; there is therefore no causal explanation that a realist can offer which an agnostic cannot also offer.

Advocates of scientific and causal realism in IR are therefore quite wrong to suggest that in order to expand the range of factors that can be brought into causal analysis in IR—in order to 'recognise the reality and causal nature of such aspects of social life as rules, norms, ideas, reasons, discourses, as well as, importantly, of "structures of social relations"' (Kurki 2008: 11)—it is necessary to embrace realism. Any evidence that a realist might cite in support of a theory concerning these aspects of social life is equally available to an agnostic, and hence the agnostic is just as capable as the realist of incorporating these aspects of social life into their explanations.[29]

Let us therefore return to our motivating question: does causal agnosticism have any negative consequences which would be alleviated by an embrace of causal realism? It should be clear, we hope, that it does not. There are no scientific inquiries that a realist can undertake which an agnostic cannot also undertake and, as we argued in Section II, there is no theory that a realist can accept which an agnostic cannot also accept. Moreover, an agnostic is just as capable as a realist of offering causal explanations and, when doing so, will invoke the picture of the world provided by theoretical science just as a realist will. Hence, there is also no substantive causal explanation that a realist can give which an agnostic cannot also give.

IV. Taking agnosticism forward: methodology, not metaphysics

As we have seen, Van Fraassen defends the possibility of agnosticism in relation to both unobservable entities and causation. He argues that it is possible to accept theories that refer to unobservable entities and to employ causal discourse, while at the same time refraining from any metaphysical judgement about whether unobservable entities or causation are (or are not) components of the world existing independently of human thought. He argues, moreover, that refraining from judgement on these metaphysical questions has no practical bearing on how scientists conduct their inquiries. This challenges the starkly binary terms in which the case for scientific and causal realism has typically been made in IR, for it suggests that scientific and causal realism are *not* the only alternatives to instrumentalism about unobservable

[29] As we have noted, Van Fraassen is concerned solely with the natural sciences and so does not discuss, for example, how an agnostic about causation might deal with the idea that reasons can be causes. In our view, however, this does not raise any issues that are uniquely challenging for agnostics. If it is coherent to refer to physical forces as causes while remaining agnostic about whether causation is part of the world existing independently of human thought, then one can also refer to reasons as causes while remaining agnostic about whether causation is part of the world existing independently of human thought.

entities and to idealism about causation (see Chernoff 2002; Humphreys 2019).

Consider three broad contentions advanced by advocates of scientific and causal realism in IR. First, they contend that mainstream IR is largely instrumentalist in how it refers to unobservable entities such as social structures (see, for example, Joseph 2007: 345–50; Wight 2007) and largely idealist in its approach to causation (see, for example, Patomäki and Wight 2000; Kurki 2008). Second, they argue that these 'positivist' commitments have 'determined what kinds of things existed in international relations', and hence 'determined what could be studied' in IR (Smith 1996: 11; see also Wendt 1987; Patomäki and Wight 2000; Kurki 2008): the combination of instrumentalism and causal idealism is said to have privileged forms of causal analysis focused on studying observable 'patterns of regularities in the world around us' (Kurki 2008: 6; see also Kurki 2006: 193–4) and led to the neglect of unobservable 'structures, processes and conditions' and also 'ideas, meanings and reasons' (Kurki 2006: 197, 203; see also Patomäki and Wight 2000). Third, its advocates argue that realism is the only viable alternative to instrumentalism and causal idealism and that it opens up new, hitherto unappreciated possibilities for causal inquiry in IR. They contend, on this basis, that developing more satisfactory forms of causal inquiry *requires* a turn to realism (see, for example, Patomäki and Wight 2000).

For the sake of argument, let us accept, for a moment, that mainstream IR is largely instrumentalist and idealist, and that this constrains causal inquiry in significant ways. Let us also grant that, as its advocates suggest, a turn to realism would open up new possibilities compared to the status quo. Even accepting all this, our analysis suggests that improving causal inquiries does not *require* a turn to realism: if, as we have argued, agnosticism does not hinder causal inquiry in any way that would be alleviated by an embrace of realism, then any new forms of inquiry that are facilitated by scientific and causal realism are equally available to agnostics. In short, even if advocates of scientific and causal realism are right to say that instrumentalism and causal idealism have hamstrung causal inquiries in IR, they are wrong to present realism as the only viable alternative.

However, if we are right in our contention that there is little, if any, difference in 'cash-value' (James 1907) between scientific and causal realism, on the one hand, and agnosticism, on the other hand, then this raises an obvious question: why has it been worth elaborating an agnostic alternative to scientific and causal realism?

The reason why we consider it worthwhile is that, in our view, debates about causation in IR have become overly preoccupied with metaphysical

concerns, at the expense of what we believe would be a more productive focus on methodological questions. Consider, once again, the contention that instrumentalism and idealism limit causal inquiries in certain problematic ways, and that in order to do better it is necessary to embrace scientific and causal realism. The effect of this argument is to give priority to metaphysics over methodology. What is being suggested is that in order to improve our causal inquiries it is necessary, *as a first step*, to adopt certain metaphysical beliefs—to believe that unobservable entities and causation really are part of the world existing independently of human thought. The corollary of this is that methodological questions are a second-order concern, to be addressed only once a satisfactory metaphysical position has been adopted.

If, as we have argued, agnosticism is a viable alternative to scientific and causal realism as a basis for causal inquiry, then this overturns the prescription 'first metaphysics, then methodology': it is *not* necessary first to adopt a realist metaphysics and only then to broach questions of methodology. In order to conduct causal inquiries, we believe that it is important to avoid the incoherence to which causal idealism gives rise but, as we hinted in the Introduction, we doubt that many of those conducting causal inquiries in IR are committed idealists, so we doubt that this will prove a controversial demand. In conducting causal inquiries, we believe that it is also sensible to construe scientific theories as meaning what they say and hence to reject instrumentalism: if a theory indicates that there is an unobservable structure to the international system, this should be taken at face value. Once again, however, we do not suppose that many readers will consider this advice especially controversial.

So long as these fairly minimal requirements are fulfilled, it is not necessary even to broach metaphysical questions, let alone to resolve them. We argued, in Chapter 2, that Bhaskar's argument for causal realism is unconvincing. Although alternative arguments may be available, we are sceptical of any claim to arrive at reliable knowledge about the real components and structure of the world as it exists independently of human thought. Simply put: we do not see how such knowledge could be developed. One attractive feature of Van Fraassen's defence of agnosticism is that it shows that these kinds of metaphysical concerns can be set to one side.

Like Van Fraassen, we, the authors of this book, are inclined to remain agnostic in relation to the reality of causation. However, like Van Fraassen, we too will stop short of actively advocating agnosticism: readers must decide for themselves whether they believe causation to be part of the world existing independently of human thought or wish to remain agnostic on that question. As Van Fraassen himself puts it: science has 'room for scientific agnostics and

scientific gnostics alike' (1994: 191). That said, we do wish to insist that belief that causation is part of the world existing independently of human thought is *not required* in order to conduct causal inquiries in IR.[30] Whether those conducting causal inquiries embrace realism or agnosticism (or prevaricate between them) has no consequence for causal inquiry itself.

What will make a difference for causal inquiries in IR is to think more carefully about methodological questions (broadly understood) in relation to causal inquiry: to consider what causal statements assert and imply, what kind of empirical evidence can best support causal claims, and what kind of reasoning will be required to develop persuasive arguments and to facilitate constructive debate about causal questions. It is to just these kinds of methodological questions which we will turn in Part II of this book. In exploring these issues, our investigations will draw on the three conceptual and methodological insights which we derived from our analysis of Bhaskar's arguments for causal realism in Chapter 2: (i) the importance of considering whether the environment in which a causal process of interest operates is an open or a closed system; (ii) the importance of controlled experiments in providing evidence for laws of science and assertions of causal relations; and (iii) the importance of recognizing that causal statements describe propensities which unfold fully only when interference is prevented.

As framed here, these insights are no less compatible with agnosticism than with scientific and causal realism. The first two are concerned with scientific practice. As we have pointed out, realists and agnostics do not differ on this issue. The third insight concerns what is being asserted in a causal statement. We will consider this in more detail in Chapter 4. For now, let us simply point out that although the agnostic, unlike the realist, will refrain from judging whether causal statements in fact latch on to part of the world existing independently of human thought, this does not imply that they should disagree about what a causal statement asserts. We therefore contend that these three insights are relevant to all those conducting causal inquiries in IR, regardless of whether they are more inclined towards realism or towards agnosticism.

How we will use causal language in the remainder of this book

It has been an important aim in this chapter to show that agnostics can coherently employ causal language even though they refrain from judging whether

[30] Moreover, as we have explained, we doubt that a convincing argument *for* causal realism can be advanced.

causation is a component of the world existing independently of human thought. Indeed, we have argued that agnostics can use causal language just as unproblematically as realists can. Our contention that those conducting causal inquiries in IR can sensibly adopt either a realist or an agnostic stance therefore carries no implication for how causal language can or should be used.

In the remainder of this book, we will continue to talk of, and endorse, *causal* inquiries and to construe causal language as describing relations of production or bringing about. We will, moreover, employ causal language not only in relation to the application of science, for example, in the context of causal explanation, but also when discussing what theories of international relations tell us about the world. What we take from Van Fraassen is not the possibility of a science shorn of causal discourse, but rather the warning that such discourse does not constitute evidence of the reality of causation; it does not even constitute evidence that those using such language *believe* causation to be a real component of the world existing independently of human thought. To remain agnostic about causation is, among other things, to recognize that our causal language takes us beyond what science, strictly speaking, reveals to us about what the world is like.

In Part II of this book, we will not endlessly repeat this point, for it has no further consequences. Hence, while some readers may wish to remind themselves, every now and again, that this is a background feature of our approach, if other readers prefer to put this issue to the back of their minds (and even to forget about it entirely!), then that is equally acceptable. In our view, it is not possible to make any further progress on our topic—causal inquiry in IR—by continuing to debate such issues. Rather, it is necessary to turn to the logic and conduct of causal inquiry itself. That is what we will do next, in Part II of this book.

PART II

METHODOLOGY

4
Causal Statements

Whereas Part I of this book considered whether causal inquiry must involve some kind of metaphysical commitment, in Part II we turn to issues which are broadly methodological in nature. Our aim is to provide an account of causal inquiry that illuminates the basic logic according to which causal knowledge is developed and that exposes some important misunderstandings which have permeated previous analyses and which have also shaped substantive causal inquiries in IR.

In developing this account, our starting point is a simple observation: a central component of causal inquiry is the search for empirical evidence capable of supporting (or undermining) causal propositions. In attempting to think systematically about causal inquiry, it is therefore sensible to start by considering what kind of causal statement could, in principle, be supported by empirical evidence. This, in turn, will require a clear understanding of causal statements themselves—of what they assert and what they imply.

We will show, in this chapter and the next, that taking the time to appreciate fully the ways in which various kinds of causal statement differ from one another is far more important than has often been realized. As we will show, this is because the basic logic of causal inquiry emerges out of the relationship between two kinds of causal statement: statements which relate to specific events, located in space and time, and statements which abstract from specific events and hence possess a broader applicability. It is crucial to appreciate the relationship between these two kinds of statement, we will argue, because direct empirical support can be provided only for causal statements which relate to specific events, but statements of this kind *imply* statements which abstract from specific events and hence possess a broader applicability; moreover, statements of this latter kind can, in turn, be drawn upon to fulfil a range of goals that are central to causal inquiry, including providing causal explanations and generating causally informed predictions and policy prescriptions.

A full account of this logic will have to await Chapter 5. For now, the key point is that it is crucial accurately to appreciate the relationship between these two basic kinds of causal statement. That will be our focus in this

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0005

chapter. We will consider in detail how statements of these two kinds differ from each other in what they assert, in the linguistic forms they can take, in the distinctive epistemic functions they perform in the context of causal inquiry, and in their implications. As will become clear, nearly all the missteps which we will go on to identify in existing work stem, at least in part, from misinterpretations of how these two kinds of causal statements differ.

Of course, our contention that there *are* two basic kinds of causal statement, and that they differ from each other in important ways, is not a novel one. We hope that even those readers who are unfamiliar with the philosophical literature in this area will accept that what is asserted in a statement such as 'short circuits cause fires' appears, prima facie, to differ from what is asserted in a statement such as 'the short circuit caused the fire'. What is distinctive about our analysis is that it highlights a common but rarely noticed problem: the terminology which is widely used, in IR and in philosophy, to distinguish statements of these two kinds is potentially quite misleading. Typically, it is said that a statement such as 'the short circuit caused the fire' is 'singular', whereas a statement such as 'short circuits cause fires' is 'general'. This familiar and apparently natural language reflects a broader 'culture of generalization' in which it is presumed that the priority in any scientific inquiry should usually be to develop 'general' knowledge.[1] As it informs the analysis of causal statements, however, this culture, and the language associated with it, proves quite problematic: it is liable to give a misleading impression of what particular causal statements assert and imply, and hence to suggest a misleading picture of the logic of causal inquiry.

To cut a long story short, the key problem with the labels 'general' and 'singular' is that they appear to suggest that the fundamental difference between the two basic kinds of causal statement is numerical: 'singular' statements refer to individual events, whereas 'general' statements refer to many events. For example, if the statement 'the short circuit caused the fire' is labelled 'singular', while the statement 'short circuits cause fires' is labelled 'general', these labels might appear to imply that whereas the first statement identifies the causes of *an individual fire*, the second statement identifies the causes of *many fires*, or of *fires* '*in general*'. This is not the case. As we will show, there *is* a fundamental difference between statements of these two kinds, but it does *not* concern the number of events to which they refer: the key difference between them is that the statement 'the short circuit caused the fire' *refers to specific*

[1] The recalcitrant belief, which we explored in Chapter 1, that causation must be somehow intimately connected to correlations or regularities is one way in which this culture shapes predominant ways of thinking about causal inquiry.

events, *located in space and time*, whereas the statement 'short circuits cause fires' *abstracts from specific events to state a causal propensity*.

We recognize that an argument of this kind may prove hard to grasp when presented in this rather compressed form. In developing a more detailed exposition, we will employ numerous illustrative examples. These will include standard examples employed by philosophers who have considered these issues and examples from the study of world politics. By the end of the chapter, we hope to have persuaded readers that there are indeed two basic kinds of causal statement, but that the predominant culture of generalization, and the language associated with it, tends to obscure the nature and significance of the distinction between them. Once this distinction is correctly grasped, it points towards a new and productive way of thinking about the logic of causal inquiry.

In making this case, we will proceed as follows. We will start by briefly illustrating the pervasiveness, in IR and political science, of what we term the 'culture of generalization'. We will then turn to the idea that there are two basic kinds of causal statement which need to be distinguished. We will show why the terms most commonly used to characterize these two kinds of statement—viz. 'general' and 'singular'—are liable to mislead, and we will propose alternative terms which more accurately convey what is distinctive about what statements of each kind assert: we will label causal statements which refer to specific events as 'concrete' and statements which abstract from such events as 'abstract'. We will also consider the variety of linguistic forms that statements of these kinds can take and the distinctive epistemic functions they perform in the context of causal inquiry. We will then ask what each kind of statement implies and indicate why this is important. Only when we have completed this task will we be in a position to elaborate the logic of causal inquiry more fully. That will be the task of Chapter 5.

I. The culture of generalization in IR and political science

In IR and political science, as in the social sciences more broadly, it is common to distinguish between two types of knowledge: 'general' knowledge and knowledge of 'singular', 'particular', 'individual', or 'unique' events or facts.

A prominent example of this distinction is found in King, Keohane, and Verba's discussion of what they describe as 'the seemingly contradictory goals of scholarship: discovering general knowledge and learning about particular facts' (1994: 35). Yet, this is far from confined to mainstream research

methods texts. For example, it is also widely invoked to characterize the relationship between social science and history, in which guise it is often described as the 'nomothetic–idiographic' distinction. Jack Levy argues, for instance, that historians tend to 'describe, explain, and interpret individual events … whereas political scientists generalize about the relationship between variables and construct lawlike statements about social behaviour' (1997: 22). Of course, many historians *deny* that they merely analyse 'individual' events: according to Paul Schroeder, for example, 'even histories that are narrative-descriptive in form, including most work in international history, are clearly nomothetic in the sense that they … establish general patterns' (1997: 66; see also Gaddis 1997: 83). Yet, what is most notable in this debate is the unquestioning acceptance by all participants that it is sensible and useful to distinguish between knowledge of the 'general' and of the 'particular'.

This distinction reflects and reinforces what we term a 'culture of generalization' which permeates the social sciences. In broad terms, this culture is characterized by a tendency to identify the distinction between knowledge of 'the general' and of 'the particular' as a key dividing line in the pursuit of knowledge and to prioritize the production of 'general' knowledge over knowledge of 'particular' facts and events as a key goal of the social sciences.[2] The presumption to which this culture gives rise is not only that, in social science, 'general' knowledge is typically more valuable than knowledge of the 'particular', but also that 'general' knowledge is expressed in 'generalizations' and, further, that 'general' knowledge, in the form of 'generalizations', is what is articulated in social scientific 'theories'. It is worth briefly elaborating this a little further.

First, it is widely assumed that, for the purpose of social science, general knowledge is typically more valuable than knowledge of the particular. For example, in his seminal work *Man, the State and War* (1959), Kenneth Waltz famously distinguished explanation of war in general from explanation of particular wars and argued that the former was more fundamental than the latter (see Suganami 2009: 376–7). More broadly, it is commonly supposed that, as King, Keohane, and Verba put it, '[g]ood social science attempts to go beyond … particulars to more general knowledge' (1994: 35; see also Bueno de Mesquita 1985: 133; Shadish, Cook, and Campbell 2002: 18–26; Lucas 2003: 236; Collier, Seawright, and Munck 2010: 35; Johnson and Reynolds 2012: 40). There is, of course, a debate to be had about what trade-offs are optimal in this regard (see Collier, Brady, and Seawright 2010a:

[2] Similarly, Jackson (2011: 8) points out the priority often 'accorded to "science" understood as the quest for generalized theoretical knowledge'.

153–9; Sil and Katzenstein 2010), but it is widely accepted that the aim of social scientific inquiry should be to develop knowledge at the most 'general' level that is compatible with the particular research aims and with the nature of the subject matter. This is taken to apply just as much to causal as to descriptive knowledge (see King, Keohane, and Verba 1994: 34).

Second, 'general' knowledge is not only assumed to be more important than knowledge of the 'particular', but is also often associated with 'generalization' and with observable regularities. For example, King, Keohane, and Verba (1994: 10) recommend dealing with 'unique' events 'by seeking generalizations: conceptualizing each case as a member of a *class of events* about which meaningful generalizations can be made'. Similarly, John Gerring (2010: 1502) argues that 'the underlying goal of social science is to generalize about social reality as much as possible, that is, to understand classes of cases (causal regularities or deterministic "laws") rather than outcomes specific to individual cases' (see also Johnson and Reynolds 2012: 41). Judith Goldstein and Robert Keohane (1993: 29; cited in Kurki 2008: 105) relate this point about the importance of generalization explicitly to *causal* inquiry: they insist that without 'valid generalizations … no causal analysis [would] be of much value'. This suggests that the knowledge which it is the task of the social sciences to develop through causal inquiries should be not only 'general', but 'general' *in the sense that it involves, or aims at, generalization*.

Third, 'general' knowledge, in the form of 'generalizations', is often linked to 'theory' and to the idea that one of the principal functions of theories is to furnish explanations of particular events. For example, Kiser and Hechter (1991: 4) argue that 'those who seek causal explanations of historical events cannot hope to dispense with general theories', while Johnson and Reynolds (2012: 43) argue that the most important component of a theory is a set of 'general, verifiable statements that explain the subject matter'. Once again, this idea is accepted not only by committed theorists but also by many historians. For instance, John Gaddis (2002: 62) insists that it is 'quite wrong to claim that historians reject the use of theory, for theory is ultimately generalization, and without generalization historians would have nothing whatsoever to say'. Applied to causal inquiry, this points towards the idea, articulated explicitly by Stephen Van Evera (1997: 40), that causal claims 'come in two broad types: theories and specific explanations. Theories are cast in general terms and could apply to more than one case … Specific explanations explain discrete events—particular wars, interventions, empires, revolutions, or other single occurrences'. We suspect that ideas of this kind are widely accepted in IR.

To many readers, there may appear to be little that is surprising or controversial in any of these tendencies and we, the authors, readily acknowledge that 'general' knowledge, in the form of 'generalizations', can be extremely valuable. We believe, however, that as a guide to understanding causal statements the culture of generalization is quite unhelpful—even harmful. When it comes to analysing causal statements, the tendencies we have just elaborated, and the vocabulary of the 'general' versus the 'particular' (or 'singular') which underpins them, point in the wrong direction.

Consider, once again, a statement of the form 'short circuits cause fires'. Clearly, such a statement is broader in scope than the statement 'the short circuit caused the fire'. This breadth may lead us to characterize it as 'general' and/or to think of it as stating a causal theory. But is it 'general' in the sense that it states a generalization? We think not. It has the look and feel of a theoretical statement, but it does not generalize about an identifiable set of events: it does not tell us, for example, that in such and such circumstances, during such and such a time period, short circuits have sometimes, often, or always caused fires. This creates something of a puzzle, at least from the perspective of the culture of generalization, for it suggests that some causal statements possess a 'general' and 'theoretical' quality, yet are not generalizations. Our task in the next section is to resolve this puzzle.

II. The two basic kinds of causal statement

Among philosophers, it is common to distinguish two basic kinds of causal statement: one kind relates to specific events, located in space and time, whereas the other kind does not relate to specific events but appears to hold more broadly. We endorse the idea that there are these two basic kinds of causal statement; moreover, we will show that the distinction between them is crucial for the logic of causal inquiry. As we shall see, if only one kind of statement relates to specific events, located in space and time, then only this kind of statement can be directly supported (or undermined) by empirical evidence. We will explore this issue in more detail in Chapter 6. For now, our aim is to interrogate the difference between the two basic kinds of causal statement. We will show that the distinction between them is a crucial one, but that the terms which are widely used, in philosophy and IR, to mark this distinction—'general' and 'singular'—are quite misleading.

We will propose alternative terminology which, in our view, indicates more clearly how the two basic kinds of causal statement differ from each other and why the distinction between them matters. It is important to emphasize,

however, that in proposing to replace one set of labels with another we are not engaged in mere semantic nit picking. What we will draw attention to is a fundamental difference in what is being asserted in statements of the two kinds, a difference which the labels 'general' and 'singular' tend to obscure. In particular, we will show that those statements which are typically labelled 'general' are *not* generalized counterparts of so-called 'singular' statements. In fact, they are not generalizations at all, but rather *propensity* statements: they abstract from specific events to describe how causal relations will unfold if the conditions are right. Although this feature of causal statements is not widely understood, it will become clear, as our discussion progresses, that it is critically important for the logic of causal inquiry and hence for the conduct of all causal investigations in IR.

The problem with the terms 'general' and 'singular'

The terms 'general' and 'singular' are widely used to differentiate two basic kinds of causal statement (see, for example, Mackie 1965, 1980; Little 1995: 34–5; Hitchcock 1995, 2018: §1.4). The paradigmatic form of a so-called 'singular' statement is '*a* caused *b*', where the lower case '*a*' and '*b*' stand for specific events, located in space and time. An oft-cited example is 'the short circuit caused the fire' (Woodward 1984: 231; see also Mackie 1965). Another example is 'the assassination of Archduke Franz Ferdinand caused the outbreak of World War I' (Little 1991: 13). By contrast, the paradigmatic form of a so-called 'general' statement is '*A* causes *B*'. In such statements, the upper case '*A*' and '*B*' stand, generically, for events of certain types (events of the types which feature in corresponding statements of the form '*a* caused *b*'); they do *not* point to specific examples of these events in the way that the lower case '*a*' and '*b*' do.[3] Thus, the (supposedly) 'general' counterpart to 'the short circuit caused the fire' is 'short circuits cause fires'. Another example of a putatively 'general' causal statement is 'hyperinflation causes political instability' (Little 1995: 35).

Before we analyse this distinction further, it is worth noting one potential source of confusion: some philosophers have used the terms 'general' and

[3] In what follows, we will always use lower case letters as placeholders for descriptions of specific events, located in space and time. Thus, for example, the letters '*a*' and '*b*' in the statement '*a* caused *b*' could be substituted for references to a specific short circuit and a specific fire. We will always use upper case letters as placeholders for generic descriptions of events of particular types—that is, descriptions which do *not* pick out specific examples of those event-types. Thus, the letters '*A*' and '*B*' in the statement '*A* causes *B*' could be substituted for references to generic short circuits and fires, references which do *not* pick out any specific examples of these phenomena.

'singular' (and the like) not to distinguish *kinds of causal statement*, but rather to distinguish *kinds, or levels, of causation*. These philosophers have debated whether there *are* different levels of causation and, if there are, which level is primary. For example, it is sometimes claimed that, in addition to causal links between specific events, located in space and time, there is also 'general causation'—that is, a level of causation which links generic 'properties' of events (Mellor 1995: 6) or resides in 'general patterns (regularities)' (Psillos 2002: 127). On the opposite side of this debate, so-called 'singularists' about causation insist that causation 'is essentially a relation between concrete individual events' and that regular patterns, if observed, are a mere 'corollary' of causal relations between concrete individual events (Ducasse 1975: 118, 114; see also Von Wright 1975: 108).

On occasion, this debate about levels of causation has made its way into IR. For example, Ned Lebow argues that 'many, if not most, international events … are best described as instances of what philosophers call singular causation'. His point is that it may be possible to 'construct causal narratives' about specific events observed in world politics even if these events 'cannot be explained or predicted by reference to prior generalizations' (2014: 5–6). We endorse the broad thrust of Lebow's view here, but in our judgement the phrase 'instances of … singular causation' muddies the water: it risks conflating a question about what causal propositions can reasonably be advanced on the basis of the available evidence with a question about kinds, or levels, of causation. We will briefly return to this issue later, but for now our focus is exclusively on how the terms 'general' and 'singular' are used to distinguish kinds of causal statement.

An indication of the difficulties which arise when the terms 'general' and 'singular' are used to distinguish the two basic kinds of causal statement is that philosophers seriously disagree about how these two kinds of statement relate to one another. Some argue that so-called 'general' statements generalize 'singular' ones. For example, Mackie (1980: 80) contends that 'general statements' are 'quantified variants of the corresponding singular ones'. On his view, the 'general' statement 'heating a gas causes it to expand' is a 'generalization' of the 'singular' statement 'heating the gas caused it to expand'. Mackie adds that what this 'general' statement asserts is that 'heating a gas always or often or sometimes causes it to expand' (see also Ducasse 1975: 124, Lewis 1975: 182; Mellor 1995: 6–7; Illari and Russo 2014: 38).

On the other side of this debate, Christopher Hitchcock explicitly challenges the idea that what he, too, terms 'general' causal statements, such as

(A) 'Smoking causes lung cancer',

are 'generalizations over' corresponding 'singular' causal statements, such as

(B) 'David's smoking caused him to develop lung cancer'.[4]

He points out that a *generalization* over statements of kind (B) might be expected to contain plural event-terms, as in

(C) 'Smokings cause lung cancers' or
(D) 'Episodes of smoking cause episodes of lung cancer' (1995: 263–4).

We agree, but we believe that his argument can and should be taken further. One might expect a generalized counterpart of (B) not only to contain plural event-terms (e.g., 'smokings') but also to be expressed in the same tense as (B)—that is, in the past tense—and to specify the population within which it holds. In other words, it should be something like

(E) '(In such and such a place, during such and such a time period), many people's smoking caused them to develop lung cancer'.

In contrast to (E), it is notable that (A) does *not* contain plural event-terms, it is *not* expressed in the same tense as (B), and it does *not* specify a relevant population!

The issue here is not a pedantic one about the precise form in which causal statements are written. As Hitchcock observes, statement (B) asserts, among other things, that 'the named events' (that is, David's smoking and his developing lung cancer) *occurred* (1995: 266; see also Mackie 1980: 31).[5] The same is true of statement (E): it asserts, among other things, that many people in a given time and place smoked and that some of them developed lung cancer. Whether these events in fact occurred could, in principle, be checked. By contrast, Hitchcock notes that a so-called 'general' causal statement, as exemplified by (A), does *not* assert 'anything about the instantiation of the event-types figuring in it'; statement (A) asserts that there is a causal relationship between smoking and lung cancer, but it does not refer to any specific

[4] We apologize to any readers who may be distressed by these kinds of examples, which feature prominently in the philosophical literature. One reason for this prominence, we surmise, is that statements of this kind are widely accepted, even by laypeople, which makes it possible to analyse their contents and implications without becoming waylaid by questions about their acceptability.

[5] Actually, Hitchcock says that statement (B) asserts that the named events 'occur'; this is puzzling, given that the statement is phrased in the past tense.

events the occurrence of which could be checked. In fact, Hitchcock points out that statements of this kind may be scientifically acceptable even if they have *never* been instantiated and there is no prospect of them *ever* being instantiated; consider, for example, the statement 'eating 1kg of uranium 235 causes death' (1995: 265–6; see also Hempel 1966: 57; Sober 1984: 409; Woodward 2003: 40)!

With this in mind, let us reconsider Mackie's contention that a 'general' causal statement of the form '*A* causes *B*' asserts that *A*-type events 'always or often or sometimes' cause *B*-type events. Statement (E) *does* make such an assertion: it asserts that, in a given time and place, smoking caused lung cancer 'always or often or sometimes'. However, statement (A) does *not* make such an assertion; it does *not* assert anything about the frequency with which the causal relationship it describes has been or will be instantiated in any specific context.[6] Great care is therefore required in analysing so-called 'general' causal statements. Statement (E) is a generalization: it generalizes across multiple events within a given population. By contrast, statement (A) is clearly *not* a generalization: it does not generalize about, or even refer to, actual events. Labelling statements of kind (A) 'general' is therefore liable to mislead, especially if, under the influence of the culture of generalization, we are accustomed to associating the term 'general' with 'generalization'.

It is also worth noting, more briefly, that the term 'singular' is somewhat problematic as a label for statements which refer to specific events. The basic problem is that the specific events to which these statements refer need not be 'single' events and they certainly need not be events concerning individual people, states, and the like. Illari and Russo highlight this by means of the following example: 'In Japan 1966 was the "Year of the Fire Horse". There was a popular myth saying it was bad luck to be born in that year, especially for women. This folk belief led to a drastic birth drop, of about 25%, in 1966' (2014: 41).

The statement 'the folk belief caused the reduced birth rate' has the same syntactical form as the statement 'the short circuit caused the fire', yet Illari and Russo point out that the reduced birth rate is a population-level event, not an event at the level of individuals.[7] We agree, but, once again, we believe that this point should be pushed further. Although the statement 'the folk belief caused the reduced birth rate' refers to a population-level

[6] Most readers will have little doubt that smoking does in fact cause lung cancer 'always, or often, or sometimes'. But the question we are addressing is not what is known, but what statement (A) asserts.

[7] This kind of population-level event may, however, emerge out of individual-level events. We consider this possibility in more detail in Chapter 6.

event, it still refers to a 'single' event and so the label 'singular' does not seem wholly unreasonable. By contrast, it would seem perverse to label the statement 'the short circuits caused the fires' as 'singular' given that it clearly refers to many short circuits and fires. On the other hand, the statement 'the short circuits caused the fires' appears similar to the statements 'the short circuit caused the fire' and 'the folk belief caused the reduced birth rate' insofar as all three statements refer to specific events, located in space and time.

The problem with the term 'singular' is hence that it obscures the important commonality that statements concerning single events (such as 'the short circuit caused the fire' and 'the folk belief caused the reduced birth rate') share with statements concerning multiple events (such as 'the short circuits caused the fires'): these statements all refer to specific events, located in space and time. In conjunction with the term 'general', the term 'singular' therefore also obscures the crucial distinction between causal statements which do refer to specific events, located in space and time, and statements, such as 'short circuits cause fires', which do *not* refer to specific events, but rather abstract from them.

Clearer terminology: 'abstract' versus 'concrete' causal statements

These considerations cast serious doubt on the aptness of the terms 'general' and 'singular' as labels for the two basic kinds of causal statement; they also undermine the idea that statements of the form '*A* causes *B*' *generalize* statements of the form '*a* caused *b*'. It is therefore notable that some philosophers either cast doubt on the utility of these terms (see, for example, Kim 1975: 57–8), propose alternative labels (see, for example, Illari and Russo 2014: 41–2), or suggest that the terms 'general' and 'singular' really mean something quite different from what one might intuitively expect. One common suggestion in this regard is that, as applied to causal statements, the terms 'general' and 'singular' *really mean* 'type-level' and 'token-level' respectively.[8] For example, Hitchcock (2018: §1.4) argues that the two main kinds of causal statement are (i) token-level statements, such as 'David's smoking caused him to develop lung cancer', which 'make reference to particular individuals, places, and times' (often labelled 'singular') and (ii) type-level statements,

[8] The type–token distinction is complex (see Wetzel 2006), but the basic idea is that a token is an instance of a phenomenon of a particular type. For example, this footnote is a *token* of a familiar *type* of phenomenon, viz. 'footnotes in academic monographs'.

such as 'smoking causes lung cancer', which 'do not refer to particular individuals, places, or times, but only to event-types or properties' (often labelled 'general').

This is an important suggestion, for it seems to make sense of the fact that statements of the form '*A* causes *B*' possess a broader applicability than statements of the form '*a* caused *b*', but without characterizing the former as generalizations. On Hitchcock's account, the statement 'short circuits cause fires' has a broader applicability than the statement 'the short circuit caused the fire' because whereas the latter refers to a specific short circuit and a specific fire, the former refers, generically, to events of these types (short circuits and fires). We endorse this basic idea. However, we will not adopt the terms 'type-level' and 'token-level'. As James Woodward points out, these terms are potentially misleading insofar as they seem to suggest that 'there are two *kinds* of causation—type and token' (2003: 40). For example, if the statement 'short circuits cause fires' is described as a 'type-level' statement, this may appear to imply that there is a distinctive *kind* of causation which holds between these event-types, independent of the relations between actual short circuits and actual fires. This strikes us as obviously wrong-headed. It is one thing to say that *events* of one type produce *events* of another type; it is another thing entirely to say that some *types* produce other *types*! If, as we have argued, causation is a relation of production, then it can only hold between concrete events of particular types (and not between the 'types' themselves).[9]

As this discussion has started to get a little complex, let us draw the threads together. We have endorsed the idea that there is a crucial distinction between two basic kinds of causal statements: those, such as 'the short circuit caused the fire', which refer to specific events, located in space and time, and those, such as 'short circuits cause fires', which abstract from any such specific events and instead refer, generically, to events of particular types. However, we consider the terms 'singular' and 'general' to be misleading as labels for these two kinds of statements. We also have reservations about the labels 'token-level' and 'type-level'. Clearer terminology is hence desirable. We believe that it is readily available. We propose labelling causal statements which refer to specific events, located in space and time, as 'concrete'. This label highlights the fact that the events to which such statements refer are specific, actual events. We propose labelling causal statements which abstract from specific events, located in space and time, and instead refer, generically, to events

[9] This is why, as we noted earlier, we find Lebow's phrase 'singular causation' (2014: 5–6) unhelpful: it appears to suggest that there are two distinctive kinds of causation, one of which links actual events (this is what Lebow terms 'singular causation') and one of which links something else, such as 'types' of events (this is what is sometimes referred to as 'general causation').

of particular types, as 'abstract'.[10] This label highlights the fact that these statements possess a broader applicability than concrete causal statements *because they abstract away from the specific events to which concrete statements refer*: the notion that these statements are *generalizations* is therefore a red herring. This is summarized in Table 4.1.

The reason why we have spent several pages discussing what terminology to use to characterize the two basic kinds of causal statement is that it is crucial, when conducting causal inquiry, to have an accurate understanding of what is being asserted in a causal statement of a given kind. The labels 'general' and 'singular' appear to suggest, incorrectly, that the two basic kinds of statement each advance assertions about causal relations linking concrete events and that the difference between them is numerical: 'singular' statements refer to individual events of particular types, whereas 'general' statements refer to many events of those types. The key advantage of the alternative labels we propose—'abstract' and 'concrete'—is that they better indicate the much more significant difference in what statements of these two kinds assert.

Table 4.1 Ways of distinguishing between the two basic kinds of causal statement

Paradigmatic form	**'*a* caused *b*'**	**'*A* causes *B*'**
Examples	'The short circuit caused the fire' 'David's smoking caused his lung cancer' 'The French Revolution caused the French Revolutionary Wars'	'Short circuits cause fires' 'Smoking causes lung cancer' 'Revolutions cause wars'
What it refers to	Specific, actual events (*a* and *b*), located in space and time	Events of particular types (*A* and *B*)—that is, events of the types referred to in statements about specific events
Common label	'Singular'	'General'
Our alternative label	'Concrete'	'Abstract'

[10] Typically, a statement or model is said to be an 'abstraction' if it is concerned only with some aspects of what it describes or represents (see, for example, Sayer 1992: 87; Chakravartty 2007: 190). In characterizing a statement of the form '*A* causes *B*' as 'abstract', we are drawing attention to the fact that although it describes a causal relation, it provides information only about the kinds of events which this causal relation links and not about when or where this causal relation has been or will be concretely manifested. However, unlike Sayer (1992: 133), we will not embrace a further distinction between 'abstract' and 'concrete' causal *explanations* and, unlike Kurki (2008: 213), we will not associate 'abstraction' with the search for 'deeper' or 'underlying' structures.

The content of what we are terming a 'concrete' causal statement is reasonably transparent: it describes the instantiation of a particular causal relationship in a particular time and place, linking two or more specific events. Thus, for example, the statement 'David's smoking caused his lung cancer' describes a concrete instantiation of the causal relation 'smoking causes lung cancer': it asserts that David smoked, that he developed lung cancer, and that the first event caused the second.

By contrast, the content of what we are terming an 'abstract' causal statement is perhaps somewhat less transparent. As we have pointed out, an 'abstract' causal statement does not advance an assertion about specific events or about concrete instantiations of the causal relationship it describes. In abstracting away from specific events, what it describes is a *causal propensity*. The statement '*A* causes *B*' describes a *propensity* for *A*-type events to produce *B*-type events: it asserts that there is a causal relationship between *A*-type and *B*-type events, such that, if this relationship unfolds without interference, an *A*-type event will produce a *B*-type event.[11]

As we will use the term, a 'propensity' should be distinguished sharply from a statistically dominant pattern, frequency, or tendency of the kind that might be observed or measured in a population.[12] Because propensity statements are abstract, they do not report anything about what may or may not be observable in specific populations; what a propensity statement describes is more like a 'capacity' (see Cartwright 1989).[13] The statement '*A* causes *B*' describes a *propensity* for *A*-type events to produce *B*-type events, but it does not assert anything about when and where this causal relationship has been (or will be) concretely instantiated. For example, the statement 'smoking causes lung cancer' describes a propensity for smoking to cause lung cancer: it asserts that there is a causal relationship between smoking and lung cancer, such that if this relationship unfolds without interference, an instance of smoking will produce an instance of lung cancer, but it does not assert anything about whether any specific instances of smoking have produced (or will produce) lung cancer. The statement 'revolution causes war'

[11] Patrick Jackson comes close to acknowledging this point when he observes that 'the generality of a causal claim' (which is his label for what we term an abstract causal statement) is an '*ideal-typical*', '*logical*', or '*formal*' quality and not a matter of mere 'empirical regularity' (2017: 704, emphasis original). However, this is somewhat undermined by his continued use of the term 'general' to describe such claims. See also Jackson (forthcoming).

[12] As we saw in Chapter 2, Bhaskar employs the term 'tendency' to mean something similar to what we mean by 'propensity'. In our view, however, the term 'tendency' is more naturally used to describe an observable pattern.

[13] Whereas Cartwright asserts that things *possess* capacities, we refrain from this kind of metaphysical assertion, which is not required for causal inquiry in IR (see Chapter 3). We are concerned with what causal statements describe and assert, not with whether there really are causal propensities in the world existing independently of human thought.

asserts that there is a causal relation between revolutions and wars, such that if this relation unfolds without interference a given revolution will produce a war, but it gives no indication of the frequency with which actual revolutions, occurring under varying conditions, have in fact produced or will produce wars.

Let us summarize our argument so far. We contend that there are two basic kinds of causal statement, which it is crucial to distinguish, but we have shown that the terms in which this distinction is most commonly drawn—as a distinction between 'singular' and 'general' statements—are misleading. We have suggested that it would be much clearer to distinguish the two basic kinds of causal statement as 'concrete' and 'abstract', where concrete statements refer to specific, actual events, located in space and time, and abstract statements refer, generically, to events of particular types. We have argued, moreover, that what a concrete causal statement describes is the instantiation of a given causal relation in a particular time and place, linking two or more specific events. By contrast, what an abstract causal statement describes is a causal propensity. This is laid out in Table 4.2.

Table 4.2 Concrete vs. abstract causal statements

	Concrete	**Abstract**
Paradigmatic form	'*a* caused *b*'[a]	'*A* causes *B*'
Examples	'The short circuit caused the fire' 'David's smoking caused his lung cancer' 'The French revolution caused the French Revolutionary Wars'	'Short circuits cause fires' 'Smoking causes lung cancer' 'Revolutions cause wars'
What it refers to	Specific, actual events (*a* and *b*), located in space and time	Events of particular types (*A* and *B*)—that is, events of the types referred to in statements about specific events
What it describes/asserts	The instantiation of a causal relation in a particular time and place, linking two or more specific events (*a* and *b*)	A propensity for *A*-type events to produce *B*-type events, such that, if this relation unfolds unhindered, an event of one type (an *A*-type event) will produce an event of another type (a *B*-type event)

[a] As we will see shortly, concrete causal statements can also take other forms, such as '*a* will cause *b*'; this statement asserts that a given causal relation *will be* instantiated in a given time and place.

Much of the confusion which we will later go on to identify in contemporary thinking about causal inquiry in IR stems from a failure clearly to distinguish these two kinds of causal statement and to appreciate how they differ. In particular, it stems from a failure to recognize that what we have termed 'abstract' causal statements are propensity statements and not generalizations. As we will show in Chapter 5, the issue is not just that abstract statements are often mistaken for generalizations, but also that the logic of causal inquiry hinges on the relationship between concrete causal statements which do refer to specific events and abstract statements which do not. However, that discussion will have to wait a while longer. Our priority at this stage is to flesh out the distinction between abstract and concrete causal statements more fully by considering the linguistic forms they can take and the roles they play in the context of causal inquiry.

The linguistic forms 'abstract' and 'concrete' causal statements can take

So far, we have said that the paradigmatic forms of abstract and concrete statements are '*A* causes *B*' and '*a* caused *b*', respectively. The examples we have considered have also taken this form. It is important to recognize, however, that statements of each kind can (certainly in English) take a variety of linguistic forms. One reason for this is that many causal statements do not involve the verb 'to cause'. Another reason is that the English language, like many others, permits some grammatical flexibility in how particular claims are expressed. A third reason is that the specific events to which concrete causal statements refer can be events in the past, present, or future; they can also be individual events, aggregations of individual events, or emergent events.

Let us start by considering the verbs which, if present, make a statement a *causal* statement. We have argued that causation is a relation of production. To say 'the assassination of Archduke Franz Ferdinand caused the outbreak of World War I' is to say that the Archduke's assassination contributed to the production, or bringing about, of the war. As this rephrasing illustrates, however, a causal statement need not contain the verb 'to cause': *any* statement that asserts, describes, or denies a relation of production is a causal statement. Some such statements may involve near synonyms for the verb 'to cause': Kurki argues that 'words such as "because", "leads to", "produces", "makes", "enables" and "constrains"' are all instances of 'causal terminology' (2006: 197).[14] An example of a causal statement employing such a term is 'a

[14] Some of these terms, such as 'because' and 'leads to', may only describe relations of production when employed in certain contexts.

rise in the temperature of a piece of metal *makes it* expand' (Mackie 1965: 253; italics added). Some other statements which describe causal relations do not even employ a near synonym for the verb 'to cause', but rather employ transitive verbs which describe specific relations of production. Examples of such statements include 'moisture rusts steel' (Mackie 1965: 253) and 'aspirin relieves headache' (Cartwright 1983: 23; see also Anscombe 1975: 68; Sayer 1992: 104).[15]

Whether or not the verb 'to cause' is present, a statement is an abstract causal statement if it describes a relation of production but does not relate to any specific events, located in space and time. Two tell-tale linguistic indicators can help to confirm this. First, because abstract statements do not refer to concrete events, they will rarely, if ever, feature terms which single out specific events—terms such as proper nouns and the definite article 'the'. Second, abstract causal statements will almost always be expressed in the present simple tense, as in 'short circuits cause fires'. This may seem surprising, but in English the present simple tense, as in 'she speaks Italian', typically describes a subject's capacity, or propensity, as distinct from their current action, which is more often expressed in the present continuous tense, as in 'she is speaking Italian'. In the context of causal statements, this generic feature of how the present simple tense is used in English illuminates why abstract causal statements are almost always expressed in this tense: they do *not* describe what is happening (or has happened or will happen) in a specific time and place, but rather describe propensities.[16]

Although they are typically expressed in the present simple tense, abstract causal statements may often be phrased in one of two ways: either 'revolution causes war' or 'revolutions cause wars'; either '*A* causes *B*' or '*A*-type events cause *B*-type events'.[17] There is no difference in meaning attached to this difference in phrasing: in each case, the statement abstracts from specific events, located in space and time. However, because abstract causal statements are often expressed using plural event terms, as in 'short circuits [plural] cause fires [plural]' or 'revolutions [plural] cause wars [plural]', it is crucial to avoid carelessly assuming that a statement containing plural terms must be a generalization: this is definitely not the case. A causal generalization, such as 'modern revolutions have often caused wars', generalizes across

[15] Although these examples are abstract statements, concrete causal statements, too, can employ such terms: for example, 'the rise in temperature made the metal expand', 'the moisture rusted the steel', and 'the aspirin relieved Nancy's headache'.

[16] Occasionally, the future simple tense might also be used. For example, the statement 'revolutions will produce wars' abstracts from concrete events. Its meaning is therefore essentially equivalent to 'revolutions produce wars'. Certainly, both of these statements are quite different from the statement 'this revolution will cause a war', which is a concrete statement referring to a specific event ('this revolution') located in time and space.

[17] Note, however, that not all abstract statements can be reformulated in this way, at least not without sounding unnatural: consider, for example, 'short circuit causes fire'.

specific events, located in space and time (in this case: modern revolutions and the wars to which they gave rise). A causal generalization is therefore a *concrete* causal statement! By contrast, abstract statements may contain plural terms, but they do not refer to specific events.

Whether or not the verb 'to cause' is present, a statement is a concrete causal statement if it describes a relation of production which links specific events, located in space and time. One linguistic indictor that a causal statement is concrete is that it features terms which single out specific events. These include pronouns (as in '*his* carelessness caused our accident'), proper nouns (such as 'Archduke Franz Ferdinand' and 'World War I'), and the definite article 'the' (as in '*the* short circuit caused a fire'). Another indicator is that such statements will typically *not* be stated in the present simple tense.

However, although we have said that the paradigmatic form of a concrete causal statement is '*a* caused *b*'—that is, a statement in the past simple tense, as in 'the short circuit caused the fire'—the following statements are also concrete:

- 'by the time they returned, the short circuit *had caused* a fire' (past perfect tense)
- 'that short circuit *is causing* a fire' (present continuous tense)
- 'that short circuit *will cause* a fire' (future simple tense)
- 'by the time they realize the risk, that short circuit *will have caused* a fire' (future perfect tense)

Admittedly, some of these statements may sound more 'natural' than others, but they illustrate that the past tense is not the only tense that concrete causal statements can take. Indeed, it is crucial to note that concrete causal statements can, potentially, also be phrased in the present continuous tense or in the future tense: the specific events to which such statements relate can be events in the past, which have already occurred, events in the present, which have begun but have not yet ended, or events in the future, which have not yet occurred.

One further layer of complexity is that the specific events to which concrete statements refer can be 'individual' events, multiple individual events aggregated together, or what are often termed 'emergent' events. We will explore these categories further in Chapter 6, so for now let us simply illustrate some of the variety this generates in the linguistic forms that concrete causal statements can take:

- 'David's smoking caused his lung cancer' asserts a causal connection between two 'individual' events

- 'the French revolution caused the French revolutionary wars' asserts a causal connection between one individual event (the French revolution) and multiple events aggregated together (the French revolutionary wars)
- 'modern revolutions have often caused wars' asserts that a specific causal connection has often been found to link events of these two kinds (modern revolutions and the wars these revolutions are said to have caused)
- 'the widespread belief … caused the reduced birth rate' asserts a causal connection between two events (the widespread belief and the reduced birth rate) that existed at the population level and which emerged out of, yet are not reducible to, individual-level events

What makes each of these statements *concrete* is that they describe causal relations linking specific events, located in space and time.

In concluding, it is important to note that while these typical linguistic features of abstract and concrete causal statements are a useful guide, some interpretation may, nonetheless, be required when considering what kind of causal statement we are dealing with. For example, suppose a researcher were to assert that 'in this sample, we found that revolutions cause wars'. The phrase 'revolutions cause wars' might appear to suggest that the researcher is advancing an abstract causal claim, but in fact what they appear to be reporting is the discovery of a causal relation linking multiple events which are each located in space and time, but which have been aggregated together for the purpose of generalization. It would therefore be clearer for the researcher to have said 'in this sample, we found that (some, many, or all) revolutions *caused* wars'. This would help to clarify that what is being asserted is a generalization about the frequency with which a causal relation of interest was instantiated within a given sample of cases. As we shall see when we consider some actual examples of causal inquiries in IR in Chapter 8, distinguishing carefully between different kinds of causal claims and appreciating how they differ in what they assert is crucial in order to be able to evaluate and describe accurately what causal investigations have revealed.

The epistemic functions of causal statements

As the final instalment in our analysis of how the two basic kinds of causal statements differ in what they assert, let us consider the distinctive epistemic functions that statements of these kinds perform in the context of causal inquiry.

Julian Reiss identifies three key aims of social science in relation to which causal statements are advanced: 'predicting events of interest, explaining individual events or general phenomena, and controlling outcomes for policy' (2009: 34).[18] As we shall see, concrete and abstract statements play distinctive roles in relation to these aims. These roles are, however, hard to grasp if the two basic kinds of causal statement are labelled 'singular' and 'general', for those labels appear to suggest, misleadingly, that the distinctive function of each kind of statement is to convey information about the causal relations that hold, respectively, between individual or multiple events. In fact, their distinctive functions are much more interesting and important than this.

We have argued that what distinguishes a concrete causal statement is that it relates to specific events, located in time and space; moreover, as we have just pointed out, these can be events in the past, present, or future. This suggests that concrete causal statements perform two important epistemic functions: they express causal explanations and they express causally informed predictions. Let us elaborate.

A concrete causal statement, expressed in the past tense, for example '*a* caused *b*', is the paradigm of a causal explanation: it asserts a causal connection between two (or more) specific events which have already occurred.[19] A statement of this kind is prima facie explanatory because it provides information about an event, *a*, which contributed to the bringing about of another event, *b*.[20] Our point is not that the *only* function of a statement such as '*a* caused *b*' is to explain *b*, but rather that one of the central aims of causal inquiry is to generate causal explanations and that the paradigmatic form of a causal explanation is a concrete causal statement, expressed in the past tense, linking two or more specific events, both of which have already occurred.[21] A statement such as '*a* is causing *b*', expressed in the present continuous tense, arguably also expresses a causal explanation: it identifies *a* (which has begun but not yet ended) as having brought about *b* (which has also begun but not yet ended). However, we will set this detail to one side; it would only complicate our analysis, without adding anything to it, to repeatedly point it out.

[18] Others unpack this slightly differently. For example, Russo and Williamson (2007: 157) argue that causal inquiry in the health sciences has two principal goals—a cognitive goal, which encompasses explanation, and an action-oriented goal, which encompasses policy and treatment decisions. In effect, they subsume prediction and prescription together as action-oriented goals.

[19] In practice, of course, causal explanations may refer to multiple events, aggregated in various ways, and also to extended sequences of events. Moreover, they may not involve the verb 'to cause'. Consequently, causal explanations may take a variety of more or less complex forms. We discuss this further in Chapter 6.

[20] Although such a statement is prima facie explanatory, it is a further question whether it is *satisfactory* as an explanation. Note, moreover, that whereas the term 'explanatory' relates, here, to *causal* explanation, not all explanations are causal. We will consider these issues in more detail in Chapter 6.

[21] In order to simplify the analysis, we will, in Chapter 5, come to refer to a concrete causal statement, expressed in the past tense, the paradigm of which is '*a* caused *b*', as an 'explanatory statement'.

A statement of the form '*a* will cause *b*', expressed in the future tense, is the paradigm of a causally informed prediction: it asserts a causal connection between two (or more) specific events, at least one of which has already occurred and at least one of which is yet to occur.[22] A statement of this kind is prima facie predictive because it purports to provide information about a future event, *b*, that will occur as an effect of event *a*. Of course, not all predictions are causally informed: predictions can be derived purely from associational knowledge. For example, knowledge that, in the past, night has always followed day appears to justify the prediction 'this night will be followed by another day', even in the absence of any *causal* understanding. We will consider, in Chapter 5, the kind of knowledge which is required in order confidently to advance a causally informed prediction. For now, the key point is that one of the central aims of causal inquiry is causally informed prediction and that the paradigmatic form of a causally informed prediction is a concrete causal statement, expressed in the future tense, linking at least one specific event which has occurred to another specific event which is yet to occur.

If the distinctive epistemic function of concrete causal statements in the past and future tenses is, respectively, to express causal explanations and causally informed predictions, we contend that the distinctive epistemic function of abstract causal statements is to express causal theories. This may seem surprising: under the influence of the culture of generalization, some readers may be accustomed to thinking of theories as being 'general' in the sense that they 'generalize'. However, the notion that theories are expressed in abstract statements is less outlandish than it may sound. Consider two well-known theoretical claims which have been prominently debated in IR: Waltz's contention (1979) that bipolarity is more stable than multipolarity, and Finnemore and Sikkink's contention (1998) that normative change occurs via norm cascades. Neither of these claims is tied to any specific event or situation. Moreover, there is an important and obvious difference between the *case-specific* claim that some particular instance of normative change *occurred* via a norm cascade and the *theoretical* claim that normative change *occurs* via norm cascades: whereas the former relates to a specific, concrete instance of normative change, the latter abstracts from specific instances.[23] What these examples illustrate is that summary statements of IR theories are often abstract in just the same way that statements of the form '*A* causes *B*' are abstract. It is therefore hardly surprising that the distinctive epistemic function of an abstract causal statement is to express a (causal) theory.

[22] Technically, the first event (which has already occurred) need only have begun; it need not yet have ended. Again, we set this detail to one side.

[23] Of course, abstract theories carry implications about what will happen, in situations of particular kinds, if the theories hold. We will explore this further in the next section of this chapter and in Chapter 5.

It will rarely be the aim of causal inquiry *merely* to arrive at a statement of a causal theory. Typically, theories are considered valuable because of how they can be applied: the ultimate aim will be to draw on a causal theory to derive a prediction, to explain the outcome observed in some specific case, or to support efforts to control outcomes for policy—that is, to justify policy prescriptions. It is, nonetheless, important to recognize that the distinctive epistemic function of an abstract causal statement is to express a causal theory, for it draws attention to a crucial puzzle, viz. how a causal theory can be justified, given that it abstracts from specific events, located in space and time. In conjunction with our earlier observation that only concrete causal statements can be directly supported by empirical evidence, this puzzle is central to the basic logic of causal inquiry; we will consider it in Chapter 5. For now, the key point is that abstract statements, like concrete statements, perform distinctive epistemic functions in the context of causal inquiry—functions which are almost wholly obscured by the misleading labels 'general' and 'singular'. This is summarized in Table 4.3.

Before we examine the basic logic of causal inquiry, one major task remains: to consider the implications of causal statements. This is crucial for an understanding of what causal inquiries reveal, and hence also for an understanding of the strengths and weaknesses of the various possible methodological approaches that may be employed in the service of causal inquiry. However, the implications of causal statements are often quite badly misunderstood. Our task, in Section III, will therefore be twofold: to set out the implications of the two basic kinds of causal statement, but also to correct some common misunderstandings. This will enable us, in Chapter 5, to develop an analysis of the logic of causal inquiry which is free of the many serious misapprehensions that have all too often dogged discussions of these issues in IR.

Table 4.3 Distinctive epistemic functions of concrete and abstract causal statements

	Concrete		Abstract
Tense	Past	Future	Present
Paradigmatic form	'*a* caused *b*'	'*a* will cause *b*'	'*A* causes *B*'
Distinctive epistemic function	Expresses a causal explanation	Expresses a causally informed prediction	Expresses a causal theory
Example	'The assassination of Archduke Franz Ferdinand caused the outbreak of World War I'	'This revolution will cause a war'	'Hyperinflation causes political instability'

III. The implications of causal statements

We started our analysis of causal inquiry, in Chapter 1, by exploring the relationship between causation and regularities, or correlations. Among other things, we drew attention to the often unnoticed subtlety of Hume's thinking about the value of regularities as evidence for causal relations. We pointed out that although Hume, at one point, appears to suggest that a regular conjunction between events of two types constitutes necessary evidence for the judgement that they are related as cause and effect, he notes that a single, carefully conducted controlled experiment can be sufficient to provide knowledge of a causal relation. Importantly, he was aware of the flip side of this: a causal statement linking events of two kinds does not necessarily imply a corresponding regularity.

Hume realized that carefully controlled experiments can provide powerful evidence for causal relations to the extent that they successfully remove factors which might interfere with how a causal relation of interest unfolds. Indeed, it is precisely because the background conditions can and do shape what outcome is observed when a particular causal factor is introduced into an open system that controlled experiments are so valuable as a method of causal inquiry. But, if the outcome observed when a causal factor is introduced depends on the prevailing conditions, then it cannot be the case that causal statements imply corresponding regularities: whether or not a causal relation '*A* causes *B*' manifests as a correlation between *A*-type events and *B*-type events will depend on the conditions.

Unfortunately, although the utility of controlled experiments for causal inquiry is widely recognized, what this indicates about the implications of causal statements is not so widely understood. In fact, it is often, though mistakenly, supposed that causal statements imply corresponding regularities: that statements of the form '*A* causes *B*' and of the form '*a* caused *b*' each imply 'whenever *A*, then *B*' or, perhaps, 'whenever *A*, then *B*, with a high probability'. In order to understand the basic logic of causal inquiry, it will be essential to appreciate why this is wrong.

Correcting misunderstandings

One well-known claim to the effect that causal statements *do* imply corresponding regularities was advanced by Carl Hempel. On his account, the assertion that a particular set of circumstances 'caused a given event … implies that whenever and wherever circumstances of the kind in question occur, an event of the kind to be explained comes about'

(2001: 278; see also 1966: 53).[24] He contends, in other words, that if event *a* caused event *b* then it must be the case that *A*-type events always produce *B*-type events, generating a regular pattern 'whenever *A*, then *B*'.[25] Claims of this kind are so commonplace in IR and political science that they often go unquestioned. For example, Stephen Van Evera argues that a statement of the form '*A* causes *B*' is a causal law and that deterministic laws 'frame invariant relationships' of the form 'if *A* then always *B*' (1997: 8).[26] Similarly, a widely used research methods textbook states that if there is a causal relation between two events then 'the emergence or presence of one condition or event will always (or with high probability) bring about another' (Johnson and Reynolds 2012: 41–2).

Hempel justifies his contention that causal statements imply corresponding regularities by invoking what he terms a 'general maxim', viz. 'same cause, same effect'. As he points out, if this principle is applied to a causal statement of the form '*f* caused *g*', then this 'yields the implied claim that whenever an event of kind *F* occurs, it is accompanied by an event of kind *G*' (1966: 53; see also 2001: 278). As a matter of logic, this is true, but the more important point is that Hempel's supposed 'maxim' is nothing of the kind! To see this, consider a well-known example. Most readers will have been taught in school that the Earth's gravity causes a falling object near the Earth's surface to accelerate downwards towards the centre of the Earth at approx. 9.8m/s^2. However, as we briefly noted in the Introduction to this book, this does *not* imply that *whenever* an object falls towards the Earth it will accelerate downwards at approx. 9.8m/s^2, for the behaviour of falling objects is affected not only by the Earth's gravity, but also by other factors, such as atmospheric resistance, which can and do interfere with the effects of gravity. In fact, it is only under special conditions, such as in a vacuum chamber (where the interference of factors such as atmospheric resistance is removed), that *any* object falling towards the Earth will accelerate downwards at approx. 9.8m/s^2. The outcome observed when a causal factor is introduced will depend on the

[24] In an earlier formulation, Hempel had argued that a statement that one event caused another event 'amounts to', rather than merely 'implies', a statement that events of the first kind are 'regularly accompanied' by events of the second kind (1942: 36). This is closely related to what we termed in Chapter 1 the 'semantic' version of the regularity view of causation.

[25] An important corollary of this is that if a statement of the form '*a* caused *b*' (or similar) is advanced as a causal explanation, then it can be tested by checking whether there is in fact a regular pattern 'whenever *A*, then *B*'. Showing that scientific explanations, expressed in causal language, can be tested in this way is a key aim of Hempel's 'covering-law model' of explanation, which we will examine in Chapter 6.

[26] Van Evera adds that a probabilistic causal law would imply 'if *A*, then sometimes *B*, with probability *X*' (see 1997: 8). On his account, therefore, deterministic causal laws imply exceptionless regularities, whereas probabilistic causal laws imply correlations. As we will go on to point out, however, the statement '*A* causes *B*' does not even imply that there will be a correlation between the occurrence of *A*-type and *B*-type events.

prevailing conditions; it is therefore highly misleading to present 'same cause, same effect', as a 'general maxim'.

This simple fact about the behaviour of falling objects reflects an important underlying feature of scientific laws. Hempel defined 'general laws' as 'statements of universal form' which assert 'a uniform connection between different empirical phenomena', such that 'whenever and wherever conditions of a specific kind *F* occur, then so will, always and without exception, certain conditions of another kind, *G*' (1966: 54; see also 1942: 35; 1966: 66). As Nancy Cartwright has shown, however, Hempel's understanding of scientific laws cannot be sustained; in fact, she argues, 'the laws of physics lie' (1983). What she means by this is that 'fundamental' (or 'theoretical') laws will often fail to offer accurate descriptions of 'what happens in concrete situations'. This is because they are 'abstract formulae which describe no particular circumstances'; they 'are designed only to tell truly what happens in each domain separately'—that is, if the relations they describe unfold without interference (1983: 12–13). This is why, as we discussed in Chapter 2, scientific experiments are often conducted under artificially controlled conditions: in order to discover scientific laws, it is often necessary to isolate the relation of interest from interference.

For our purposes, what is most valuable about Cartwright's analysis of scientific laws is her acknowledgement that, in open systems, in which interferences are not prevented, even the most fundamental laws of physics may not offer a reliable guide as to the outcome that will be observed. As Cartwright puts it: 'If we follow out their consequences [in open systems], we generally find that the fundamental laws go wrong; they are put right by the judicious corrections of the applied physicist or the research engineer' (1983: 13).

This critical difference between open and closed systems, which we emphasized in Chapter 2, is as important for understanding causal statements as it is for understanding fundamental scientific laws. Scientific laws do not imply exceptionless regularities; rather, they imply that patterns conforming to the laws should be observable under controlled conditions. Similarly, but contrary to what Hempel supposed, the statement '*A* causes *B*' does *not* imply 'whenever *A*, then *B*'; rather, it implies something more qualified: that an *A*-type event will produce a *B*-type event if this causal relation unfolds without interference. This crucial qualification is often overlooked, but Elizabeth Anscombe made the point more than fifty years ago. She argued that 'if you take a case of cause and effect, and relevantly describe the cause *A* and the effect *B*, and then construct a universal proposition, "Always, given an *A*, a *B* follows" you usually won't get anything true'. As she went on to

explain, a universal proposition of this type will be true only in the absence of interference (1975: 69; see also Mackie 1980: 77). When other factors can potentially interfere with the unfolding of this causal process, then it may be the case *both* that '*A* causes *B*' *and* that not all *A*-type events are followed by *B*-type events. As Cartwright points out, for example, it is 'generically true that aspirin relieves headache even though some particular aspirins fail to do so' (1983: 23).[27]

In fact, as Cartwright points out, the statement '*A* causes *B*' does not even imply that there should be an observable *correlation* between *A*-type events and *B*-type events, at least not in open systems. She offers an example which proceeds roughly as follows. Suppose that 'smoking causes heart disease' but that '[e]xercise prevents heart disease'. Suppose, further, that 'exercise is more effective at preventing heart disease than smoking [is] at causing it' and that, in the population being studied, smoking and exercising are highly correlated. In such a situation, even though 'smoking causes heart disease', no positive correlation between them may be observable; there may even be a negative correlation! A positive correlation between smoking and heart disease is guaranteed to appear only if the interfering effects of exercise are controlled for. (One simple way to do this would to be to segment the population into exercisers and non-exercisers and consider the correlation between smoking and heart disease within each group.) To express the point more broadly: causal relations are *guaranteed* to manifest as observable correlations only in 'situations in which all other causal factors are held fixed' (Cartwright 1983: 23–5).

To most quantitatively trained social scientists, this will be a very familiar point, but its implications for how we interpret causal statements have not always been quite so fully appreciated in IR. To give just one example: Van Evera argues that if 'democracy causes peace' this should show up as a correlation: it should be possible to observe that 'democratic states are involved in fewer wars than authoritarian states' (1997: 36). This is not right, at least not if it is assumed, as we recommended in Chapter 2 that it should be, that the world being observed is likely an open system with respect to this causal process. Even if it is the case that 'democracy causes peace', in an open system democracy may also be correlated with other factors which cause war; if so, then democracies could potentially be involved in more wars than non-democracies *even though* democracy causes peace (see, for example, Risse-Kappen 1995; Müller 2004; Geis, Brock, and Müller 2006)!

[27] Cartwright might have said, more accurately, that it is generically true that taking aspirin relieves headache even though, *under some conditions*, taking an aspirin may fail to relieve a headache.

The true implications of concrete and abstract causal statements

What, then, *are* the implications of causal statements? If a causal statement is accepted, what else can be inferred from it? We contend that concrete and abstract causal statements each have one core implication which is central to the logic of causal inquiry.

Concrete causal statements assert a causal connection between specific events, located in space and time. Crucially, they also *imply* corresponding abstract statements: '*a* caused *b*' *implies* '*A* causes *B*'.[28] This should be immediately clear when we realize, for instance, that if this short circuit caused that fire, then it *must* be the case that there are *some conditions* under which short circuits produce fires. To put this differently: if this short circuit caused that fire, then there must be a *propensity* for short circuits to produce fires. If there were no propensity for short circuits to cause fires—if there were no conditions under which a short circuit would produce a fire—then it simply could not be the case that this short circuit caused that fire.[29] What this reveals is that if the statement '*a* caused *b*' is accepted, then, *logically*, the statement '*A* causes *B*' must also be accepted. In order to grasp this, it is crucial to have recognized, as we argued earlier, that statements of the form '*A* causes *B*' are propensity statements and not generalizations; once this is recognized, then it should immediately be clear that there is nothing problematic about the suggestion that '*a* caused *b*', a statement about a single historical episode, implies '*A* causes *B*', a statement with a much broader applicability.[30]

As we shall see in Chapter 5, the reason it is important to appreciate that concrete causal statements imply corresponding abstract statements is that if evidence can be provided in support of a concrete causal statement, then this evidence will also, indirectly, support the corresponding abstract statement. To put this differently: if evidence can be provided in support of a causal statement relating to specific events, located in space and time, then it will also, indirectly, support a corresponding causal theory. For example, if evidence can be found which demonstrates that 'this short circuit caused that fire', then this evidence also, indirectly, supports the causal theory 'short circuits cause fires'. This is central to the logic of causal inquiry, for it indicates

[28] Recall that the reverse is not true: an abstract causal statement does not imply any particular concrete statement; rather, it abstracts from the specific events to which concrete statements refer.

[29] Note that what is at stake here is not the truth of an individual causal statement, but rather its logical implications. Whether one specific event, *a*, in fact caused another specific event, *b*, will have to be determined on the basis of the available evidence. We will consider, in Chapter 5, how this may be done.

[30] We suspect that those who dismiss the value of single case studies for causal inquiry may have failed to recognize this!

how empirical support can be provided for causal theories even though they abstract from specific events.

With this in mind, let us now consider the implication of abstract causal statements. The crucial implication of the statement '*A* causes *B*' is that an *A*-type event will produce a *B*-type event *if the causal process of interest unfolds without interference*. In other words, the abstract causal statement '*A* causes *B*' implies the conditional statement 'an *A*-type event will produce a *B*-type if the conditions are right', the 'right' conditions being those conditions under which this process unfolds without interference. For example, the statement 'revolution causes war' implies a corresponding conditional statement: 'a revolution will produce a war if the conditions are right'—that is, if the conditions are such that the causal relationship between revolution and war unfolds without interference. Note the significant difference between this conditional statement and the regularity statement 'whenever a revolution, then a war'. It is crucial to understand that even if there is a propensity for revolutions to produce wars, the consequences of any specific revolution will depend on whether the prevailing conditions allow this propensity to unfold without interference.

For any readers who are used to supposing that causal statements imply regularities or correlations, this implication may appear disappointingly weak. It is, however, a corollary of the methodological assumption which we recommended in Chapter 2 as appropriate for most causal inquiries in IR: that unless there is a persuasive reason to suppose otherwise, it should be assumed that the system being investigated is open with respect to the causal process of interest. If we may be dealing with an open system, in which potential interference with the unfolding of a causal process of interest has not been fully prevented, then this process cannot be expected to unfold without interference and hence cannot be expected to manifest as a regularity.

In any case, the significance of the fact that the abstract statement '*A* causes *B*' implies a conditional statement 'an *A*-type event will produce a *B*-type event if the conditions are right' should not be underestimated, for conditional statements of this kind can, as we shall see, be drawn on to support causal explanations, causally informed predictions, and policy prescriptions, thereby fulfilling some of the core goals of causal inquiry. We will explore this further in the next chapter.

Conclusion

Our analysis in this chapter has put us in a position to show how the basic logic of causal inquiry emerges out of the relationship between concrete and

abstract causal statements. Before we elaborate that logic further in Chapter 5, let us conclude this chapter by offering a brief summary of what we have argued and noting one dimension of our topic which we have not covered, but which we address in the Appendix.

We started by noting the ubiquity of what we termed the 'culture of generalization' in IR and political science and by pointing out the confusion which is liable to be generated by the tendency to label the two basic kinds of causal statement 'general' and 'singular'. We suggested alternative labels which we consider clearer and more precise: we termed causal statements which refer to specific events, located in space and time, 'concrete', and termed statements which abstract from such events 'abstract'. We pointed out that the key difference between statements of these two kinds is not the number of events to which they relate, but rather the categorical difference in what they assert: concrete statements describe the instantiation of a given causal relation in a particular time and place, whereas abstract statements describe causal propensities.

Having further clarified the variety of linguistic forms in which these two kinds of statements can arise and the functions they perform in the context of causal inquiry, we then explored their implications. Correcting the common, but mistaken, supposition that causal statements imply corresponding regularity statements, we identified a core implication of each kind of causal statement: concrete statements imply corresponding abstract statements, while an abstract statement of the form '*A* causes *B*' implies that an *A*-type event will produce a *B*-type event if the conditions are right. As we hinted, these implications are central to the logic of causal inquiry.

Before we move on to present that logic, it is worth briefly noting one topic that we have not discussed: the relationship between causal statements and statements of necessary and/or sufficient conditions. Some readers may wonder why we have neglected this topic, given that the vocabulary of necessary and sufficient conditions often crops up in the context of causal inquiry. We have refrained from getting drawn into this discussion for two main reasons. First, developing knowledge of necessary and/or sufficient conditions is far harder than is commonly acknowledged; indeed, it is probably harder than generating causal knowledge. Second, causal statements cannot straightforwardly be analysed as statements of necessary and/or sufficient conditions, for causal statements have a direction, or asymmetry, which statements of necessary and/or sufficient conditions lack. In our view, although it is possible to draw some systematic conclusions about what causal statements imply about necessary and/or sufficient conditions, doing so is complex and does not offer much of value to those conducting causal inquiries in IR.

Despite our scepticism about whether it is worthwhile analysing the relationship between causal statements and statements of necessary and/or sufficient conditions, we acknowledge that this relationship is often considered to be an important one. We have therefore defended our own view in more detail in the Appendix. Interested readers may wish to turn to that now. Other readers can simply press on to Chapter 5, where we will show how the logic of causal inquiry emerges out of the relationship between concrete and abstract causal statements.

5
The Logic of Causal Inquiry

In Chapter 4 we distinguished between two basic kinds of causal statement: concrete statements, which relate to specific events, located in space and time, and abstract statements, which abstract from specific events to describe causal propensities. We argued that understanding this distinction accurately is important in its own right—because it facilitates a correct understanding of what is asserted in and implied by statements of these two kinds, which are routinely advanced in IR—and also because it underpins the logic of causal inquiry. That logic is the focus of this chapter. We will contend that there is a 'deep logic' to causal inquiry—a logic that is independent of the purpose of any particular causal inquiry, of its subject matter, and of the methods of data collection and analysis that it employs—and that this logic emerges out of the relationship between concrete and abstract causal statements. In subsequent chapters, we will show how this logic shed lights on other topics which are of central importance for causal inquiry, including causal explanation (Chapter 6) and causal reasoning (Chapter 7). We will also demonstrate how it can illuminate substantive examples of causal inquiry in IR (Chapter 8).

As we conceive of it, the logic of causal inquiry concerns how empirical evidence can support causal claims. This logic is important because whatever the immediate purpose of causal inquiry may be—whether it is to generate knowledge of causal theories, or, say, to advance causal explanations, predictions, or policy prescriptions—an underlying aim is to advance causal statements as knowledge claims. Yet, no causal statement can be advanced as a knowledge claim unless it is adequately supported by empirical evidence.[1] It is therefore essential for those engaged in causal inquiry to consider how empirical evidence can support different kinds of causal statement.[2]

There is, of course, a substantial existing literature on what is often called 'causal inference'—that is, inferences, drawn from analysis of empirical evidence, to the effect that there is a causal relationship between events of two

[1] We assume that causal knowledge cannot be developed through a priori reasoning alone.

[2] Although we prefer to remain agnostic about whether causation is a part of the world existing independently of human thought, we take seriously the idea that causal statements are knowledge claims which can be justified by reference to empirical evidence. In this sense our position differs significantly from Lebow's contention (2020: 310) that causal claims are 'rhetorical at best'.

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0006

(or more) kinds and/or that such a relationship has manifested in a particular setting. However, our approach will differ from that which is typically employed in this literature.

One way in which our approach will differ relates to our terminology. Under the influence of the culture of generalization, existing discussions of how empirical evidence can support causal claims often use key terms such as 'theory' and 'explanation' in ways that we consider imprecise and misleading. In particular, causal theories are often construed as being 'general', while causal explanations are construed as 'singular', 'specific', or 'particular' (see, for example, King, Keohane, and Verba 1994: 9–12, 35–49; Van Evera 1997: 7–8, 15–16, 40). This may appear natural, but we pointed out in Chapter 4 that causal theories are expressed in abstract statements (paradigmatically: '*A* causes *B*'). These are not generalizations, but rather describe causal propensities. Meanwhile, causal explanations are expressed in concrete statements (paradigmatically: '*a* caused *b*'). These are not intrinsically 'singular', but are rather distinguished by the fact that they relate to specific events, located in space and time. We endorse the common supposition that a key aim of causal inquiries is often to provide support for causal theories which can then be applied to generate explanations and predictions (see, for example, King, Keohane, and Verba 1994: 19–20; Van Evera 1997: 4), but we will show that providing empirical support for causal theories is much less straightforward than is commonly supposed, for empirical evidence relates to specific events, located in space and time, whereas causal theories abstract from such events to describe causal propensities.

A second way in which our approach will differ from that adopted in existing discussions is that our analysis does not privilege any particular method of data collection or analysis. Existing work tends to adopt a particular account of causation and to show how a particular approach can be used to generate empirical support for causal claims framed in accordance with that account. For example, the so-called 'potential outcomes' framework (also known as the Neyman–Rubin causal model) defines a causal effect in counterfactual terms and shows how statistical analysis can be used, in experimental and observational settings, to estimate the counterfactual effect of a treatment (or causal factor) in such settings (see, for example, Rubin 1974, Holland 1986).[3] This is a powerful framework which has been drawn upon to generate

[3] As we saw in the Introduction, an alternative to a counterfactual account of causation might be a regularity account (see Chapter 1), an account of causation as rooted in powers, capacities, or mechanisms (see Chapter 2), or an account of causation as a kind of probabilistic dependency (for an accessible overview, see Reiss 2009: 21–6; for more detail, see Beebee, Hitchcock, and Menzies 2009a, parts II and III).

some robust and important findings in IR.[4] However, its application is closely associated with the use of statistical methods: it does little to illuminate other ways in which causal inferences might also be derived (see Brady and Collier 2010). We believe that, whatever the strengths and weaknesses of particular methods of data collection and analysis may be, all efforts to show how empirical evidence can support causal claims involve a common, underlying logic which is rarely unpacked—a logic which is *not* tied to, and does not favour, any particular method.

Our account of this logic will build on the preliminary analysis of the two basic kinds of causal statement which we provided in Chapter 4. In the first section of this chapter, we will show that the only kind of causal statement which can be directly supported by empirical evidence is what we will term an 'explanatory statement'–that is, a concrete statement relating to events that have already occurred, the distinctive epistemic function of which is to express a causal explanation. Causal theories, which are expressed in abstract statements, can only be indirectly supported, by providing evidence that persuasively supports explanatory statements from which they can be inferred. In Section II, we will argue that the kind of evidence which will persuasively support an explanatory statement is evidence which rules out competing explanatory statements. We will show, moreover, that this is true irrespective of what methods of data collection and analysis are employed. In Section III, we will consider how causal theories which have been indirectly supported can then be applied to generate causal explanations, causally informed predictions, and causally informed policy prescriptions. Section IV will draw these threads together and summarize what we are describing as the 'deep logic' of causal inquiry.

I. What kind of causal statement can be empirically supported?

Searching for empirical evidence which supports or undermines causal propositions is the central activity of causal inquiry. This activity cannot take place in a conceptual and theoretical vacuum: defining concepts, delimiting the 'events' which are postulated as 'causes' or 'effects', and formulating candidate causal propositions are also crucial ingredients of causal inquiry.[5] Yet,

[4] In Chapter 8, we examine in detail two examples of research informed by this underlying framework: one experimental study (Fair et al. 2018) and one observational study (Lyall 2009).

[5] We acknowledge, moreover, that empirical evidence is always to some extent 'theory-laden' (see Hollis and Smith 1991a: 55–7). It is an open question how much of a problem this may be (see Boyd and Bogen

whether the ultimate goal is explanation, prediction, policy prescription, or just the generation of knowledge for its own sake, causal propositions can only be advanced with confidence when they are backed by evidence.

It is therefore necessary to consider what kind of causal statement *can* be supported by empirical evidence. Here, the distinction we drew in Chapter 4 between concrete and abstract causal statements is crucial, for it is only concrete statements, and not abstract statements, which can be directly supported (or undermined) by empirical evidence. This is the basic insight from which, we will show, the logic of causal inquiry emerges.

Why abstract statements cannot be directly supported by empirical evidence

By definition, empirical evidence relates to actual objects, events, or states of affairs which have occurred or obtained. Such evidence is potentially relevant to causal inquiry if it relates to events which could be linked together as cause or effect.[6] These events may be large or small in scale and they may have occurred anywhere in time from the present to the distant past. They must, however, be actual events—events which are concretely located in time and space. This is true irrespective of what methods of data collection and analysis are used: whether it is a single-outcome process tracing study or a survey experiment with thousands of observations, the empirical evidence that is drawn upon in causal inquiry must relate to specific events, located in space and time. It could not be otherwise!

We emphasize this fairly obvious point in order to highlight an important disconnect between causal theories and empirical evidence. We argued, in Chapter 4, that the key epistemic function of an abstract causal statement, paradigmatically '*A* causes *B*', is to express a causal theory. As we explained, such a statement is not a generalization, but rather describes a causal propensity. This is important because whereas the empirical evidence which is sought in the course of causal inquiry necessarily relates to specific events, located in space and time, causal theories do *not* relate to specific events. It is fairly easy to see how empirical evidence can speak to concrete statements such as 'David's smoking caused his lung cancer' or 'modern revolutions have

2021), but if it is a problem, it is likely to be equally problematic for any account of how empirical evidence can support or undermine causal propositions, so we set it to one side.

[6] As a reminder, we construe 'events' very broadly to include not only discrete occurrences but also shorter- and longer-term states of affairs, such as the fleeting possession of an intention or the endurance of a deeply established social norm or structure.

often caused wars': because these statements refer to specific events, empirical evidence can, among other things, establish whether these events in fact occurred. By contrast, it is much less obvious how empirical evidence can speak to abstract statements such as 'smoking causes lung cancer' and 'revolution causes war', for these statements *do not* refer to specific events about which evidence could be gathered, but rather abstract from such events to describe causal propensities.

One way in which this disconnect between empirical evidence and abstract causal statements manifests practically is that causal theories, which are expressed in abstract statements, cannot be falsified. To see the basic point, suppose that a particular short circuit does not lead to a fire. This will refute the concrete statement 'that short circuit caused a fire', but it will not refute the theory 'short circuits cause fires'. In fact, this theory will not be refuted even if a huge number of short circuits do not lead to fires. As we pointed out in Chapter 4, there may be a propensity for short circuits to cause fires even though the vast majority of actual short circuits have not in fact produced fires. This is a familiar feature of how well-understood causal relationships play out in everyday life, but its implications for causal inquiry are rarely noted: all too often, it is casually supposed that a principal task of causal inquiry must be to seek empirical evidence which directly supports or refutes causal theories (see, for example, King, Keohane, and Verba 1994: 100–5; Van Evera 1997: 20, 27–30).

Despite this disconnect between causal theories and empirical evidence, causal theories can become well supported. To see how, it is necessary to recognize that they are expressed in abstract causal statements and also to understand the relationship between abstract and concrete causal statements which we pointed out in Chapter 4: the fact that a concrete statement, which relates to specific events, *implies* a corresponding abstract statement. For example, the concrete statement 'this short circuit caused a fire', which relates to a specific short circuit and a specific fire, *implies* the abstract statement 'short circuits cause fires'.[7] This implication is important, for concrete causal statements, which relate to actual events, clearly *can* be supported or undermined by empirical evidence. Hence, even though causal theories cannot be *directly* supported by empirical evidence, they can be *indirectly* supported: empirical evidence *indirectly* supports an abstract statement, expressing a causal theory, if (and only if) it supports a concrete statement from which this abstract statement can be inferred. If evidence can be found which

[7] As we explained, if this short circuit caused a fire, then there must be a propensity for short circuits to cause fires; if there were no such propensity (if there were no conditions under which a short circuit will produce a fire), then this short circuit could not have caused a fire.

persuasively supports the statement 'this short circuit caused a fire', then the theory 'short circuits cause fires' will also, thereby, become well supported.[8]

This suggests that there is an inherent logic to how causal knowledge must be developed. Suppose that one aim of a research project is to provide support for a particular causal theory, perhaps with a view to applying this theory to generate explanations, predictions, or policy prescriptions (we will discuss, in Section III, how this can be done). In order to provide support for this theory, it will be necessary to provide support for a concrete statement which describes how the causal relation of interest manifested in a specific time and place. If, and only if, evidence is found which persuasively supports this concrete causal statement can the theory which this statement implies be considered well supported.

In fact, it is possible to specify more precisely what kind of concrete statement must be supported: it must be what we are calling an 'explanatory statement'. Let us elaborate. We pointed out in Chapter 4 that concrete causal statements come in two main varieties: statements typically expressed in the past tense, which relate to specific events that have already occurred, and statements expressed in the future tense, which relate to events some of which are yet to occur. We pointed out, moreover, that these two varieties of concrete statement perform two distinctive epistemic functions: the former express causal explanations, while the latter express causally informed predictions.[9] In order to simplify the analysis, we will, from this point on, refer to the former—concrete statements which relate to specific events that have already occurred, the distinctive epistemic function of which is to express causal explanations—as 'explanatory statements'.[10] It is important clearly to distinguish this category of causal statement from all other types because only statements of this kind can be directly supported by empirical evidence.

The reason for this is relatively simple. An explanatory statement, paradigmatically '*a* caused *b*', relates to events which have already occurred and describes a causal relationship which has unfolded fully, bringing about an

[8] Recall, though, that the statement 'short circuits cause fires' is a propensity statement, not a generalization. Hence, if the evidence shows persuasively that 'this short circuit caused a fire', what this implies is that a short circuit will produce a fire under the right conditions; it does not imply that short circuits will often or always produce fires.

[9] As we noted in Chapter 4, statements such as 'this short circuit is causing a fire', relating to events which have begun but not yet ended, also express causal explanations. However, as we explained, we set this issue to one side. Doing so does not significantly affect our analysis.

[10] In labelling such statements 'explanatory', we are drawing attention to their distinctive epistemic function, which is to express causal explanations. We do not mean to imply that any such statement is or should be considered satisfactory when offered as an explanation: its adequacy as an explanation will depend on whether it is convincingly supported by empirical evidence and on the pragmatic relevance of the information it provides. We will consider this issue further in Chapter 6.

effect: *b*. In relation to such a statement, it is possible to seek evidence that *a* and *b* have occurred and that there was a causal relationship between them. By contrast, a concrete statement expressed in the future tense, paradigmatically '*a* will cause *b*', refers to at least one event, *b*, which has not yet occurred, and it describes a causal relationship which has not (yet) unfolded fully. Clearly, it will not be possible to provide evidence that this relationship has unfolded fully, giving rise to *b*, unless and until *b* has actually occurred. However, if and when this causal relationship does unfold fully, giving rise to *b*, the evidence will no longer speak to the statement '*a* will cause *b*'; rather, it will speak to the explanatory statement '*a* caused *b*'.[11]

Combining this analysis with our previous observation that a causal theory can only be supported indirectly gives rise to the following insight: a causal theory becomes well supported only when evidence is identified which persuasively supports an explanatory statement from which the theory can be logically inferred.

Why this matters

Some readers may accept our contentions that explanatory statements are the only kind of causal statements that can be directly supported by empirical evidence and that a causal theory can only be supported indirectly, and yet wonder why this matters. Why is it important to differentiate so sharply between the direct support that empirical evidence can provide for an explanatory statement and the indirect way in which causal theories become well supported? Surely the key point is just that empirical evidence can, in one way or another, support both kinds of statement?

The reason why it is important to be accurate about this is that, under the influence of the culture of generalization, the terms 'theory' and 'explanation' are often used in ways which suggest a seriously misleading picture of causal inquiry. We pointed out in Chapter 4 that statements of the paradigmatic form '*A* causes *B*', which express causal theories, are often, though misleadingly, labelled 'general', while statements of the paradigmatic form '*a* caused *b*', which express causal explanations, are often, though misleadingly, labelled 'singular'. In corresponding fashion, causal theories are often supposed to be 'general', whereas causal explanations are often supposed to be 'singular' or 'particular' and to relate to 'individual' events (see, for example,

[11] This is not to deny that a causally informed prediction, '*a* will cause *b*', can be generated; as we will show, however, it is generated by applying a well-supported causal theory. We will discuss this further in Section III.

King, Keohane, and Verba 1994: 9–12, 35–49; Van Evera 1997: 7–8, 15–16, 40; Gerring 2006; Johnson and Reynolds 2012: 43; Beach and Pederson 2013: 3).

Confronted with these labels, it might be natural to suppose that the principal aim of causal inquiry will ordinarily be to provide support for 'general' causal theories, the provision of support for 'singular' explanatory statements being a secondary concern. It might also appear natural to suppose that the evidence which is most valuable for the purpose of causal inquiry is evidence which supports 'general' causal theories, *rather than* evidence which supports explanatory statements relating to specific events. This is liable to lead to a mistaken supposition that it is possible to provide support for causal theories *without* providing support for explanatory statements, perhaps by providing 'general' evidence about the co-occurrence of *A*-type and *B*-type events.

Although it may seem quite intuitive, this picture of causal inquiry has things precisely the wrong way around.[12] As we have repeatedly emphasized, a causal theory is not a generalization, but rather describes a causal propensity, and hence it *cannot* be directly supported by empirical evidence, for empirical evidence necessarily relates to the specific events from which causal theories abstract. A causal theory can become well supported only when persuasive support is provided for a corresponding explanatory statement. Consequently, the evidence which is most important for causal inquiry is *not* 'general' evidence about the co-occurrence of *A*-type and *B*-type events, but rather evidence which shows that one or more specific *A*-type events led to *B*-type events, this evidence justifying the inference that there is a corresponding propensity, '*A* causes *B*', and hence establishing the theory '*A* causes *B*' as well supported.

We will consider, in Section II, what kind of evidence will persuasively support an explanatory statement. For now, the crucial point is that identifying evidence which supports a relevant explanatory statement constitutes the *means by which* support can be provided for a causal theory of interest.[13] This 'means–ends' relationship between, on the one hand, providing support for explanatory statements (the means) and, on the other, providing support for causal theories (the end), is obscured by characterizations of theories as 'general' and of explanations as 'singular' or 'particular'; consequently,

[12] As Lipton (2004: 56) puts it: a common picture of inference is 'inference first, explanation second', but this 'seriously underestimates the role of explanatory considerations in inference'. We will consider this further in Chapter 7.

[13] Clearly, providing support for a concrete causal statement may sometimes be an end in its own right, but what we wish to emphasize here is that it is a necessary part of causal inquiry even when it is not an end in its own right.

it is overlooked in most research methods texts. It is significant not only because it underpins the logic of causal inquiry, but also because it has several important methodological corollaries. Let us briefly consider these.

Methodological corollaries

Noting the emphasis we place on identifying evidence which supports explanatory statements, some readers may wonder whether our account of the logic of causal inquiry is relevant principally to efforts to generate causal explanations for particular outcomes of interest, using methods which are especially suited to this goal. That is not the case. We characterize the logic we are elaborating as a 'deep logic' precisely to emphasize that it is not tied to, and does not favour, any particular aim of causal inquiry (such as, say, explaining rather than predicting) or any particular method of data collection and analysis. Let us illustrate this by considering how support for a causal theory may be derived from a randomized controlled trial (hereafter, RCT), which is often identified as constituting the 'gold standard' among methods of causal inquiry (see Cartwright 2010: 59; Imai 2017: 49).

In order to make the key issues as comprehensible as possible, imagine that an RCT is conducted to test whether aspirin relieves headache in adults—to provide support for the causal theory 'aspirin relieves headache'. Suppose that an accurate measure of the degree of relief from headache is available and that a suitably sized test population of adults suffering from headache is randomly given either aspirin (this is the treatment group) or a placebo (this is the control group). Suppose, further, that this is an ideal experiment in which the treatment and control groups are perfectly balanced across all possible confounders (see Cartwright 2010: 64).[14] After an appropriate period of time has elapsed, the degree of relief from headache is measured, for each participant. Suppose that the average degree of relief is noticeably higher in the treatment group than in the control group. This outcome appears to provide compelling empirical support for the theory 'aspirin relieves headache': there is a noticeable difference in outcomes across the two groups, all possible confounders have been controlled for, and any problems in the running of the experiment have been avoided.[15] But why does the observed result provide support

[14] For the purpose of this discussion, we also assume that any other problems arising from the running of the experiment, such as attrition (see Shadish, Cook, and Campbell 2002: 56–62, 323–40), have been avoided.

[15] If the causal relations of interest are deterministic (eliminating chance from the equation), if the two groups are perfectly balanced across all possible confounders, and if all other problems arising from the running of the experiment have been avoided, then *any* difference in the average outcome measured

for this theory, given that 'aspirin relieves headache' is an abstract statement and that, as we have argued, there is an inherent disconnect between abstract statements and empirical evidence?

The key point to note is that an RCT consists of a sequence of events. The first event is the random assignment of test subjects to two groups and the application of the treatment to one group, but not to the other. The second event is the experimental outcome, which is the average value measured on the outcome variable in each group after an appropriate period of time has elapsed. The empirical evidence generated in an RCT pertains directly to these events. What our hypothetical RCT establishes is hence that one event, the giving of aspirin to the treatment group but not to the control group, caused another event, the higher average degree of relief measured in the treatment group as compared to the control group. In other words, this hypothetical RCT provides persuasive support for an explanatory statement, viz. 'giving aspirin to members of the treatment group (but not to members of the control group) caused them, on average, to experience a higher degree of relief from headache (as compared to members of the control group)'. From this explanatory statement, the theory 'aspirin relieves headache' can be logically inferred.

Although we will not elaborate this line of argument in relation to other methods of data collection and analysis which are also employed in the service of causal inquiry, we contend that all methods which utilize empirical evidence to generate support for causal theories must do so by providing support for explanatory statements from which the theories of interest can be inferred. How could it be otherwise? How could empirical evidence provide support for a statement concerning anything other than the actual events to which this evidence relates? Of course, researchers do not typically spell this out: it is natural to interpret our hypothetical RCT as providing persuasive support for the theory 'aspirin relieves headache' without explicitly going through a process of reasoning, first, that it provides persuasive support for the explanatory statement 'giving aspirin to members of the treatment group caused them, on average, to experience a higher degree of relief from headache' and, second, that this indirectly supports the theory 'aspirin relieves headache'. It is, however, important to be aware of the underlying logic, which holds regardless of what method of data collection and analysis is employed.

across the two groups, however small, must be due to the treatment; if such conditions are fulfilled, it is not even necessary to consider whether the difference is statistically significant. As we do not wish to get distracted by such questions, we simply assume, for the purpose of this illustration, that there is a 'noticeable' difference in the outcome measured across the two groups.

It is also important to be aware that in order for a causal theory to become well supported it is not necessary to examine a large number of cases or to consider a broad evidence base. This may appear counter-intuitive, especially to those who are accustomed to thinking of causal theories as being 'general'. The key point to note, however, is that a causal theory will become well supported when it is inferred from *just one well-supported explanatory statement*. If the evidence demonstrates persuasively that 'a short circuit caused the fire', then this establishes the theory 'short circuits cause fires' as well supported: it is not necessary to examine many short circuits and fires. Moreover, what matters is that the evidence persuasively supports the explanatory statement from which the theory is inferred, not how broad the evidence base was nor what method for gathering and analysing this evidence was used. Whether it is generated through an experiment, through process tracing in relation to a single historical event, or through the analysis of a large, multi-country dataset, if the evidence persuasively supports an explanatory statement, then it will also provide strong support for the corresponding causal theory.

Finally, it is important to note that when persuasive support is provided for an explanatory statement, '*a* caused *b*', thereby also supporting the corresponding theory, '*A* causes *B*', this only reveals one set of conditions under which the propensity described by the theory unfolds fully: the conditions under which *a* caused *b*. Again, this is true irrespective of what method is used to generate the supporting evidence. However many cases are examined, the evidence can only show that the propensity of interest unfolds fully under the conditions which have been studied; it cannot show whether the propensity of interest will unfold fully under conditions which have not been studied. It is therefore important to be aware that even when a causal theory is well supported, little may be known about the frequency with which, or the variety of conditions under which, the propensity it describes will unfold fully. We will consider this issue further in Section III of this chapter and in Chapter 7, where we discuss 'external validity'. At this point, however, we will turn to a central question raised by our analysis so far: what evidence will persuasively support an explanatory statement?

II. What evidence will persuasively support an explanatory statement?

It is widely known that correlation is not causation. In an open system, observing a regular pattern, in which an event of kind *A* is always or often followed by an event of kind *B*, does not constitute a sound basis

on which to assert that any, let alone all, of these *A*-type events caused the ensuing *B*-type events. At best, a correlation of this kind constitutes *suggestive* evidence that there may be a causal connection between *A*-type and *B*-type events—that '*A* causes *B*'. This is one reason why controlled experiments, including RCTs, are so powerful: if an *A*-type event leads to a *B*-type event under controlled conditions—when interference by other causal processes has been prevented—then this constitutes persuasive evidence of a causal connection between events of these kinds, for if the *A*-type event did not cause the *B*-type event, then how was it brought about? This is also why large-*n* observational studies focus on controlling for possible confounders. In seeking to replicate experimental control through statistical control, this practice implicitly acknowledges that a mere correlation, observed in an open system, offers, at best, only limited support for a causal claim.[16]

The weakness of a mere correlation, observed in an open system, as evidence for a causal claim is perhaps the best-known example of a broader and very important rule: evidence that is consistent with a causal claim, but which provides no further reason to accept that claim beyond the (mere) fact that the evidence is consistent with it, does not provide persuasive support for that claim.[17] To get a sense of why this is, suppose that those investigating a particular house fire suspect that it may have been caused by a short circuit and hence seek evidence to support this—evidence which will provide persuasive support for the explanatory statement 'a short circuit caused the fire'. Suppose, further, that the investigators discover that rats were present in the property. Now, it is widely known that rats can chew through electrical insulation, exposing live wires which can cause a short circuit. But how strongly does evidence that rats were present support the claim 'a short circuit caused the fire'? We hope it is uncontroversial to suggest that this evidence supports the claim only weakly at best. Although the presence of rats is *consistent* with the causal claim of interest, the evidence is, to borrow a well-known expression, 'circumstantial': it shows that rats were present, but not that they actually chewed through any important electrical insulation, let alone that this caused the fire. Even if evidence could be found of rats having chewed through some electrical insulation, this would still constitute

[16] We emphasize 'observational' studies here because quantitative analysis is also used to report and analyse the results in experimental studies which seek to control for confounders through the research design, rather than through purely statistical controls. On the recent shift towards design-based approaches in the social sciences, see Dunning (2010, 2012).

[17] It is especially important to be aware of this rule in relation to so-called 'congruence' methods (see, for example, Van Evera 1997). George and Bennett (2005: 181–8) explicitly draw attention to the risk, when employing such methods, that the 'congruence' of evidence with a putative causal claim may be spurious.

merely suggestive, and not persuasive, evidence that a short circuit caused the fire.[18]

The reason why such evidence is not persuasive is that it does not rule out competing accounts of what caused the short circuit. Evidence that rats were present, or even that they chewed through some electrical insulation, suggests that a short circuit *may* have caused the fire, but it does not rule out other possible causes, such as a cooking-related incident, a piece of clothing left over an electric heater, or an unextinguished cigarette falling onto a curtain.

If evidence is found which is (merely) *consistent* with an explanatory statement of interest, then this statement passes what Van Evera (1997: 31) terms a 'hoop test' in relation to that evidence (see also Collier 2011; Mahoney 2012). Had the evidence been *inconsistent* with the explanatory statement, then the statement would have been refuted, so in order for the statement to remain viable it 'must jump through the hoop' which the evidence creates (Van Evera 1997: 31), but this does little to establish that the statement should actually be accepted. By contrast, if evidence can be found which rules out competing accounts of how the outcome of interest was brought about, then this will provide persuasive grounds on which to accept this statement. Let us elaborate.

If evidence is sought in support of an explanatory statement, '*a* caused *b*', but the evidence that is found fails to rule out the possibility that *b* was brought about by a combination of events which did not include *a*, then this evidence fails to provide persuasive support for the statement. The reason is simple: if the possibility remains open that *b* was brought about by a combination of events which did not include *a*, then we cannot be sure that *a* was among the causes of *b* and hence we lack persuasive grounds on which to assert that '*a* caused *b*'. This is the scenario we illustrated above: evidence that rats were present, or even that they had chewed through some electrical insulation, constitutes suggestive, rather than persuasive, evidence in support of the statement 'a short circuit caused the fire' because it fails to rule out the possibility that the fire was brought about in some other way. Yet, if evidence could be found which was consistent with the statement 'a short circuit caused the fire' *and also* ruled out other, competing accounts of how the fire was brought about, then such evidence would provide persuasive grounds for accepting this statement, for if other ways in which the fire might have been brought about have been ruled out, then it *must have been* caused by a short circuit.

[18] In what follows, we will draw a simple distinction between evidence which is (merely) 'suggestive' and evidence which is 'persuasive'. We acknowledge that in practice this is a continuum, not a binary division. We will return to this issue in Chapter 7.

This suggests that in order to provide persuasive support for a given explanatory statement, evidence must not only be consistent with this statement, but must also rule out competing explanatory statements. Given that, as we argued in Section I, it is necessary to provide support for an explanatory statement in order to provide support for a causal theory, this indicates that ruling out other explanatory statements which compete with an explanatory statement of interest is central to the logic of causal inquiry. In order to provide support for a causal theory, it is necessary to provide support for an explanatory statement from which that theory may be inferred, and in order to providing persuasive support for this explanatory statement it is necessary to rule out competing explanatory statements. As we shall see, however, what this involves must be explicated with care.

Ruling out competing explanatory statements

It is often suggested that ruling out 'alternative explanations' (or the like) constitutes a key component of causal inquiries (see, for example, Mackie 1965: 246; Mackie 1980: 74; Shadish, Cook, and Campbell 2002: 6, 53–4; George and Bennett 2005: 117–19; Cartwright 2007: 30; Elster 2007: 17, 20; Bhaskar 2008: 125; Brady 2008: 225; Reiss 2009: 34; Collier, Brady, and Seawright 2010b: 161; Bennett 2010; Bennett and Checkel 2015b: 23–4).[19] Yet, existing discussions tend to be somewhat imprecise in how they describe what kind of statement must be ruled out.

One indication of this imprecision is that a wide variety of terms is used to describe what it is that must be ruled out. The term that is most commonly used is probably 'alternative explanations' (see, for example, Elster 2007: 20; Reiss 2009: 34; Collier, Brady, and Seawright 2010b: 181; Bennett 2010: 207; Bennett and Checkel 2015b: 18, 23–4), but others include 'rival explanations' (see Collier, Brady, and Seawright 2010b: 161; Bennett 2010: 212; Mahoney 2012: 573), 'competing historical explanations' (George and Bennett 2005: 118), 'competing explanatory claims' (Bennett 2010: 207), 'alternative causes' (Bhaskar 2008: 125), 'alternative causal accounts' (Reiss 2009: 34), 'rival accounts' (Elster 2007: 17), 'alternative hypotheses' (Bennett 2010: 209; Collier 2011: 826), 'rival hypotheses' (Collier 2011: 825; Mahoney 2012: 589; Rohlfing 2014; Zaks 2017: 344), 'contending hypotheses' (Bennett 2010: 207;

[19] Some readers may notice that this list of sources includes philosophical works and qualitative methods texts, but not quantitative methods texts. As we shall see, however, it would be a mistake to infer from this that ruling out competing explanatory statements is a feature only of qualitative approaches to causal inquiry, or only of those research projects which explicitly aim to generate explanations of particular outcomes.

Zaks 2017), 'competing hypotheses' (Mahoney 2012: 589), and 'alternative theor[ies]' (Zaks 2017: 348).

This range of terminology obscures two fundamental issues. First, what needs to be ruled out in order to provide persuasive support for an explanatory statement, '*a* caused *b*', are not just *other, alternative* statements describing how *b* was brought about, but rather statements which offer *competing* accounts of how *b* was brought—that is, statements which are *inconsistent* with '*a* caused *b*'.[20] Second, the statements which need to be ruled out *really are explanatory statements* and not, say, theories or hypotheses. Let us explore both of these issues in more detail.

First, whereas some explanatory statements concerning the occurrence of a given event, *b*, compete with one another, others are complementary. Let us define an explanatory statement, '*a* caused *b*', as *competing with* another such statement, '*c* caused *b*', if the two statements could not be consistently accepted—if it could not be the case both that '*a* caused *b*' and that '*c* caused *b*'.[21] By contrast, an explanatory statement, '*a* caused *b*', is *complementary to* another such statement, '*c* caused *b*', if both statements could consistently be accepted—if it could be the case both that '*a* caused *b*' and that '*c* caused *b*'.[22]

To illustrate this distinction, consider a series of explanatory statements concerning the outbreak of World War I:

a) the rigid alliance structure which took shape in the late nineteenth century caused the war
b) the assassination of Archduke Franz Ferdinand on 28 June 1914 caused the war

[20] Whereas the existing literature talks of ruling out alternative (or competing) *explanations*, we will talk of ruling out competing *explanatory statements*. This enables us to maintain a clear focus on the question of what evidence persuasively supports such a statement, setting to one side the question of whether such a statement is offered as an explanation and, if it is, whether it is satisfactory as an explanation. This is important for two reasons. First, we argued in Section I of this chapter that providing persuasive support for explanatory statements is the means by which causal theories become well supported. Identifying what evidence will persuasively support an explanatory statement is therefore important even if the purpose is not to offer an explanation. Second, whether an explanatory statement, offered as an explanation, is *satisfactory as an explanation* raises a question about its pragmatic adequacy. We will discuss this issue in Chapter 6.

[21] Mahoney (2012: 573) refers, in passing, to 'contradictory alternatives': this captures what we mean by the term 'competing', used in relation to explanatory statements.

[22] Bennett and George (2005: 188) also emphasize the importance of determining whether 'causal arguments are complementary or competing'. However, they do not draw as sharp a distinction as we believe is necessary. For example, they say: 'Sometimes competing explanations can be equally consistent with the available historical evidence; this makes it difficult to decide which is the correct explanation or, alternatively, whether both interpretations may be part of the overall explanation … Another possibility is that each of the ostensibly competing explanations in fact addresses different parts of a complex longitudinal development' (2005: 92). But if two explanatory statements are both part of an 'overall' explanation (that is, both can consistently be accepted), or if they have different explananda (because they relate to distinct parts of a longitudinal development), then they are complementary, not competing.

c) Austria's ultimatum to Serbia on 23 July 1914 caused the war
d) Russia's partial mobilization against Austria on 28 July 1914 caused the war

Although there is an obvious sense in which these statements might be said to offer 'alternative explanations' for the war, it is prima facie plausible to believe that *all* the factors they identify contributed to the outbreak of World War I. If so, then these statements are complementary—they can all, consistently, be accepted and hence there is no need to choose between them.[23] By contrast, consider a further statement:

e) whereas Russia's partial mobilization against Austria on 28 July 1914 caused Germany to renew her efforts to avoid war, Russia's general mobilization on 30 July 1914 made war inevitable.[24]

This statement appears to be complementary to (a)–(c), but it clearly competes with (d): if Russia's partial mobilization against Austria led to a softening of Germany's position, making the avoidance of war more likely, then it is hard to maintain that this partial mobilization caused the war. Hence, if (d) is accepted, then (e) cannot be accepted, and vice versa.[25]

The reason this distinction is important is that in order to provide persuasive support for an explanatory statement of interest, evidence must rule out *competing* explanatory statements; ruling out *complementary* explanatory statements is not required. This thought is worth pursuing a little further.

All events are brought about by a more or less complex conjunction of other events, each of which has its own more or less complex causal history (cf. Lewis 1993). Consequently, an explanatory statement, '*a* caused *b*', does not identify one event, *a*, as the *sole* cause of another event, *b*, but rather identifies *a* as one of the (potentially very many) events which contributed to the bringing about of *b*.[26] But if many events helped to bring about *b*, then many

[23] Of course, if we are seeking to explain the outbreak of World War I to someone, then we may find one of these statements more apt for our purpose than the others, but this is a pragmatic issue, which we will consider further in Chapter 6.

[24] This is roughly the argument advanced by Trachtenberg (1990–1: 129–37). For our purpose here, what matters is not whether his argument is right, or even whether we have summarized it correctly, but merely that (e) appears inconsistent with (d).

[25] Note that a simple statement to the effect that Russia's general mobilization on 30 July 1914 caused the war would not compete with (d): what makes (e) compete with (d) is (e)'s implication that claims about the causal significance of Russia's partial mobilization against Austria are mistaken.

[26] As Lewis points out, people often speak 'of *the* cause of something', as if suggesting 'that there is only one', but this is misleading: what is being emphasized in such statements is the *salience* of this cause 'for the purposes of some particular inquiry' (1993: 183). We will consider this issue further in Chapter 6, when we discuss the pragmatic dimension of causal explanation. However, it is worth emphasizing at this point that the idea that something could have a 'sole' cause is nonsensical. All causal relations unfold (or fail to unfold) in some context and this context contributes causally to the outcome: if it did not, then it would not

explanatory statements concerning the bringing about of *b* will be acceptable. Hence, when we ask whether an explanatory statement, '*a* caused *b*', is supported by the empirical evidence, what we are asking is whether *a* was among the (potentially very many) causes of *b*.

In attempting to answer this question, there is little to be gained by ruling out 'alternative' explanatory statements unless the acceptability of these alternatives has some direct bearing on the acceptability of '*a* caused *b*'. To frame this point in relation to our illustration, if the aim is to make the case for (d), there is little to be gained by seeking to rule out (a): showing that the rigid alliance structure *was not* among the causes of World War I does nothing to show that Russia's partial mobilization against Austria *was* among the causes of the war. Certainly, it is not *necessary* to rule out (a) in order to make the case for (d), for these statements are complementary—both can be accepted.

In order to provide persuasive support for '*a* caused *b*' it is necessary to rule out precisely and only those 'alternative' explanatory statements which *compete* with '*a* caused *b*'—those which indicate, explicitly or implicitly, that *a* was *not* among the events which contributed to the bringing about of *b*. We indicated above why this is: if the possibility that *a* did *not* contribute to the bringing about of *b* can be eliminated, then it must be the case that *a did* contribute to the bringing about of *b*, in which case the explanatory statement '*a* caused *b*' should be accepted (as should the theory '*A* causes *B*'). To frame this, once again, in relation to our illustration, above: in order to make a persuasive case for (d), it is necessary to rule out (e), for if (e) is acceptable, then (d) is not. It is also necessary to rule out any other possible explanatory statements which compete with (d). However, it is *not* necessary to rule out any of (a)–(c).

What this reveals is that talk of ruling out 'alternative explanations' (or the like) is insufficiently precise as a guide to how persuasive support for an explanatory statement of interest can be generated. Not all 'alternatives' are equally significant. In order to provide persuasive support for the statement '*a* caused *b*', evidence must rule out *competing* explanatory statements; ruling out 'alternative', but *complementary* statements is not required and is not likely to achieve very much.[27]

be possible to make sense of why a particular causal relation unfolds fully under controlled, experimental conditions but fails to unfold fully outside of such conditions. Moreover, all events which are identified as 'causes' have their own causal history: there is no such thing as an uncaused cause. These considerations are overlooked by those who criticize causal analysis as being intrinsically bound up with politically problematic 'origins' thinking (see, for example, Zehfuss 2003; Kurki and Suganami 2012: 406–11).

[27] We therefore have some doubt about the utility of an investigative strategy which seems to be employed in at least some instances of process tracing in IR, in which the 'alternative' explanations which

The second way in which existing discussions of this issue are insufficiently precise is that they fail clearly to distinguish between explanatory statements, which can compete with one another, and causal theories, which cannot compete with one another. Some readers may be surprised at the idea that causal theories cannot compete with one another: if there can be competing explanatory statements, then it may seem natural to suppose that there can also be competing theories.[28] It is certainly common, in the literature on process tracing, for the terms 'explanation' and 'theory' to be used fairly interchangeably. For example, Collier, Brady, and Seawright (2010b: 170) talk of testing 'well-developed theories against clear alternative explanations', while Mahoney (2015: 201) describes the tasks of identifying 'alternative explanations of World War I' and of assessing these alternative explanations as tasks of '*theory construction*' and '*theory testing*' respectively (italics in original).[29] This kind of conflation is quite unhelpful.

In order to understand the deep logic of causal inquiry, it is crucial to recognize that an explanatory statement identifies one or more causes of a specific event, or set of events, located in space and time. Explanatory statements compete with one another insofar as they offer inconsistent accounts of what brought about specific events. By contrast, causal theories abstract from specific events to describe causal propensities. Statements of causal theories do *not* offer inconsistent accounts of what brought about any specific event, because they do not even *refer* to specific events! While it might be the case that two theories describe propensities which could not both have unfolded fully in one specific time and place, this does not mean that these two theories compete with one another. An illustration will help to make this clear.

Take two explanatory statements which provide competing accounts of the origins of the same fire: 'a short circuit in a kitchen appliance caused the fire' and 'an unattended candle in the bedroom caused the fire'. These statements compete with each other because if one is accepted the other must be rejected: the fire might have been caused by a short circuit in

are compared are selected because they are associated with certain prominent IR theories, without explicitly considering whether they compete with one another (for examples, see Bennett 2010). Although our emphasis is somewhat different to hers, we agree with Zaks (2017) that it would be a serious mistake simply to *assume* that any two 'alternative' explanations for a given event, *b*, are 'competing' and hence simply to *assume* that if one explanation can be ruled out then the other should therefore be accepted.

[28] The frequently encountered supposition that theories can compete with one another once again reflects the influence of the culture of generalization: the background presumption is that theories state 'generalizations' which can be directly tested against empirical evidence. As we have emphasized, however, theories are typically expressed in *abstract* statements.

[29] Moreover, the series of tests that are now often associated with ruling out 'alternative explanations' in process tracing studies (see, for example, Collier 2011; Mahoney 2012; Rohlfing 2014) were introduced by Van Evera (see 1997: 17–40) in the context of a discussion of how *theories* can be tested.

a kitchen appliance or by an unattended candle in the bedroom, but it cannot have been caused by both. However, while these two explanatory statements cannot both be accepted, the causal theories which they imply *can* both be accepted: there is nothing inconsistent about asserting both that 'short circuits cause fires' and that 'unattended candles cause fires'. Admittedly, it is unlikely that both of these propensities unfolded fully in any specific time and place: any specific fire will have been caused by a short circuit *or* by an unattended candle, but not both, which is why the two explanatory statements we have been considering compete with one another. However, it makes no sense to suggest that the theory 'short circuits cause fires' can be accepted only if the theory 'unattended candles cause fires' is rejected. In order to provide persuasive support for an explanatory statement, what needs to be ruled out are competing explanatory statements concerning specific events which have occurred, and which are located in time and space, *not* 'alternative' (or supposedly 'competing') *theories*.

Before we move on, it is worth entering one important additional observation. We recognize that, when they seek empirical evidence in support of causal claims, those conducting causal inquiries may not describe, or even conceive of, evidence which appears to support explanatory statements of interest as 'ruling out competing explanatory statements'. Let us give a quick example. Suppose that burn patterns indicate overwhelmingly that a house fire originated behind an electrical socket in the kitchen.[30] It would be natural simply to treat this as persuasive evidence *in favour* of the statement 'a short circuit caused the fire' without explicitly thinking through which competing explanatory statements this evidence rules out (and, perhaps, without even thinking through the fact that such a statement will be acceptable only if competing explanatory statements can be ruled out). Note, though, that if the evidence does indicate overwhelmingly that the fire was caused by a short circuit, then it must *serve to* rule out competing explanatory statements concerning how the fire was brought about, regardless of whether these competing possibilities are explicitly considered.[31] In other words, although evidence which appears persuasively to support an explanatory statement of interest may not be self-consciously conceptualized as ruling out competing explanatory statements, reflecting on why such evidence appears

[30] Lentini (2013) discusses some of the difficulties that arise when interpreting such patterns.

[31] In this respect, some discussions of the tests often used in process tracing are confusing. Contrary to what Bennett (2010: 210) and Collier (2011: 825) assert, so-called 'smoking gun' evidence (that is, evidence which counts decisively in favour of a given explanatory statement) must serve to eliminate competing explanations: how could it be otherwise?

persuasively to support the explanatory statement of interest confirms that it does so because it serves the function we have identified—because it rules out competing explanatory statements.

Logic, not methods

Our contention that the evidence which is required persuasively to support an explanatory statement is evidence that rules out competing explanatory statements is based on logic—on thinking through what must be the case for a given explanatory statement to be acceptable—and is not tied to any particular method of data collection and analysis.

Although ruling out 'alternative explanations' may be associated predominantly with qualitative methods, quantitative methods, too, are geared towards generating evidence which counts persuasively in support of explanatory statements of interest because it rules out competing explanatory statements. This may not be immediately apparent because the stated aim of quantitatively oriented studies is often to provide support for a causal theory and because the language of 'ruling out alternative explanations' (and the like) is not explicitly used. However, if the evidence generated through the application of quantitative methods did not rule out competing explanatory statements, then it would not provide persuasive support for explanatory statements of interest and hence would not succeed in demonstrating that the theories which can be inferred from such statements are well supported. In fact, 'controlling for confounders', which *is* often discussed in quantitatively oriented studies, is properly understood as a *means by which* competing explanatory statements can be ruled out. Let us briefly illustrate this by returning to the hypothetical example of an RCT which we discussed in the previous section, the aim of which is to determine whether 'aspirin relieves headache'.

We supposed that this RCT was an ideal experiment, in which the treatment and control groups are balanced across all possible confounders; we also stipulated that the treatment group, on average, experienced a noticeably higher degree of relief from headache than the control group. We argued that this outcome supports the explanatory statement 'giving aspirin to members of the treatment group caused them, on average, to experience a higher degree of relief from headache', from which the theory 'aspirin relieves headache' can be inferred. What we wish to point out now is that the evidence strongly supports this explanatory statement because, given the design of the RCT, it serves to rule out competing explanatory statements.

In order to see this, consider what kind of explanatory statement would compete with the statement 'giving aspirin to members of the treatment group caused them, on average, to experience a higher degree of relief from headache'. A competing explanatory statement would hold that giving aspirin to members of the treatment group, but not the control group, did *not* have any causal effect and hence that something else caused the differential outcome experienced across the two groups. Because we have specified that the RCT was an ideal experiment, any such competing explanatory statements can be ruled out: by design, there is no difference between the treatment and control groups which could have contributed to the differential outcome observed across these two groups, other than the fact that the former, but not the latter, received the treatment. Consequently, it must be the case that giving aspirin to members of the treatment group contributed causally to their relief from headache. It is, in fact, this capacity definitively to rule out competing explanatory statements which makes the design of an RCT so powerful: because an ideal RCT removes all differences between the treatment and control groups other than the fact that the former, but not the latter, received the treatment, when a differential outcome is observed across the two groups, all explanatory statements which compete with the claim that this was caused by the treatment can be ruled out.

Of course, we acknowledge that researchers who conduct RCTs do not typically reason in this way, at least not explicitly. It is important to recognize, however, that insofar as it controls for possible confounders, the design of an RCT does thereby *serve to rule out competing explanatory statements*. The important underlying point is that whatever methods of data collection and analysis are used in the context of causal inquiry, the requirement is the same: in order to provide persuasive support for a causal theory, it is necessary to provide persuasive support for an explanatory statement, and in order to provide persuasive support for an explanatory statement it is necessary to find evidence which rules out competing explanatory statements, whether this is explicitly recognized by those conducting causal inquiries or not. This is a matter of logic, not of methods.

III. Applying causal theories to generate explanations, predictions, and policy prescriptions

In Sections I and II, we showed that a causal theory becomes well supported only when persuasive support is provided for an explanatory statement from which it may be inferred, and that in order to provide persuasive support

for an explanatory statement it is necessary to identify evidence which rules out competing explanatory statements. One reason why it is important to be clear about this is that once they become well supported, causal theories can be applied to generate causal explanations, causally informed predictions, and causally informed policy prescriptions; in fact, this is often a central goal of causal inquiry. It is therefore important to consider what is involved in applying a causal theory and also to consider how reliable the resulting explanations, predictions, and policy prescriptions are likely to be. As we shall see, while the idea of 'applying' causal theories is a familiar one, the requirements that must be fulfilled in order to generate causal knowledge in this way are more demanding than is often realized.

If a theory, '*A* causes *B*', is well supported, it might be supposed that this theory can be drawn upon fairly straightforwardly to generate causal explanations and causally informed predictions and policy prescriptions. This supposition is likely to prove especially seductive when causal theories are described using language associated with the culture of generalization: if theories are conceived of as 'general', then it might seem obvious that the way to generate explanations, predictions, and policy prescriptions is to 'apply' the knowledge encapsulated in well-supported 'general' theories to 'particular' cases of interest. Thus, if '*A* causes *B*' is a well-supported theory, it might be supposed that this theory can be applied to generate the prediction that some *A*-type event of interest will cause a *B*-type event and/or the prescription 'to produce a *B*-type event, perform an *A*-type event'. Equally, if an *A*-type event has occurred, followed by a *B*-type event, and the theory '*A* causes *B*' is well supported, it might be supposed that this theory can be applied to generate an explanation: the *A*-type event caused the *B*-type event.

Such suppositions are mistaken. As we have pointed out, a causal theory, '*A* causes *B*', does *not* imply a corresponding regularity statement, 'whenever *A*, then *B*'. The theory '*A* causes *B*' is a propensity statement; what it implies is that an *A*-type event will produce a *B*-type event *if the conditions are right*—that is, if interference with the full unfolding of this propensity is prevented. When applying a causal theory, it is therefore necessary to consider the context in relation to which it is being applied—to consider whether the prevailing conditions are such as to allow the propensity described by the theory to unfold fully. Let us elaborate.

Suppose that empirical evidence has been found which persuasively supports an explanatory statement, '*a* caused *b*', thereby providing strong indirect support for the theory '*A* causes *B*'. It is important to appreciate that *a* caused *b* under some set of conditions. Let us call these conditions *c*, using the lower-case *c* to refer to the actual conditions, located in time and space, under

which *a* caused *b*, and the upper-case *C* to refer, generically, to conditions of this kind. What the evidence shows is not only that *a* caused *b*, and hence that there is a propensity for *A*-type events to produce *B*-type events, but also that this propensity unfolds fully under conditions of kind *C*—that conditions of this kind are among the (potentially many) sets of conditions which are 'right' for this propensity to unfold fully. It follows from this that another *A*-type event that occurs under conditions of kind *C* will also produce a *B*-type event.

This knowledge provides strong grounds on which to advance an explanation, prediction, or policy prescription to the effect that some *A*-type event did, will, or would produce a *B*-type event, *but only if this A-type event occurred (or will occur) under conditions of kind C*. It is much harder to justify a claim that an *A*-type event which occurred (or will occur) under *different* conditions—conditions which are not of kind *C*—did, will, or would produce a *B*-type event. As this point may be difficult to grasp when expressed in the abstract like this, let us spell it out by means of a hypothetical example.

Suppose that empirical evidence has been found which persuasively shows that a revolution in country *P* caused a war, thereby also supporting the theory 'revolutions cause wars'. This evidence demonstrates not only that there is a propensity for revolutions to produce wars, but also that this propensity unfolds fully under conditions of the kind which prevailed when the revolution in country *P* occurred. Armed with this information, the theory 'revolutions cause wars' can be applied to generate a reliable explanation, prediction, or policy prescription relating to the occurrence of some other revolution, but only if this revolution also occurs (or occurred) under conditions of the kind which prevailed when the revolution in country *P* occurred.

Suppose that a further revolution occurs in country *Q* under conditions of just the kind which prevailed when the revolution in country *P* occurred. The theory 'revolutions cause wars' being well supported, it is known that revolutions have a propensity to produce wars. It is also known that this propensity unfolds fully under the conditions which prevail in country *Q*. In this scenario, the theory 'revolutions cause wars' can be applied with a high degree of confidence to generate a prediction that the revolution in country *Q* will cause a war.

Relatedly, suppose that rogue elements in country *R* wish, for nefarious reasons, to provoke a war and that the conditions which prevail in country *R* are of just the kind which prevailed when the revolution in country *P* occurred. The theory 'revolutions cause wars' being well supported, it is known that revolutions have a propensity to produce wars. It is also known

that this propensity unfolds fully under the conditions which prevail in country *R*. This indicates that if a revolution were to occur in country *R*, it would produce a war. In this scenario, the theory 'revolutions cause wars' can be applied with a high degree of confidence to generate a policy prescription, albeit a distasteful one: 'foment a revolution, and this will cause a war'.

Finally, suppose that in country *S* a revolution occurred, followed by a war. Suppose, moreover, that the conditions which prevailed in country *S* at the time of the revolution were of just the kind which prevailed when the revolution in country *P* occurred. The theory 'revolutions cause wars' being well supported, it is known that revolutions have a propensity to produce wars. It is also known that this propensity unfolds fully under the conditions which prevailed in country *S* at the time of the revolution there. In this scenario, the theory 'revolutions cause wars' can be applied with a high degree of confidence to generate an explanation for the war in country *S*: it was caused by the revolution which preceded it.[32]

These hypothetical examples illustrate how a well-supported causal theory might, in principle, be applied to generate a reliable explanation, prediction, or policy prescription. However, it is important to appreciate that a causal theory can be applied in this way only when we are quite confident that the conditions which prevail (or prevailed) in the scenario of interest are conditions under which the propensity described by the theory unfold fully. The evidence which shows persuasively that a revolution in country *P* caused a war, and which thereby also demonstrates that there is a propensity for revolutions to produce wars, reveals that this propensity unfolds fully under the conditions which prevailed in country *P* at the time of the revolution, but it does *not* reveal what *other* conditions may also be right for this propensity to unfold fully. Consequently, if a revolution were to occur in country *Q*, but the prevailing conditions were *not* of the kind which prevailed when the revolution in country *P* caused a war, the theory 'revolutions cause wars' could *not* be applied to generate a prediction that this revolution will cause a war. This is because we lack evidence about whether the propensity for revolutions to produce wars will unfold fully *under these conditions*. The same applies, *mutatis mutandis*, in relation to efforts to generate explanations and policy prescriptions.

[32] We can have high confidence in this explanation because it is supported by clear evidence that there is a propensity for revolutions to produce wars and that this propensity unfolds fully under the relevant conditions. This explanation may, however, not be satisfactory in the sense that it may not tell someone everything (or, indeed, anything) that they wish to know about what caused the war in country *S*. We will explore this thought in more detail in Chapter 6.

Cartwright and Hardie (2012: *passim*) frame the point we are making specifically in relation to policy prescription by drawing attention to the gap between the claim that a particular intervention 'worked there' and the claim that it 'will work here'. In relation to our example, the point they are making is that we cannot be confident that just because a revolution in country *P* caused a war, a revolution in country *R* will also cause a war, for the conditions which prevail in country *R* may be different from the conditions which prevailed at the time of the revolution in country *P*. We are making a similar, but broader, point: if a well-supported causal theory is to be applied to generate an explanation, prediction, or policy prescription, two requirements must be fulfilled. First, evidence must persuasively support an explanatory statement from which the theory may be inferred: this is what establishes the theory as well supported. Second, we must have a good reason to believe that the propensity described by this theory will unfold fully under the conditions which prevail in the target situation—in the situation about which an explanation, prediction, or policy prescription is to be generated.

In practice, of course, we may not know, for sure, that the conditions which prevail in a target situation are of the same kind as conditions under which the propensity described by the theory we wish to apply is known to unfold fully. This raises the question of when it might be possible to infer, or judge, that a propensity which is known to unfold fully in one scenario will also unfold fully in another scenario. We will examine this issue in more detail in Chapter 7, when we discuss the oft-heard idea that causal inferences can possess 'external validity', allowing their conclusions to be generalized.

For now, the point we wish to emphasize is that generating new causal knowledge (in the form of predictions, policy prescriptions, and explanations) by applying a causal theory is more difficult than is commonly appreciated. That is because a causal theory, '*A* causes *B*', is a propensity statement, not a generalization, and so in order to be confident that any given *A*-type event will produce a *B*-type event it is necessary to be confident not only that there is a propensity for *A*-type events to produce *B*-type events, but also that the conditions are such as to allow this propensity to unfold unhindered. In IR, there are few causal theories which are convincingly supported by empirical evidence, and it seems reasonable to wager that even less is known about the various sets of conditions under which the propensities they describe unfold fully. This imposes a significant constraint on the causal knowledge that can be generated by applying causal theories, even well-supported ones.

Before we conclude, it is worth emphasizing one further point. In Section I, we argued that the only kind of causal statement which can be directly

supported by empirical evidence is an explanatory statement; in Section II, we showed that evidence will persuasively support an explanatory statement when it rules out competing explanatory statements. This suggests that it is possible to generate a causal explanation directly, without having to apply a causal theory; this is done by identifying evidence which persuasively supports the explanatory statement we intend to offer as an explanation because it rules out competing explanatory statements. We have shown, in this section, that a causal explanation can *also* be generated by applying a well-supported causal theory, if the evidence shows that the propensity this theory describes unfolds fully under the conditions which prevail in the target situation. There are therefore two different ways in which a causal explanation can be generated: by identifying evidence which directly supports an explanatory statement that is to be offered as an explanation, or by applying a well-supported causal theory.[33] However, both of these ways of generating a causal explanation require that persuasive support be provided for an explanatory statement, by identifying evidence which rules out competing explanatory statements.[34]

By contrast, there is only one way in which a causally informed prediction or policy prescription can be generated: by applying a well-supported causal theory. As we explained earlier, this is because it is not possible to provide direct empirical support for statements relating to events which have not yet occurred. However, it is crucial to recognize that, just as both ways of generating a causal explanation require that persuasive support be provided for an explanatory statement, this is also required in order to generate a causally informed prediction or policy prescription by applying a well-supported causal theory, for this theory can be well supported only if persuasive support has been provided for an explanatory statement from which that theory has been inferred. This reveals a crucial feature of the logic of causal inquiry: causal theories, causal explanations, causally informed predictions, and causally informed policy prescriptions *all* rest upon the

[33] In broad terms, it is probably the case that historians typically construct explanations by providing empirical evidence which supports an explanatory claim directly (because it rules out competing explanatory statements), whereas social scientists typically construct explanations by applying well-supported causal theories. These two possible approaches to the explanation of a particular outcome are not often explicitly noted in debates about the relationship between 'history' and 'theory' in IR, where the focus is more often on alleged differences in the aims of (causal) inquiry: constructing explanations versus constructing theories (see, for example, Levy 1997, 2001).

[34] Not noticing that it is possible to generate a causal explanation directly, by ruling out competing explanatory statements, Jackson (2017) contends that a causal explanation always invokes (what he misleadingly characterizes as) a more 'general' causal claim (see also Jackson, forthcoming). It is crucial to recognize, as we have pointed out in this chapter, that the logic of causal inquiry begins with the provision of support for an explanatory statement, from which an abstract causal theory is inferred, not the other way around.

identification of evidence which persuasively supports explanatory statements because it rules out competing explanatory statements.[35] With this in mind, we are now in a position fully to elaborate the logic of causal inquiry.

IV. The deep logic of causal inquiry, elaborated

As we have shown, the logic of causal inquiry emerges out of the distinction between the two basic kinds of causal statement. The starting point is a simple observation: if causal statements are to be supported by empirical evidence, they must refer to specific events about which evidence can be gathered. It follows from this that empirical evidence can directly support concrete causal statements but not abstract statements. In fact, the only kind of causal statements that empirical evidence can directly support are *explanatory statements*—that is, concrete statements which relate to specific events that have already occurred, the distinctive epistemic function of which is to express causal explanations. Providing support for explanatory statements therefore lies at the heart of causal inquiry. Two further questions flow from this. First, what kind of evidence will provide persuasive support for an explanatory statement? Second, how can other kinds of causal statements become well supported?

In relation to the first question, we have argued that evidence will only count persuasively in favour of an explanatory statement if it rules out competing explanatory statements. If evidence is sought in support of the statement '*a* caused *b*', but the evidence which is found fails to rule out competing statements (statements which assert or imply that *b* was caused by some combination of events which does *not* include *a*), then this evidence is not persuasive: if the possibility that *b* was brought about by a combination of events which does not include *a* cannot be ruled out, then the claim that '*a* caused *b*' cannot be advanced with confidence.

In relation to the second question, the key point is that concrete statements imply corresponding abstract statements; hence, providing support for an explanatory statement, '*a* caused *b*', also, indirectly, supports the corresponding theory, '*A* causes *B*'. Once it is well supported, this theory can be applied to generate explanations, predictions, and policy prescriptions. However, it is crucial to be aware that a causal theory, '*A* causes *B*',

[35] Lipton (2004: 65) makes a similar point: 'Even when our main interest is in accurate prediction or effective control, it is a striking feature of our inferential practice that we often make an "explanatory detour". If I want to know whether my car will start tomorrow, my best bet is to try to figure out why it sometimes failed to start in the past'.

is a propensity statement: it implies that an *A*-type event will produce a *B*-type event if the conditions are right, but not that every *A*-type event will produce a *B*-type event. It is therefore crucial to consider what is known about the conditions under which the propensity described by a theory of interest unfolds fully. A causal theory can be applied to generate a robust explanation, prediction, or policy prescription only if the propensity it describes has been shown to unfold fully under the target conditions.

Given this logic, causal inquiry can be conceptualized as proceeding in three phases, each of which builds on the previous one. This is laid out in Figure 5.1.

In arguing that the later phases build on the earlier ones, we do not mean to suggest that every instance of causal inquiry involves all three. If the aim is to generate a causal explanation and this is done by providing evidence which directly supports an explanatory statement, '*a* causes *b*', then there is no need for Phase 1 to be followed by Phases 2 and 3. If it is widely known that a particular causal theory is well supported, then this theory may be applied to generate an explanation, prediction, or policy prescription, as described in Phase 3, without going through Phases 1 and 2. Note, however, that Phases 1

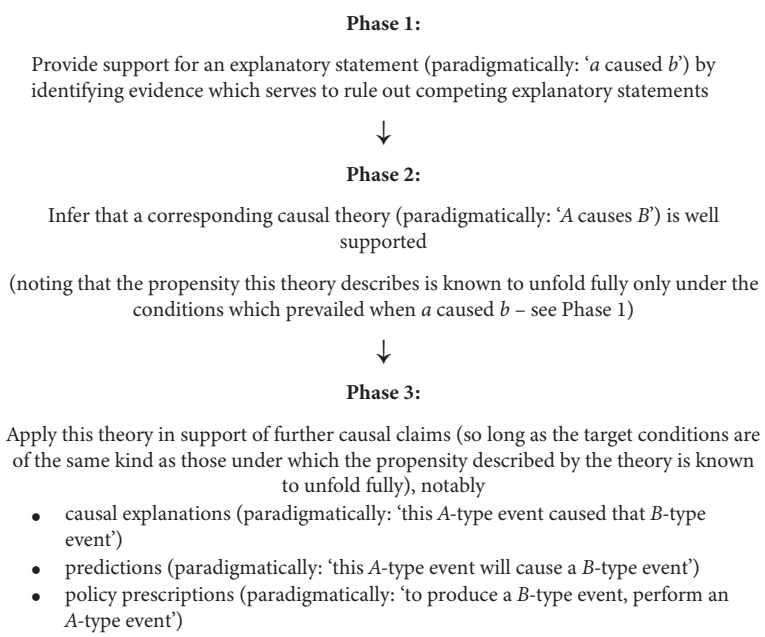

Figure 5.1 The three phases of causal inquiry

and 2 must have been completed at some point, in order for the application of a causal theory, as described in Phase 3, to be credible.

We also recognize that, in practice, these phases may not be explicitly identified or distinguished by those conducting causal inquiries. Our point is not to contend that those conducting causal inquiry in fact think of their inquiries as being divided into these phases (most do not!), but rather to emphasize that, logically speaking, Phase 3 depends on Phases 1 and 2. In other words, the provision of support for explanatory statements constitutes the fundamental building block of causal inquiry.

Conclusion

By way of conclusion, it is worth emphasizing three features of the logic we have elaborated.

First, providing support for explanatory statements, by identifying evidence which rules out competing explanatory statements, is required not only when the aim is to generate a causal explanation, but also when the aim is to demonstrate support for a causal theory or to generate a causally informed prediction or policy prescription. This radically undermines the widespread assumption, informed by the culture of generalization, that a principal aim of causal inquiry will ordinarily be to provide support for causal theories and that providing support for explanatory statements relating to specific events is a secondary concern, to be broached only when causal theories have been established. It is crucial to recognize that causal theories are abstract, not 'general', and that knowledge of a causal theory can be produced only by providing persuasive support for an explanatory statement from which this theory can be inferred.

Second, what we are calling 'the logic of causal inquiry' is methodologically neutral: it is a 'deep logic' which is not tied to any particular aim of causal inquiry, nor to any particular method of data collection or analysis. In fact, the logic we have elaborated is what underpins the application of all methods that are used to generate causal knowledge. One important implication of this is that any method for analysing empirical evidence, in the context of causal inquiry, is subject to the same evaluative standard—viz. can it generate persuasive support for explanatory statements of interest?

Third, and harking back to Part I of this book, the logic of causal inquiry is equally consistent with both causal realism and causal agnosticism. Causal realists hold not only that causal propensities (and underlying causal powers)

are real, but also that belief in their reality is required by science.[36] However, realists still face a practical question: how can empirical evidence support claims about causal propensities, given that these are not directly observable? The answer, as we have shown, is that a statement of a causal propensity can be considered well supported when empirical evidence persuasively supports an explanatory statement from which it can be inferred.

Now, when a causal theory is well supported, realists will doubtless hold that the propensity it describes is real and infer, further, that it is underpinned by a real causal power. Causal agnostics, of course, will wish to resist this suggestion. Agnostics accept causal discourse as a deeply embedded feature of human life and thought, which it would be hard to do without, and yet doubt that this reveals anything about the world existing independently of human thought. From the perspective of an agnostic, the logic we have described reveals when it is proper to assert statements of different kinds, but it does not reveal anything further than this; to characterize a particular kind of causal statement as describing a propensity is not to assert (but neither is it to deny) that causal propensities are part of the world existing independently of human thought.

This difference between causal realists and causal agnostics is a significant one, yet it has no bearing on the logic of causal inquiry which we have just elaborated, for this logic carries no implication about whether causal propensities are real. The logic of causal inquiry concerns how causal statements can be supported by empirical evidence and this issue is no less important for realists than it is for agnostics. It should therefore be possible to discuss this logic without being waylaid by discussions about the merits of causal realism. Indeed, one problem with the debate between causal realists and their critics in IR is that it has diverted attention away from the deep methodological issues we have discussed in this chapter.

These three features of the logic we have elaborated offer an indication as to why it is not more widely recognized and discussed. First, too little attention is paid to the basic kinds of causal statement and the relationship between them. An important culprit here is the culture of generalization, which supports the misleading characterization of causal theories as 'general' and hence obscures the crucial 'means-to-an-end' relationship that we noted between providing support for explanatory statements and providing support for causal theories. Second, many accounts of how empirical evidence can support causal statements are designed to explicate how particular

[36] Recall, from Chapter 2, that Bhaskar used the term 'tendencies' to refer to what we are here terming 'propensities'.

methods of data collection and analysis can be applied in the service of causal inquiry, neglecting the 'deep' logic which underpins the application of all such methods. Third, the debate about causal realism has distracted attention from the question of how empirical evidence can support causal statements.

In our view, the logic of causal inquiry which we have elaborated should be much more central to discussions of causal inquiry in IR. We recognize, however, that we have presented this logic in a fairly condensed fashion, skating over some key questions that it raises. The task of the next three chapters is hence to flesh it out in more detail, both by explicating key elements more fully and by showing how it can be drawn upon to illuminate the strengths and weaknesses of some substantive examples of causal inquiries in IR.

6
Causal Explanation

In Chapter 5 we argued that all causal knowledge claims rest on the provision of persuasive evidence in support of explanatory statements. As we pointed out, there is a crucial 'means-to-an-end' relationship between providing support for an explanatory statement (the means) and developing knowledge of a corresponding causal theory (the end). This is because causal theories become well supported only when they are inferred from persuasively supported explanatory statements.

A key implication of our analysis in Chapter 5 was that providing evidence which persuasively supports particular explanatory statements is fundamentally important for causal inquiry even if the aim is *not* to generate a causal explanation but rather to demonstrate support for a causal theory or to develop causally informed predictions or policy prescriptions. What, though, if the aim *is* to offer a causal explanation? What, if anything, is required to make such an explanation satisfactory, over and above the obvious requirement that it should be persuasively supported by empirical evidence? These are the questions we take up in this chapter. We have three principal reasons for devoting a whole chapter to them.

First, given our account of the logic of causal inquiry, it is important to spell out the relationship between, on the one hand, providing empirical support for an explanatory statement as a means to an end (that is, in order to establish indirect support for a causal theory) and, on the other hand, advancing a causal explanation.

Second, causal explanation is a rather neglected topic in IR. For the most part, this is not because it is considered unimportant, but rather because it is considered straightforward! If, as is widely supposed, causal explanation consists simply in the 'application' to 'particular' events of more 'general' causal theories, then there will be little to say about causal explanation as such. This supposition is, however, quite wrong. As we showed in Chapter 5, an explanatory statement is the only type of causal statement which can be directly supported by empirical evidence. The idea that generating reliable explanatory statements is a second-order concern, to be pursued only once reliable theories have been established, is therefore badly misleading. But this

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0007

is not the only problem with the conventional thinking. Existing discussions largely ignore a further, yet crucial, component of causal explanation: what we will term its 'pragmatic' component (see Humphreys 2017b). This concerns how well attuned a causal explanation is to the needs of its intended audience.

Third, these limitations in existing discussions of causal explanation have led to some quite significant misunderstandings. These stem both from a failure to appreciate the pragmatic component of causal explanation and from a failure to be precise about how causal explanations relate to other kinds of knowledge claims. Our aim in this chapter is therefore to provide an account of causal explanation which brings out its pragmatic component and which also addresses the many confusions that can and do arise when the topic is discussed more superficially.

We will proceed as follows. Having highlighted the relative lack of attention paid to the topic of causal explanation in IR, the first section will define what a causal explanation is and ask what makes it satisfactory. To answer this question, we will introduce the idea that causal explanation has a pragmatic component: a satisfactory causal explanation must, among other things, help resolve whatever puzzlement has motivated the demand for an explanation. In the second section, we will show how recognizing the pragmatic component of causal explanation can bring clarity to a series of otherwise potentially confusing puzzles about the content of causal explanations and the form in which they should be presented. These include the notion that causal explanations should be 'complete', questions about the forms which causal explanations can take, and the widespread demand that causal explanations should identify causal mechanisms. The final section will address a series of further red herrings which often arise when the topic of causal explanation is discussed in IR.

I. What is a causal explanation and what makes it satisfactory?

In IR, and especially in the mainstream research methods literature, it is striking how little attention is paid to 'causal explanation' as a topic worthy of discussion in its own right.[1] Consider, for example, the following influential methodology texts: King, Keohane, and Verba's *Designing Social Inquiry*

[1] Some recent exceptions to the general neglect of this topic in IR (which are, however, few and far between) include Suganami (2008, 2011), Grynaviski (2013), Chernoff (2014), Humphreys (2017b), Jackson (2017), Norman (2021), and Jackson (forthcoming).

(1994), Charles Ragin's *The Comparative Method* (2014) and *Redesigning Social Inquiry* (2008), and John Gerring's *Social Science Methodology* (2012) and *Case Study Research* (2017). These works contain, *between them*, just one index entry for 'explanation' or 'causal explanation': an entry in *Designing Social Inquiry* pointing to just a single page of that book (see King, Keohane, and Verba 1994: 241)!

What accounts for this pattern of neglect? These authors do not consider causal explanation irrelevant or only marginally relevant to their concerns: they are all concerned with how empirical methods can contribute to causal inquiry, thereby generating, among other things, explanations. King, Keohane, and Verba in fact describe 'explanation' as 'the ultimate goal' of social science research and explicitly endorse the search for 'meaningful causal explanations' (1994: 34). As Henry Brady (2010: 70–3) and James Johnson (2006: 234–7) argue, this literature neglects 'causal explanation' as a topic because it reduces 'explanation' to 'inference': it supposes that the way to provide an account of '(causal) explanation' is to provide an account of '(causal) inference' (see also Lipton 2004: 55). Thus, having identified 'explanation' as a key goal, King, Keohane, and Verba go on to assert that 'real explanation is always based on causal inferences' (1994: 75).[2] From that point on, they focus on 'inference' *rather than* 'explanation' (see Waldner 2007: 152).[3]

This way of thinking about causal explanation reflects, once again, the baleful influence of the culture of generalization. If it is supposed (wrongly!) that what is produced by means of a causal inference must be the kind of 'general' knowledge that is (supposedly) summarized in causal theories, and if it is supposed (wrongly!) that a 'specific causal explanation' can be generated quite straightforwardly by applying the 'generalized explanation' contained within a causal theory to a 'particular' subject of interest, then it may also appear (wrongly!) that there is little to be gained by addressing 'causal explanation' as a significant subject in its own right—that is, separate from 'causal inference'.[4]

There are lots of things wrong with this way of thinking. In Chapter 5, we highlighted an important underlying issue: causal theories are not 'general', but rather 'abstract', and they become well supported only when they

[2] Their one index entry for 'explanation' (noted above) points to the page on which this observation is made (see also Goertz and Mahoney 2012: 236).

[3] The idea that political scientists should focus on 'inference' rather than 'explanation' is reinforced by the widespread supposition that there is an implicit division of labour between political scientists and historians in which political scientists focus on constructing the kinds of 'lawlike statements about social behaviour' that are, on this way of thinking, applied in explanations, whereas historians draw on such statements to 'describe, explain, and interpret individual events' (Levy 1997: 22).

[4] We will discuss 'causal inference' in more detail in Chapter 7.

are inferred from persuasively supported explanatory statements. We also pointed out there are two ways in which a causal explanation can be generated: either by providing direct empirical support for an explanatory statement, or by applying a well-supported causal theory (which must, in turn, have been inferred from a persuasively supported explanatory statement). However, this observation, combined with our analysis of the kind of evidence which will persuasively support an explanatory statement, does not yet provide a full account of causal explanation. It is also necessary to consider what makes an explanatory statement that is offered as a causal explanation *satisfactory as an explanation*. It is crucial to recognize that, in addition to being supported by empirical evidence, a satisfactory causal explanation must fulfil the brief, addressing the needs of its intended audience.

Before we develop this line of argument any further, it is worth briefly pointing out that, in philosophy, there have been two broad approaches to causal explanation (see Salmon 1998). The first involves, in Nancy Cartwright's words, 'reducing causation away' (2002: 8): the aim is to show that although many explanations employ causal language, a causal explanation can be reduced to a non-causal statement (see Achinstein 1993a: 139). For example, Hempel famously argued that a statement that one event caused another event 'amounts to' a statement that events of the first kind are 'regularly accompanied' by events of the second kind (1942: 36).[5] This contention is closely linked to the so-called 'covering-law model' of explanation, of which Hempel was a key exponent. According to this model, an event is explained when it is subsumed under relevant scientific laws.[6] In the final part of this chapter, we will show how this model fails to provide an adequate account of causal explanation.

The second broad approach to causal explanation emerged as a response to the apparently insuperable difficulties with the first (see Achinstein 1993a). It accepts that the concept of 'cause' is irreducible (any attempt to define it wholly by reference to non-causal concepts will fail) and hence abandons the idea that causal explanations can be wholly reduced to non-causal statements. According to this second approach, a causal explanation identifies certain events as *causes* of the event(s) being explained: to explain such events is

[5] As we noted in Chapter 4, Hempel later amended this claim slightly, arguing that a statement that one event caused another event 'implies' (rather than 'amounts to') a statement that events of the first kind are regularly accompanied by events of the second kind (see Hempel 1966: 53, 2001: 278).

[6] This model was strongly informed by positivist suspicion of metaphysics and of concepts, such as that of causation, which might appear to open the door to metaphysics (Cartwright 2002: 4–5). Although Hempel acknowledged that explanation is 'one of the primary objectives' of science (1966: 47; see also 2001: 276; Hempel and Oppenheim 1948: 135) and that scientific explanations are often *framed* in causal language (Hempel 1966: 47; see also see 1942: 36, 2001: 278), he sought to show that this language is *only a framing* (see Dessler 1991: 344; Ruben 1993a: 2).

to provide information about what *brought them about* (see, for example, Lewis 1993: 185; Lipton 1993: 207; 2004: 30; Kim 1993: 233; Woodward 1993; Little 1995; Salmon 1998: 69).

We will adopt the second of these approaches. This reflects several strands of our argument up to this point. In the Introduction to this book, we argued that the verb 'to cause' *means* 'to produce' or 'to bring about' and that the concept of 'cause' is irreducible: causal statements cannot be wholly reduced to non-causal statements. In Chapter 1, we therefore rejected the 'semantic' version of the regularity theory of causation, according to which the verb 'to cause' means the same as 'to be regularly followed by'. In Chapter 3, we noted, with approval, that one of the key features of Van Fraassen's defence of causal agnosticism is that he acknowledges that the concept of cause is irreducible. Rather than seeking to deny this, he insists that this *semantic* fact does not have any intrinsic *ontological* consequence: the fact that we, as human beings, cannot do without causal language does not mean that causation really is part of the world existing independently of human thought.[7] In our view, therefore, there is no reason to shy away from the intuitive idea that a causal explanation provides information about how an event, or set of events, was brought about—information about its causes.[8]

Returning to the question of what makes a causal explanation satisfactory, two preliminary points should be noted. First, what is causally explained is the occurrence of a specific event, or set of events, located in time and space. In order to provide information about how a given event was brought about, this event must have occurred, and it must have been picked out as an event, located in time and space.[9] Second, and consequently, a causal explanation will typically be provided by means of an explanatory statement—that is, a concrete causal statement relating to specific events which have already occurred. As we argued in Chapter 4, the distinctive epistemic function of such a statement is to express a causal explanation. And, as we also argued

[7] Hence, although the first of the two broad approaches to causal explanation was motivated by an anti-realist suspicion of metaphysics, the second does not require causal realism. One of the ways in which Van Fraassen's approach departs from earlier forms of anti-realism, such as Hempel's, is that Van Fraassen accepts that the concept of cause is irreducible, but regards this as a semantic issue which does not have any direct metaphysical implications.

[8] Whereas Jackson (2017) describes a causal explanation as providing a 'recipe' for producing an outcome of a particular kind (see also Jackson forthcoming), we distinguish sharply between a causal explanation, which relates to events that have already occurred, and a policy prescription.

[9] As Mink points out, historical 'events' are not ontologically distinguished as discrete 'events'; the historian determines the contours, in space and time, of the 'events' they include in their narratives and gives them a relevant description (1987: 199–201). We believe that this is also true in the context of causal explanation. Unlike Mink, however, we do not believe that in this respect history is any different from natural science.

in Chapter 5, the kind of evidence which will persuasively support such a statement is evidence that rules out competing explanatory statements.

It is crucial to recognize, however, that a causal explanation is not a 'mere explanatory statement', but rather an explanatory statement put forward in a specific context and for a specific purpose (see Achinstein 1977: 1). As Lewis puts it: 'In an act of explaining, someone who is in possession of some information about the causal history of some event … tries to convey it to someone else' (1993: 185; see also Achinstein 1983: 63; Suganami 2008: 328; Jackson 2017: 696). In short, a causal explanation is (implicitly or explicitly) addressed to an audience. For a causal explanation to be required, the intended audience must be puzzled, to some extent, about how the event(s) to be explained were brought about. For a causal explanation offered in response to this puzzlement to be satisfactory, the information it provides about the causes of the event(s) to be explained must be relevant to this puzzle: it must help to resolve the audience's puzzlement.

We term this—the requirement that the information provided in a causal explanation be *relevant* to the needs of an intended audience—the 'pragmatic' component of causal explanation (cf. Van Fraassen 1980; Achinstein 1993b).[10] It is crucial because it shapes the information an explanatory statement must convey in order for it to be satisfactory as an explanation.[11]

The pragmatic component of causal explanation

To be satisfactory, a causal explanation must fulfil two basic requirements (see Humphreys 2017b). First, the information it provides must be *reliable*: explanatory statements which are offered as causal explanations must be supported by empirical evidence. We showed, in Chapter 5, that the kind of evidence which will provide persuasive support for an explanatory statement is evidence which rules out competing explanatory statements. This is not well understood. Indeed, one symptom of the relative neglect of causal explanation as a topic in IR is the lack of attention paid to how explanatory statements can be supported empirically. We hope that it is, nonetheless, uncontroversial to insist that when an explanatory statement is offered as a causal explanation it must, among other things, be reliable.

[10] It is 'pragmatic' insofar as it relates to human interests: it concerns the relationship between an explanation and what its intended audience is puzzled about (see Van Fraassen 1980: 89–91).

[11] Hence, we are not just saying that the *act of explanation* is a pragmatic one—that when one offers an explanation, one does so with an audience in mind—but rather that the *content* of a causal explanation is always shaped, in part, by pragmatic considerations.

If an explanatory statement is considered in isolation, then it makes sense to focus on its reliability as a measure of how satisfactory it is. However, when an explanatory statement is offered as a causal explanation, a second requirement arises: as well as being reliable, the information it provides must be *relevant*.

It is crucial to recognize that even if an explanatory statement is persuasively supported by empirical evidence, and hence reliable, it may nevertheless fail to tell the intended audience what they want to know; it may fail to resolve their puzzlement about the occurrence of the event(s) purportedly being explained. As Lewis points out (1993: 194), even if the information being provided is reliable, it may be disorganized, or lack sufficient detail. Alternatively, it may be 'stale news', adding 'little or nothing to the information the recipient possesses already'. Even if the information provided is new, detailed, and well organized, it may still not be information 'of the sort the recipient wants'. For example, it might be addressed to an aspect of how the event of interest was brought about that is different from the one the intended audience is puzzled by: it might be information which would help resolve a *different* puzzle that some *other* audience might have about the occurrence of this event, yet fail to address, let alone resolve, the puzzle which the *intended* audience has.[12]

What this highlights is the inherent 'contextuality of explanation' (Suganami 2008: 337). The information that will have to be provided in order to make a causal explanation of a given event satisfactory 'differs from context to context' (Van Fraassen 1980: 156): it depends, in part, on the interests of the intended audience—on what they already know and what they find puzzling.[13] In other words, whether a causal explanation is satisfactory is always, in part, a pragmatic question.[14]

[12] Although it is rarely discussed, the same considerations arise when someone offers a causally informed prediction, '*a* will cause *b*'. Just as the explanatory statement '*a* caused *b*' identifies one of the many causes of *b*, so the prediction '*a* will cause *b*' identifies one of the many consequences of *a*. From the audience's point of view, *b* may not be the consequence they are interested in. Which consequences are of interest, and which constitute mere 'side effects' (see Jervis 1997: 10), is a wholly pragmatic matter.

[13] One way of clarifying what a particular audience finds puzzling about the occurrence of an event of interest is to frame their (actual or surmised) request for an explanation in contrastive terms. If what is requested is not simply an explanation of why *x* occurred, but rather an explanation of why, given the background conditions, *x rather than z* occurred, then this will help clarify what seems puzzling about the occurrence of *x* and will hence clarify what information about the causes of *x* is likely to be most informative (see Van Fraassen 1980: 97–157; Lewis 1993: 196–7; Lipton 1993). This idea has been picked up in a number of recent discussions of causal explanation in IR (see, for example, Grynaviski 2013; Humphreys 2017b; Jackson 2017: 700; Norman 2021; see also Morgan and Patomäki 2017).

[14] We acknowledge that this pragmatic requirement, combined with the virtual impossibility of meeting, with absolute certainty, the epistemic requirement of reliability on evidential grounds, means that the credibility of any causal claim must rest, to a certain extent, on the judgement of whoever is advancing it (see Chapter 7). However, we believe that Ned Lebow goes too far when he asserts that causal claims are necessarily 'rhetorical at best' (2020: 310); indeed, we will warn against a discernible tendency, even in

Recognizing that there are always these two dimensions—reliability and relevance—on which a causal explanation can be evaluated is especially important when we are confronted by 'alternative' explanations for the same event. In this situation, one possibility is that these explanations are all addressed to the same puzzle, but only one of them is *reliable*—only one is adequately supported by empirical evidence. However, a second possibility, which is often encountered in IR, is that, reliable or not, these explanations are addressed to different puzzles.

We illustrated this latter possibility in Chapter 5, when we considered a series of 'alternative' explanatory statements concerning the outbreak of World War I:

a) the rigid alliance structure which took shape in the late nineteenth century caused the war
b) the assassination of Archduke Franz Ferdinand on 28 June 1914 caused the war
c) Austria's ultimatum to Serbia on 23 July 1914 caused the war
d) Russia's partial mobilization against Austria on 28 July 1914 caused the war

As we pointed out, these statements appear to complement, rather than to compete with, one another: it is prima facie plausible to believe that *all* the factors they identify contributed causally to the outbreak of World War I. If so, then each of these statements may turn out to be reliable, yet they may not be equally valuable if offered as explanations. This is because the information they provide may not be equally *relevant*: some may do a better job than others in addressing a particular audience's puzzlement about the outbreak of the war.[15]

To see how causal inquiry is liable to go wrong if we fail to recognize this, consider the long-standing debate about the causal mechanisms which underpin the so-called 'democratic peace'. Numerous studies have analysed and compared 'normative' and 'institutional' (or 'structural') explanations for the empirically observed democratic peace (see, for example, Maoz and Russett 1993; Owen 1994; Risse-Kappen 1995; Chan 1997; Ray 1998; Rosato 2003; Doyle 2005; Hayes 2012). However, it would be an obvious mistake to

high-quality research outputs in IR, to express causal claims with unwarranted 'rhetorical flourish' (see Chapter 8).

[15] Dray (1978) and Suganami (1996) discuss the different kinds of things we might be interested in when we seek an explanation of something like 'the outbreak of a war'.

suppose that if democratic norms contributed causally to some instance of democratic peace, then democratic institutions could not have done so, and vice versa. To reason in this way would be to preclude the obvious possibility that these factors both contributed.[16] It is crucial to recognize that if a given causal explanation emphasizes the contribution of one particular factor (say, democratic norms) to an outcome of interest, this does not imply that another factor (say, democratic institutions) made no contribution. The decision to focus on the contribution made by one particular factor will often be a pragmatic choice, informed, perhaps, by the neglect of this factor in other explanations.

The underlying point is that when we offer an explanation for the occurrence of some event of interest, we are always responding to a real or hypothetical puzzle about why it occurred; when we are puzzled about why some event occurred, this is usually because we have failed to recognize one or more of its causes; causal inquiry therefore often highlights and provides evidence for causes which have not been adequately examined before. But if a particular explanation stresses the contribution of one particular cause, this does not imply that there were no other causes; the decision to focus on one cause rather than another is a pragmatic one.

II. Taking the pragmatic component of causal explanation seriously

When we realize that a satisfactory causal explanation must be relevant as well as reliable, and hence appreciate that it is necessary to take the pragmatic component of causal explanation seriously, this helps to resolve a number of potential questions about the content of causal explanations and how they are presented. In this section we will consider three such questions: whether a satisfactory causal explanation must be 'complete'; what forms causal explanations can take; and whether causal explanations must identify causal mechanisms. The overarching theme of our discussion will be that pragmatic considerations play a much more substantial role in shaping the content of causal explanations and the form in which they are presented than is often realized.

[16] It would be an even more egregious mistake to suppose that it is necessary to identify a single explanation for the whole of the empirical phenomenon known as the 'democratic peace'—that is, for every instance of peaceful relations among pairs of democracies (Suganami 1996: 107).

Why the notion of a 'complete' causal explanation is misguided

It is sometimes suggested that causal explanations should be complete in the sense that they identify all the causes of the event(s) being explained. According to John Gerring, for example:

> Wherever a study focuses on a single outcome, the reader quite naturally wants to know everything, or almost everything, about the causes of that outcome (leaving aside the obvious background factors that every causal argument takes for granted). Thus, single-outcome studies usually seek to develop a more or less 'complete' explanation of an outcome, including *all* causes that may have contributed to it. (2006: 716; see also Waldner 2007: 146; Beach and Pedersen 2013: 18–21)[17]

This raises an obvious question: how is it to be determined what is a 'background factor', which can be taken for granted, and what is a 'cause', about which the reader will want to know?

The challenge is outlined by David Lewis (1993: 182–3). He argues that '[a]ny particular event that we might wish to explain stands at the end of a long and complicated causal history'. To illustrate this, he describes a car crash which is known to have been caused by 'the icy road, the bald tire, the drunk driver, the blind corner, the approaching car, and more'. He contends that although an explanation for the crash may well focus on one or more of these causes, these 'are by no means all the causes of the crash'. Causal histories extend backwards in time: 'each of these causes in turn has its causes; and those too are causes of the crash. So in turn are their causes, and so, perhaps, *ad infinitum*'. Moreover, causal histories are dense: each of these causes can be broken down into its constituent parts and described in more and more detail, as can the interconnections between them (see also Lipton 2004: 32–3).

Four important points emerge out of Lewis's argument. First, the causal histories of *all* events extend almost endlessly back in time and *every* cause can, potentially, be described in ever greater detail. Second, although it is better, other things being equal, to provide more information about a causal history rather than less (Lewis 1993: 194), *any* account of a causal history will be 'partial' (1993: 204): it is always possible to provide more detail and/or to go further back in time. Third, offering a causal explanation therefore

[17] Gerring is expressing a common sentiment here, but in later work he backtracks and wonders 'what a complete explanation might consist of, given the problem of endless causal regress' (2017: 227).

requires a choice to be made about which causes to focus on and which to treat as taken-for-granted background conditions; it also requires a choice about how much information to provide about the causes which are focused on. Fourth, and most importantly, there is no objective basis for these choices. As Lewis puts it: 'There is no one right way … of carving up a causal history' (1993: 183).

What Lewis is describing is sometimes termed (the problem of) causal 'selection' (see, for example, Lewis 1975: 182–3; Martin 1982: 55–6; Hart and Honoré 1985; Hesslow 1988; Schweder 1999; Carr 2002: 97–100; Lipton 2004: 47–8; Schaffer 2005; Broadbent 2008), this term highlighting the fact that a causal explanation must focus 'selectively' on certain causes of the event(s) being explained, with other causes being treated as taken-for-granted background conditions (see also Dray 1978). In some cases, however, this feature of causal explanations can be easy to overlook.

Often, it will be obvious that certain causes of the event(s) being explained are of particular interest. These may include: causes which are unusual and hence illuminate why an event of the kind being explained (such as the car crash) occurred in the scenario of interest but not in other scenarios; causes which could potentially have been manipulated, changing the outcome; and causes which allow judgements of culpability to be made (Lewis 1993: 183; see also Hart and Honoré 1985; Dray 1978; Mackie 1980: 117–42). When it is obvious that causes of a particular kind are the most relevant, the decision to focus on them may appear so 'natural' that the 'selectiveness' of the resulting explanation is obscured. Moreover, as we noted in Chapter 5, it is common for people to describe a cause that is of particular interest to them as '*the* cause' of the event(s) being explained, thereby obscuring the complexity of the histories from which such causes are selected (Lewis 1993: 183). It is important to recognize that no matter how 'natural' the selection may seem, the causes identified in a causal explanation have always been selected from a denser and more extensive causal history.

It follows from this that 'completeness' is an impossible standard, at least if the demand is that a causal explanation must identify 'all causes' of the event(s) being explained. It is also undesirable. As Suganami has pointed out, in providing a causal explanation 'it is necessary to assume a certain level of understanding' on the part of the audience: 'we cannot explain anything to someone who understands nothing' (2008: 334). A satisfactory explanation helps fill gaps in the audience's understanding, but this presumes that some things are already understood and can hence be left untouched: some information about the causes of the event(s) being explained does *not* need to be provided (Suganami 2008: 337; see also Hanson 1955: 293). In fact,

trying to include everything may undermine the capacity of an explanation to resolve the audience's puzzle, for the information that is most relevant to this puzzle may get lost in the noise. We might hope that a causal explanation will prove *completely satisfying* to the intended audience, fully resolving their puzzlement, but no more stringent demand for 'completeness' is sustainable.

The forms that causal explanations can take

We have characterized a statement of the form '*a* caused *b*' as the paradigmatic form of an explanatory statement. As we pointed out in Chapter 4, however, concrete causal statements, of which explanatory statements are a subset, can take a wide variety of linguistic forms. It is important to appreciate that this applies equally to causal explanations. Moreover, and contrary to common supposition, choices about the form in which a particular causal explanation is presented are wholly pragmatic.

One common misapprehension, especially in the mainstream research methods literature (see, for example, Van Evera 1997: 15, 40; Gerring 2006; Beach and Pedersen 2013: 18–21, 63–7), is that causal explanations necessarily relate to 'single' or 'individual' events.[18] This is not true. The objects of causal explanations must be specific events which have occurred, and which have been picked out as events for the purpose of explanation, but no further a priori restriction on which events can be objects of causal explanations is warranted. Let us elaborate.

The examples of causal explanations which are often used for illustrative purposes are, by design, relatively simple; the objects of these explanations are therefore often 'individual' events. Examples we have discussed previously include 'the short circuit caused the fire' and 'David's smoking caused his lung cancer'. A typical example from IR would be 'German expansionism caused World War II' (Van Evera 1997: 16).[19] However, explanations might also be sought for multiple events that have been aggregated together for the purpose of explanation. For example, someone might request an explanation not for the occurrence of 'the fire', but rather for the occurrence of 'the multiple fires (which occurred within a defined area and a defined time period)'. These events may be individually named, or they may be picked out under some generic description. For example, an explanation might be sought for the occurrence of 'World Wars I and II' or for the occurrence of 'the World Wars' or even for the occurrence of 'the major wars of the twentieth century'.

[18] This is intimately connected to the frequent characterization of concrete causal statements as 'singular', which we criticized in Chapter 4.

[19] Note that that the 'individual' events which are the objects of many causal explanations can and do vary widely in scale, from the proverbial 'loss of a nail' to the 'loss of a kingdom'.

A causal explanation might also be sought for the occurrence of a kind of event which is not straightforwardly either an 'individual' event or an aggregation of multiple events, viz. an 'emergent' event: an event which *emerges out of* and yet is *conceptually irreducible to* an identifiable set of other events (see Benton 1977: 85; Jervis 1997: 12–17; Wendt 1999: 1; Harrison 2006b: 7; Kurki 2008: 164; Cudworth and Hobden 2012: 168–9; Gunitsky 2013: 39).[20] This category includes statistically summarized population-level outcomes, such as the reduced birth rate in Japan in 1966 (an example we considered briefly in Chapter 4), the near-absence of war between democracies since 1945 (see, for example, Reiter 2017), or the increase in the use of bilateral investment treaties in the 1990s (see Skovgaard Poulsen 2014). It also includes 'system-level' properties of social systems and of how they are organized and sustained, such as structures, rules, and collective identities (see, for example, Wendt 1999).[21]

Whether the object of a particular causal explanation is an individual event, an emergent event, or multiple events aggregated together is a pragmatic issue: it is determined by the inquirers' interests. Of course, each event that features in a causal explanation must have been 'singled out' or 'individuated' as an event, but this should not mislead us into supposing that only 'single' events, let alone single events at a particular 'level', can be objects of causal explanations.

Given what we have said about the problem of causal selection, it should be clear that pragmatic considerations not only determine which events become the objects of causal explanations, but also determine which of the causes of these events are included within causal explanations. However, it is worth being aware of the confusions that sometimes arise in relation to this issue.

One area of confusion concerns how complex a causal explanation should be. As we have noted, many of the most common examples of causal explanations take the form '*a* caused *b*': they identify a single event, *a*, as the cause of another event, *b*, which is being explained. Yet, as Lewis points out, there are many other possibilities:

> Most simply, an explainer might … specify one of the causes of the explanandum. Or he [*sic*] might specify several. And if so, they might comprise all or part of a cross-section of the causal history: several events, more or less simultaneous and

[20] An oft-used illustration of the idea that 'societal' phenomena *emerge out* of the activities of society members, yet are *conceptually distinct from them*, is Durkheim's analysis of social suicide *rates*, as distinct from individual cases of suicide. He claimed that suicide *rates* remained relatively stable within each society (when at peace) but differed significantly across societies (Durkheim 1952).

[21] Although we refer here to 'levels', unlike some scientific realists (see, for example, Wight and Joseph 2010: 12) we do not suppose that there are, objectively speaking, 'levels' to reality; rather, we employ the term as a metaphor (see Patomäki 2002: 130), indicating the scale at which events are picked out for explanation.

> causally independent of one another, that jointly cause[d] the explanandum. Alternatively, he might trace a causal chain. He might specify a sequence of events in the history, ending with the explanandum, each of which is among the causes of the next. Or he might trace a more complicated, branching structure that Is likewise embedded in the complete history. (1993: 187)

How much of the complexity of the causal history of the event(s) being explained to incorporate into a causal explanation is a pragmatic judgement. It would be incorrect to assume that, in general, more or less complex causal explanations are to be preferred.[22]

A related area of confusion concerns the 'level' at which causal explanations should focus. In broad terms, this relates to the so-called 'levels-of-analysis' debate (see Waltz 1959; Singer 1961; Hollis and Smith 1991a, 1991b; Wendt 1991; Suganami 1996), but a particular question which has arisen in IR over the last couple of decades concerns the 'level' at which 'emergent' events should be explained. Many of those who emphasize the significance of such events suggest that these events should be analysed and explained at the population- or system-level: they prioritize the search for system-level causes of system-level outcomes (see, for example, Jervis 1997: 13; Harrison 2006a, Wendt 1999; Leon 2010; Cudworth and Hobden 2012; see also Waltz 1959, 1979).[23] It is important to recognize, however, that any decision to focus specifically on the system-level causes of system-level outcomes is a pragmatic one. It may well be that system-level causes of emergent events have been neglected in IR and hence that significant value can be found in seeking to understand them better. Yet, this is a pragmatic judgement, just as it is a pragmatic judgement what 'level' of event to seek to explain in the first place.[24]

A final issue concerning how causal explanations are presented is whether they should take a narrative form. There is an extensive debate in the philosophy of history concerning the role of narrative in historical understanding and how this relates to the aim, which not all historians share, of providing causal explanations (see, for example, Mink 1966, 1970, 1987; Dray 1971;

[22] Despite what complexity theorists sometimes suggest (see, for example, Harrison 2006b: 12), it does not follow from the fact that a social system is complex that explanations of events within that system must also be complex. In a more complex environment, it may be more difficult to identify evidence which persuasively supports a given explanatory statement, but it is a further question how complex an explanation will have to be to satisfy its intended audience: this will depend entirely on what it is that this audience finds puzzling about the occurrence of the event(s) being explained.

[23] In this sense, a commitment to the study of 'emergent' properties or events often stands in contrast to a commitment to the study of 'micro-foundations' (see, for example, Elster 1989; Little 1995).

[24] Singer's original formulation of the so-called 'levels-of-analysis' problem conflates these two issues, viz. what 'level' of event to seek to explain and what 'level' to focus on when constructing explanations (see Suganami 2009: 374).

White 1973, 1984; Stone 1979; Ankersmit 1986; Carr 2008). This debate has sometimes made its way into IR (see, for example, Campbell 1998c; Suganami 1999, 2008; Roberts 2006; Kurki and Suganami 2012; Lawson 2012), though that need not detain us here. The point we wish to make is simple: the choice of whether to present a causal explanation in the form of a narrative is a pragmatic one. We recognize, of course, that different scholars, operating in different contexts, and addressing different audiences, are likely to make different choices.[25] Yet, whatever choice is made on this issue, it remains necessary to ensure that persuasive evidence is provided in support of any claim that a particular factor *caused* the event(s) being explained.

Causal explanation and causal mechanisms

Finally, recognizing the pragmatic component of causal explanation also sheds light on the real nature of the oft-heard demand that causal explanations should identify causal mechanisms (see, for example, Elster 1989; Kiser and Hechter 1991: 5; Little 1991, 1995; Bunge 1997: 455; Wendt 1999: 81; Machamer, Darden, and Craver 2000: 1; Johnson 2006: 235; Wight 2006a: 32; Waldner 2007: 146; Beach and Pedersen 2013: 18–21).[26] We showed, in Chapter 2, that Bhaskar's contention that causal mechanisms are real is unpersuasive: how could we know that a mechanism, identified in a causal explanation, really is part of the world existing independently of human thought? Yet, this raises an important question: if Bhaskar's position is unsatisfactory, then how should the widespread call for causal explanations to identify mechanisms be understood?

It is helpful to start by noting that a 'concern with causal mechanisms' often arises out of a contrast 'between correlational analysis and causal analysis' (Mahoney 2001: 577; see also Gerring 2008: 163–4). For example, David Dessler's (1991) critique of the Correlates of War project trades on a contrast between 'two explanatory formats … correlational and causal'. Dessler argued that the project's limitations stemmed from its focus on developing correlational knowledge; he contrasted this with 'integrative reasoning in natural science', which seeks 'to identify the mechanisms through which specified outcomes occur, when they do' (1991: 346, 343). Similarly, Falletti and Lynch (2009: 1146) argue that identifying causal mechanisms serves 'to open the

[25] History and IR have different disciplinary norms around these kinds of issues. However, it would be a mistake to infer from this that fundamentally different accounts of causal inquiry and causal explanation are required for each discipline (see Suganami 2008).

[26] Sometimes, the term 'social mechanism(s)' is preferred (see, for example, Hedström and Swedberg 1998; Guzzini 2012).

black box of lawlike probability statements that simply state the concurrence or correlation of certain phenomena or events'.[27] In such accounts, the term 'mechanism' means something closely akin to 'cause' (see Kurki 2008: 233), and the call for causal inquiry to focus on identifying mechanisms amounts to a call for causal inquiry to identify causes and not just correlations. This is reflected in definitions of 'mechanisms' as 'processes in the world that generate outcomes' (or similar) (Bennett 2013: 465; see also Bunge 1997: 414; Goldthorpe 2001: 9; Wight 2006a: 34; Waldner 2007: 153; Imai, Keele, Tingley, and Yamamoto 2011: 765) and in the contrast that is often drawn between the covering-law model of explanation and explanation by reference to mechanisms (see, for example, Wendt 1999: 81; George and Bennett 2005: 127–49).

There is, however, more to the demand that causal explanations should identify causal mechanisms than the simple idea that explanations should identify causes. A common theme of such demands is that causal explanations should provide knowledge of 'how' events were brought about and not just of 'why' they occurred (see, for example, Kiser and Hechter 1991: 5; Falletti and Lynch 2009: 1147; Imai, Keele, Tingley, and Yamamoto 2011: 765; cf. Elster 1989: 4); the implication is that causal mechanisms provide knowledge of 'how', as distinct from 'why'. In our view, this demand for knowledge of 'how' and not just 'why' events were brought about is best understood as a *pragmatic* demand: a demand for 'more fine-grained explanations' (Johnson 2002: 230; see also Checkel 2006: 363; Brady and Collier 2010: 317).

In order to see this, it is important to appreciate that, in causal inquiry, information about 'how' is qualitatively indistinguishable from information about 'why': whether a piece of causal information *counts* as information about 'how'—as information about 'mechanisms'—depends on the *context* in which it is provided.

Suppose someone asks 'why' World War II occurred and they are told 'German expansionism caused World War II'. They then request more detailed information about 'how' German expansionism brought about World War II and are told 'German expansionism caused German aggression, causing World War II'.[28] In the context, this further statement might well be described as providing information about the 'mechanism' through which German expansionism caused World War II. Suppose, however, that someone asks 'why' World War II occurred and they are told 'German expansionism caused

[27] The motif of opening, unpacking, or getting inside the (black) box features quite widely in discussions of causal mechanisms (see, for example, Elster 1989; Salmon 1998: 77; Goldthorpe 2001: 9; Gerring 2008; Imai, Keele, Tingle, and Yamamoto 2011).

[28] The example comes from Van Evera (1997: 15–16)

German aggression, causing World War II'. Here, the same information is provided as in the first case, but here it is packaged as information about 'why' rather than as information about 'how'. Now suppose that someone requests more detailed information about 'how' German aggression caused World War II. If provided, this *further* information might well be described as information about the 'mechanism' by which German aggression brought about World War II. This illustrates that it is the context in which information is provided, and not its content, which makes it information about 'how' an event of interest was brought about, and hence about 'causal mechanisms': what makes it information about 'mechanisms' is that it gives *more detail* about how the causal process of interest unfolded than had previously been provided.

We contend, therefore, that whether a specific piece of information, provided as part of a causal explanation, counts as information about 'mechanisms' depends on what information has previously been provided.[29] A request for information about 'causal mechanisms' is a request for *more detailed information* than has so far been provided. The request for information about mechanisms is, in short, a pragmatic one: it indicates that the information that has so far been provided does not yet resolve whatever is puzzling about the occurrence of the event(s) being explained.

References to 'underlying' mechanisms (see, for example, Little 1991: 15; Johnson 2006: 246; Wight 2006a: 32; Joseph 2007: 346; Bennett 2013: 467) should therefore be understood as metaphorical: when information is provided about what, in the context, are characterized as 'mechanisms', this more detailed information may be said to 'underpin' the less detailed information that has previously been provided or to give greater 'depth' to the resulting explanation, but we see no reason to interpret this language in ontological terms, as revealing something about 'levels' of reality in the world existing independently of human thought. In fact, it is important to recognize that the request for more detailed information could, potentially, continue ad infinitum (see Machamer, Darden, and Craver 2000: 13; Goldthorpe 2001: 9; Johnson 2006: 249; Wight 2006a: 36; Bennett 2013: 467); we never reach bedrock, but only exhaust our capacity to generate further information (and, in all likelihood, our interest in it).

Insofar as the demand that causal explanations should identify 'causal mechanisms' is a demand that such explanations should identify causes, or a demand for *more information* about how the event(s) being explained came

[29] We therefore reject the distinction that is sometimes drawn between 'causal description' and 'causal explanation', according to which a causal explanation is distinguished by the fact that it provides information about mechanisms (see Shadish, Cook, and Campbell 2002: 9, 11; see also Jackson 2017).

about, we endorse it: causal explanations should, as we have argued, identify causes, and it is always appropriate to request more information if a causal explanation fails to resolve our puzzlement. The idea that a causal explanation is only satisfactory if it provides information about mechanisms reflects the fact that most puzzles which motivate requests for explanations will be resolved only through the provision of more detailed information than was previously available. That said, we have not used the term 'mechanism' in developing our account of causal explanation. Let us briefly explain why this is.

One obvious indicator that the term 'mechanism' may be problematic is that numerous competing definitions have been offered (see Mahoney 2001: 579–80; Gerring 2008; Hedström and Ylikoski 2010: 51). Some of these are more plausible than others. For example, whatever a 'mechanism' is, we are quite sure that it is not itself a kind of 'explanation' (see Gerring 2008): explanations may invoke mechanisms, but they are not themselves mechanisms. In our view, however, the key problem with the term 'mechanism' is that we can request more information about how *any* causal process unfolded, and yet the term that is often used to frame this request, viz. as a request for information about '(causal) mechanisms', has some potentially quite misleading connotations, especially in view of the close linguistic association between a 'mechanism' and the working of a 'mechanical device'.

One initial point to note is that many of the causes which are likely to be of interest in a discipline such as IR have nothing whatsoever to do with the operation of mechanical devices: these include rules, norms, ideas, reasons, and discourses, as well as social structures (see Kurki 2008: 11). Yet, there is also a deeper point: many discussions of causal mechanisms elide the important distinction between a mechanical device and the causal processes which contribute to its operation (see Suganami 1996: 164). Consider, for example, Hedström and Ylikoski's insistence that 'there is nothing in the notion of a mechanism that would imply that it is by definition unobservable', and an example they give of observable mechanisms: those 'constituting an automobile's engine' (2010: 50–1). This example concerns the (observable) parts of a mechanical device—an engine—rather than the (unobservable) causal processes taking place within it. It is crucial to recognize, however, that when we ask for information about '(causal) mechanisms', it is (unobservable) causal processes we are interested in (see Patomäki 2002: 130; Bunge 2004: 183; Wight 2006a: 34; Kurki 2008: 233–4). The term 'mechanism' not only fails to clarify this important point, but actively muddies the water.[30] Hence, while

[30] Admittedly, those who frame causal inquiry as a search for mechanisms often warn against associating 'mechanisms' too closely with the operation of mechanical devices (see, for example, Machamer,

we endorse the idea that it is often appropriate to request more information about the unfolding of a causal process of interest, we mostly refrain from using the term 'causal mechanism'.

III. Other red herrings in discussions of causal explanation in IR

So far, our focus has been on the importance of the pragmatic component of causal explanation and on correcting the misunderstandings that can flow from its neglect. However, it is also worth highlighting a further series of red herrings that often emerge when causal explanation is discussed in IR. These stem from a failure to recognize that a causal explanation is an explanatory statement—a concrete causal statement relating to specific events which have occurred—and from a failure to recognize that *all* explanations of how specific events were brought about are *causal* explanations.

Causal explanations and causal theories

In IR, mainstream theories such as neorealism and neoliberalism are often described as 'explanatory' theories (Smith 2000: 380; see also Humphreys 2011). They are also often described as providing 'explanations'. To give just one well-known example, Kenneth Waltz defined theories as statements which 'explain laws', adding that this definition of the term 'theory' 'satisfies the need for a term to cover the explanatory activity we persistently engage in' (1979: 5–6). More broadly, philosophers of science often identify 'explanatory power' as a key virtue of scientific theories.[31] This suggests that 'explaining' may be a key part of what theories do. Yet, this apparently commonsensical idea is something of a red herring. Let us explain.

If asked for a causal explanation, it will not be sufficient to respond simply by stating a causal theory, even a well-supported one. As we have pointed out, theories are propensity statements: they abstract from specific events to describe the causal relationship that will obtain between events of particular kinds if the conditions are right—that is, if all interference with the unfolding of this relationship is prevented. By contrast, a causal explanation provides information about what actually brought about some specific

Darden, and Craver 2000: 2), but this merely confirms the point we are making: the term has unhelpful connotations!

[31] There is considerable debate in the philosophy of science about what this signifies (see Van Fraassen 1980: 87–96). For our purposes, the crucial point is that in the philosophy of science, as well as in IR, 'theories' are often described as 'explaining'.

event (or set of events) located in space and time. There is therefore a big and important difference between stating a causal theory and providing a causal explanation. If someone were to ask for an explanation of World War II, it would not be sufficient simply to state the theory 'expansionism causes war', as if this constituted an explanation. Even if this theory were well supported, it would indicate only that there is a *propensity* for episodes of expansionism to produce wars, not that an episode of expansionism in fact contributed to the bringing about of World War II.

We recognize, of course, that causal explanations may sometimes be generated by *applying* a well-supported causal theory: we discussed, at the end of Chapter 5, how this may be done. As we pointed out, however, in order to generate a reliable explanation by applying a causal theory it is necessary not only that the theory should be well supported and that events of the relevant kinds should have occurred, but also that the conditions under which these events occurred should be conditions under which it is known that the propensity described by the theory unfolds fully. In order to generate a reliable causal explanation of the occurrence of World War II by applying the theory 'expansionism causes war', it would be necessary not only for this theory to be well supported and for World War II to have been preceded by an episode of expansionism, but also for the conditions under which this episode occurred to be conditions under which it is known that the propensity for episodes of expansionism to produce wars unfolds fully. Moreover, even if such an explanation were reliable, it may not be pragmatically adequate: it may not provide information which helps to resolve whatever puzzle motivated the request for an explanation of World War II.

With this in mind, let us return to the idea that theories can 'explain'. One way of interpreting this is as shorthand for the claim that causal explanations can be generated by applying well-supported causal theories, subject to the proviso outlined above. Yet, if the claim that theories 'explain' is used in this way, it will be essential not to lose sight of the crucial distinction between (i) possessing a well-supported causal theory and (ii) possessing a satisfactory causal explanation of the occurrence of the event in question. To suppose that if we possess a well-supported causal theory, '*A* causes *B*', we also, thereby, possess an explanation for the occurrence of any *B*-type event of interest, or even for the occurrence of any *B*-type event which follows an *A*-type event, would be a serious error.

There is also a somewhat different sense in which theories may be said to 'explain': they are said to explain not 'events', but rather 'laws'. As noted earlier, this is the view famously propounded by Kenneth Waltz in *Theory*

of International Politics (1979), but it is also defended by Stephen Van Evera, who spells the idea out in more detail.[32] Van Evera argues that a theory states a causal law (or hypothesis), the paradigmatic form of which is '*A* causes *B*', and also provides 'an explanation' of this law which 'explicates how *A* causes *B*'. On his account, this kind of explanation takes the paradigmatic form '*A* causes *B* because *A* causes *q*, which causes *r*, which causes *B*' (1997: 9). Applying this formula to one of Van Evera's own examples, we might say that the theory 'expansionism causes war' states a causal law, viz. 'expansionism causes war', and also contains an explanation for this lawlike relation, viz. 'expansionism causes aggression, causing war' (1997: 15).[33]

We agree with Van Evera that if one is presented with a causal theory, '*A* causes *B*', it is always appropriate to ask for more information about how (or why) *A*-type events produce *B*-type events—about the process by which *A*-type events produce *B*-type events, when they do.[34] However, we consider it unhelpful to characterize a statement such as 'expansionism causes aggression, causing war' as an 'explanation', for while it *fleshes out* the theory 'expansionism causes war', it remains an abstract, theoretical statement: like the theory 'expansionism causes war', the statement 'expansionism causes aggression, causing war' abstracts from specific instances of expansionism and war.[35] Terming such a statement an 'explanation' risks subverting the crucial distinction between abstract propensity statements and concrete explanatory statements. In our view, a better characterization of such statements is suggested by Van Evera himself (see 1997: 9). Rather than being an 'explanation', the statement 'expansionism causes aggression, causing war' *explicates* the theory 'expansionism causes war' by providing more information about the process through which this propensity unfolds, when it does.[36] While it may sometimes be valuable to explicate a theory in more detail, it is crucial to recognize that providing this kind of 'explication' is very different from providing an explanation of the occurrence of a specific event.

[32] This is also a recurrent theme in the philosophy of science (see, for example, Hempel 2001: 278).

[33] Van Evera characterizes this kind of 'explanation' as 'a theory's explanation' (1997: 12, 13, 19) in order to distinguish it from what he terms a 'specific explanation': an explanation 'that accounts for a distinctive event' (1997: 15).

[34] This highlights the fact that requests for further information about the unfolding of causal processes are not limited to the context of causal explanation, where they are often described as requests for information about 'causal mechanisms' (see above).

[35] This is reflected in Van Evera's characterization of 'expansionism causes aggression, causing war' as a 'theory' (see 1997: 17), even though, on his own account, it is (also) the 'explanation' contained within the theory 'expansionism causes war'.

[36] Machamer, Darden, and Craver (2000: 4) use the term 'explicate' in a similar fashion in relation to laws.

Causal explanations and regularities: The 'covering-law model of explanation'

A second red herring that often arises in discussions of causal explanation in IR is the idea that causal explanation requires knowledge of the empirical regularities associated with scientific laws.[37] David Dessler asserts, for example, that in order to explain 'a concrete occurrence or happening at a particular place and time' it is necessary to 'appeal to one or more laws, since without a knowledge of regularities and recurring patterns in the world, we would have no reason to expect particular happenings at particular times' (2003: 387). Although this kind of view is commonly encountered in discussions of causal inquiry in IR, it is badly misleading. There are two basic problems with it.

First, as we showed in Chapter 1, associational knowledge is not causal knowledge. Knowledge that there is a regular pattern in which *A*-type events are always or often followed by *B*-type events may often provide us with a reason to *suspect* that there may be a causal relationship between events of these two kinds, but there is a big and important difference between, on the one hand, pointing out that if an *A*-type event occurs and *A*-type events are regularly followed by *B*-type events, then a *B*-type event can also be expected to occur, and, on the other hand, providing an explanation of how a specific *B*-type event of interest was brought about. Second, in order to advance a reliable causal explanation, '*a* caused *b*', it is not *necessary* to know that there is a regular pattern in which *A*-type events are always or often followed by *B*-type events. As we showed in Chapter 5, the kind of evidence that will persuasively support an explanatory statement, '*a* caused *b*', is evidence that rules out competing explanatory statements—that is, statements to the effect that *b* was brought about without any contribution from *a*. If such evidence can be found, then the statement '*a* caused *b*' will constitute a reliable explanation for *b*, regardless of whether *A*-type events are regularly followed by *B*-type events.[38]

We recognize, however, that the view we have just outlined runs counter to the conventional wisdom in IR. It is therefore worth exploring in more detail the idea that knowledge of regularities may be required for the development of causal explanations.

The best-known articulation, in philosophy, of this idea is the so-called 'covering-law model of explanation' (hereafter CLM), according to which an event 'is explained by subsuming it under general laws, i.e., by showing that it

[37] This idea reflects the residual influence of the regularity view of causation, which we discussed in Chapter 1.

[38] As we discussed in Section I of this chapter, it is a further question whether this statement will be *satisfactory* as an explanation, given the interests of the intended audience.

occurred in accordance with these laws, by virtue of the realization of certain specified antecedent conditions' (Hempel and Oppenheim 1948: 136). This model was most fully elaborated by Carl Hempel (1942, 1965, 1966, 2001; see also Hempel and Oppenheim 1948).[39] He contended that a scientific explanation of the occurrence of a particular event (in his terminology: 'an event of kind *E*') requires knowledge only of 'general laws' and of the occurrence of particular events, and that it has the structure of a deductive argument with two premises and a conclusion (1942: 36; 1966: 50–1), each of which is empirically verifiable (1942: 38). This deductive structure is summarized in Table 6.1:

Table 6.1 The deductive structure of the CLM (covering-law model of explanation)[a]

Premise 1	A statement that a particular antecedent event (in Hempel's terminology: 'an event of kind *C*') occurred
Premise 2	A statement of one or more 'general laws expressing uniform empirical connections' (Hempel 1966: 50), these laws indicating that 'whenever' an event of kind *C* occurs 'an event of the kind to be explained [an event of kind *E*] will take place' (Hempel 1942: 36)
Conclusion (deduced from Premises 1 and 2)	A statement that the event to be explained ('an event of kind *E*') occurred

[a] In addition, Hempel discussed two other forms of CLM. In one, the object of scientific explanation is not a particular event but a subsidiary law (see, for example, 1966: 51; 2001: 278). We regard this as irrelevant to our purposes and hence set it to one side. In the other, the regularities used to explain the occurrence of particular events are probabilistic rather than deterministic (see, for example, Hempel 1966, 2001). We believe that our objections to the CLM apply *a fortiori* to its probabilistic version: if, as we shall argue, knowledge of deterministic laws is not in itself sufficient for causal explanation, then knowledge of probabilistic laws can hardly fare better.

The basic idea behind this model of causal explanation is that when a *C*-type and *E*-type event have each occurred and there is an established pattern 'whenever *C*, then *E*' which accords with one or more general laws (stated in Premise 2), then this pattern will 'cover' the *E*-type event of interest, thereby explaining its occurrence.

This model has, over the last sixty years, been the subject of extensive debates and criticisms in the philosophy of science (see, for example, Ruben 1993b) and is now largely discredited.[40] However, it still attracts considerable attention in IR theory (see Suganami 1996: 116–25; see also Smith 1996; Wendt 1999: 79–81; George and Bennett 2005: 132–5; Kurki 2008; Lebow 2014; Jackson 2017: 692–5), in part because its core ideas continue to exert a strong influence on how causal explanation

[39] The expression 'covering-law model' was, however, coined by Dray (1957: 1).

[40] The CLM has also been heavily criticized in philosophy of social science (see, e.g., Keat and Urry 1975; Benton 1977; Bhaskar 1998; Manicas 2006: 8–12; Winch 2008) and philosophy of history (see, e.g., Dray 1957; Gardiner 1959: 344–475; Mink 1967–8).

is conceived of in IR. While it is, on occasion, explicitly advocated (see, for example, Van Evera 1997: 40; Elman and Elman 2001b: 14), its continuing influence is better illustrated by some of the taken-for-granted ways in which its central ideas are invoked. For example, Hollis and Smith note 'the familiar dictum that science explains particular events by generalizing and making them cases of laws at work' (1991a: 3), James Lee Ray asserts that an explanation must 'demonstrate that the outcome of a particular event is … an example of an established pattern' (1995: 138), and Alexander George and Andrew Bennett assert that 'historical explanations use theoretical generalizations to argue why in a particular context certain outcomes were to be expected' (2005: 131). Such remarks present core tenets of the CLM as if they were a matter of simple common sense.

One criticism of the CLM that is sometimes advanced is that this model is *irrelevant to IR* because there are no known laws about world politics (see, for example, Kratochwil 1993: 66; Ruggie 1998: 861; Lebow 2014: 49; see also Little 1995: 42; Elster 2007: 35).[41] However, while Hempel's examples of general laws and of their deployment in scientific explanations are primarily drawn from natural science, he contended that the CLM is applicable to 'any … branch of empirical science' (1942: 39; see also 2001). When explanations do not *explicitly* invoke known laws, he argued, this is either because they have not been fully spelled out or because they constitute promissory notes contingent on a fuller elaboration and testing of the lawlike regularities to which they nonetheless appeal (see 1942; 2001). Hence, the fact that IR has not (yet) established unambiguous knowledge of 'general laws' does not provide any real challenge to the CLM as a model for how causal explanations should be developed in IR.[42]

The basic problem with the CLM is that it does not succeed in showing that knowledge of regularities *can* be drawn upon to explain the occurrence of specific events, let alone that causal explanations *require* such knowledge. To see this, consider, once again, the deductive argument laid out in Table 6.1. What this argument establishes is that the occurrence of an *E*-type event

[41] On the other hand, it has also sometimes been supposed, in IR, that what is required to explain specific events in accordance with the CLM is *merely* an empirically well-confirmed generalization (see, for example, Ray 1995: 138; see also Waldner 2007: 152–3; Jackson 2017: 690). This is not so. According to Hempel, what is required is a generalization 'according to' a scientific law (1942: 36; see also 1966: 50); he explicitly distinguished 'general laws' from what he termed 'accidental generalizations' (1966: 55; see also 1965: 377, 417–18), a point which often goes unacknowledged by critics in IR (see, for example, Patomäki and Wight 2000: 228; George and Bennett 2005: 132–3; Bennett 2013: 465; Jackson 2017: 693–4).

[42] Wendt argues that Hempel did not intend the CLM as a 'prescription for how to do science' and hence criticizes social scientists who have treated it as a model 'of what scientific explanations should look like' (1999: 80). We consider this somewhat misleading. Hempel recognized that many explanations are not presented as deductive arguments (see 2001: 281–2), but he *did* insist that *any* scientifically acceptable explanation in *any* scientific field must, ultimately, fit the model.

(stated in the conclusion) after a *C*-type event (stated in Premise 1) conforms to the pattern associated with one or more general laws (stated in Premise 2). To show that the occurrence of an event conforms to a known pattern is not, however, to explain it (see Bunge 1997: 412).

Consider the following experiment, which we briefly referenced in the Introduction to this book. A bowling ball and a feather are dropped, together, in a vacuum chamber; they accelerate downwards together at an identical rate. This sequence of events fits a known pattern—in a vacuum, objects of different shapes and weights accelerate downwards at an identical rate—for experiments of this kind have been performed before. Moreover, this pattern accords with general laws concerning the effects of gravity in the absence of atmospheric resistance. However, pointing out that the observed sequence of events conforms to a known pattern does not explain *why* the two objects accelerated downwards at an identical rate: it merely provides a further data point in support of the known pattern. In order to explain why this event occurred, it will be necessary to provide information about its causes, viz. the two objects were dropped in a vacuum and were hence subject to the influence of gravity, but not of atmospheric resistance.

To be fair to Hempel, he does not simply claim that an event is explained by showing that it fits a known pattern. According to him, what the CLM reveals is that the occurrence of an event of the kind being explained was precisely of the sort that was 'to be expected in the circumstances' (1966: 48)—that is, given the occurrence of an appropriate antecedent event and given the general laws linking events of these two types (Hempel and Oppenheim 1948: 136). He suggests, in other words, that showing that an event of interest 'was to be expected', and hence that its occurrence 'was not "a matter of chance"', *amounts to* explaining it (1942: 39; see also Webb 1995: 132; Dessler 2003: 386–7). There is, however, a clear flaw in this suggestion. Once it is appreciated that a given event is of a kind that was to be expected in the circumstances, its occurrence will no longer be *surprising*, but it may still be *puzzling*; the fact that such an event was to be expected still does not reveal *why* it occurred (see Wendt 1999: 81).[43]

These brief arguments seriously undermine the CLM and the associated idea that knowledge of regularities can be drawn upon to generate a causal explanation of the occurrence of a specific event of interest. Such

[43] It is sometimes claimed that the CLM explains the occurrence of a given event insofar as its deductive structure reveals why this event *had to happen* (see, for example, Taylor 1970: 7–9, 11–12, 75; Salmon 1993: 78–84; Lewis 1993: 195). There is a serious flaw in this argument, too. What the deductive structure of the CLM provides is not a *causal* explanation of *why an event of interest had to happen*, but rather a *logical* explanation of *why a statement that it occurred must be true* (see Suganami 1996: 124).

knowledge does not even provide a *sufficient* basis on which to construct such an explanation, let alone a *necessary* basis. To return to where we started this discussion, Dessler (2003: 378–8) suggests that there are two possible approaches to explaining specific events: a 'generalizing strategy', in which the event to be explained is treated 'as an instance of a certain *type* of event, which is then shown to accompany or follow regularly from conditions of a specified kind', and a 'particularizing' strategy, in which information is provided about 'the sequence of happenings leading up to' the event of interest (2003: 387–8). Yet, as we have shown, identifying an event of interest as an event of a certain type, which occurs regularly under certain conditions, does not explain its occurrence. We recognize the importance, for a range of scientific purposes, of identifying when particular kinds of events regularly co-occur, but in our view *all* causal explanations are, in Dessler's terms, 'particularizing'. The continuing influence of the CLM on how people think about causal explanation in IR reflects the insidious power of the regularity view of causation, but this, as we have shown, is a red herring.

'Causal' and 'constitutive' explanations

The two red herrings we have considered so far in this section have in common a failure to recognize that causal explanations are explanatory statements which relate to specific events, located in space and time: they are not 'theories' and they are not statements of 'regularities'. The final red herring we will discuss is the idea that some explanations which might appear, on an account such as the one we have given, to be 'causal' are in fact 'constitutive'. The basic mistake here is a failure to appreciate that all explanations of how specific events were brought about are, by their nature, *causal* explanations.

In order to unpack this issue, it is helpful to start by noting that many statements which are quite properly described as 'explanations' are concerned with something other than how specific events were brought about. When we (attempt to) explain something to someone, we (attempt to) improve their understanding of it (Achinstein 1977: 1–2), but the range of things we can improve someone's understanding of is very broad indeed, and hence the range of things that can potentially be 'explained' is also very broad (Suganami 2011: 721–2; see also Ruben 1993a: 4; Jackson 2017: 698). There are, in other words, many kinds of explanations which are not causal explanations.[44] This is illustrated in Table 6.2.

[44] Note, however, that the term 'explanation' is often used in IR as a shorthand specifically for *causal* explanation (see, for example, Hollis and Smith 1991a; Van Evera 1997; Chernoff 2014).

Table 6.2 Some explanations which are not causal explanations[a]

Context	Explanation
An explanation to a tourist, of how to get to the museum:	'Turn left at the crossroads, and it's on your left.'
An explanation to a child, of what the green thing on their plate is:	'It's broccoli. Try it!'
An explanation, to an Officer Cadet, of what International Humanitarian Law requires:	'Civilians may not be directly attacked.'
An explanation, by an academic, of the meaning of the verb 'to explain':	'The standard dictionary meaning … is "to make plain" or "to render more intelligible"'. (Suganami 2008: 329–30)

[a] The very possibility of 'noncausal explanation' is sometimes questioned (see, for example, King, Keohane, and Verba 1994: 75). However, it is important to consider what is really at stake in such debates. For example, Jon Elster insists that 'all explanation is causal. To explain a phenomenon … is to cite an earlier phenomenon … that caused it' (2007: 7). He is not really rejecting the possibility that we can explain things other than how specific events were brought about. Rather, he is arguing that all explanations of *how specific events were brought about* are causal explanations. His point is that the category 'causal explanation' does not 'exclude the possibility of intentional explanation of behaviour. Intentions can serve as causes' (2007: 7). We agree with him on this point, to which we will return later.

We will not seek to categorize such explanations any further, but we accept that were a typology to be produced, it would include what might sensibly be called 'constitutive' explanations—that is, explanations of what things are constituted of. However, it seems to us that while explanations of what things are constituted of may sometimes be very valuable, they are categorically distinct from 'causal explanations'—explanations of how specific events were brought about. Moreover, we see no great difficulty in distinguishing between these two kinds of explanations. The notion that some explanations of how specific events were brought about are in fact 'constitutive' explanations is, in our view, mistaken. Let us elaborate.

In IR, the most prominent discussion of the relationship between these two kinds of explanation has been provided by Alexander Wendt (1998; see also 1999: 77–88). He draws a contrast between two kinds of questions: causal and constitutive. On his account, 'causal' questions are about 'why?' and 'how?', as in 'why did Gorbachev move to end the Cold War?' and 'how did the Germans conquer France in 1940?'; he argues that answers to such questions seek to 'explain *changes* in the state of some variable or system'. By contrast, constitutive questions are about 'how-possible?' and 'what?', as in 'How was it possible for Stalin, a single individual, to exercise so much power over the Soviet people?' and 'What kind of political system is the European Union?'; Wendt argues that answers to such questions seek 'to show how the properties

of things are constituted' (1998: 104–5). He emphasizes that questions of both kinds are important and are properly part of science, including social science; we agree with him on these points. However, we believe that the dividing line he draws between 'causal' and 'constitutive' explanation is misplaced.

One superficial issue is that how research questions are phrased—that is, as 'why …?/how …?' questions or as 'how-possible …?/what …?' questions—will not always be a reliable guide to how these questions should be answered, for it will often be possible to rephrase them without significantly changing their meaning. Let us set this aside, however.

A deeper issue is that, in defining 'constitutive' inquiry, Wendt 'equivocates between two different meanings of the word "constitute"' (Jackson 2011: 106; see also Kurki 2008: 180). On the one hand, he describes 'constitution' as being 'conceptual or logical'. He argues, for example, that 'the factors constituting the social kind "Cold War" … define what a Cold War *is* … when they come into being, a Cold War comes into being with them, by definition and at the same time' (1998: 106). In particular, he emphasizes the role of shared ideas in constituting the social kind 'Cold Wars' (see 1998: 109); he seems to be thinking that part of what makes the relationship between two great powers a 'Cold War' is that the protagonists share a distinctive set of ideas about the nature of their relationship and how they stand in relation to one another, this set of ideas being different from the set of ideas associated with, say, a hot war, or peaceful coexistence. On the other hand, Wendt also describes constitutive relations as contributing to the *bringing about* of specific outcomes. He argues, for example, that shared ideas 'generated or produced the Cold War' between the US and the USSR in the twentieth century (1998: 109).

There are two problems with this. First, what Wendt describes as 'constitutive' inquiry appears to include two quite different tasks: explaining what certain things are constituted of, and explaining how certain outcomes were brought about. Second, contrary to what Wendt suggests, explanations of how specific events, such as 'the Cold War', were brought about are *causal* explanations and not *constitutive* explanations.

These problems arise because Wendt defines causal inquiry too narrowly, leading him, in turn, to define constitutive inquiry too broadly. On his account, when one advances a causal claim, 'X causes Y', one assumes 'three things: 1) that X and Y exist independent of each other, 2) that X precedes Y in time, and 3) that but for X, Y would not have occurred' (1998: 105; see also 1999: 79).[45] By contrast, he argues, constitutive inquiry does not

[45] It is unclear what Wendt's (non-italicized) capital letters should be taken to refer to. 'X causes Y' is an abstract statement, so it is natural to suppose that X and Y refer, generically, to events of particular types.

involve the first or second of these assumptions (1998: 105; 1999: 79). To frame this in terms of his discussion of the Cold War, what he is pointing out is the following: if the presence of certain shared ideas constitutes a particular instance of great power rivalry as a Cold War, then these ideas, and the Cold War they constitute, are not logically and conceptually independent of each other; moreover, these ideas, when they exist, do not precede the Cold War in time. This is encapsulated in his claim that when these ideas 'come into being, a Cold War comes into being with them, by definition and at the same time' (1998: 106). So far, so good: we agree with Wendt that the claim that the presence of certain shared ideas constitutes an instance of great power rivalry as 'a Cold War' is not a causal claim, but rather a constitutive one. However, Wendt appears to believe that *whenever* there is a logical or conceptual connection between one event and another then these events are not 'independent', and hence there cannot be a causal relationship between them; the relationship must, instead, be constitutive. This is a mistake.

In effect, Wendt interprets assumption (1) as implying that in a causal claim 'there should not be any logical or conceptual connection whatsoever between the cause-event and the effect-event' (Suganami 2002a: 32). This is too demanding: causes and effects often *are* logically or conceptually related to one another. Suppose someone offers the following explanation: 'reading the text in the footnote through a magnifying glass made the words appear larger'. Here, there is a clear conceptual connection between the cause (looking at the text through a magnifying glass) and the effect (the words appearing larger), for to 'magnify' something is to make it appear larger! However, this conceptual connection between the cause and the effect does not make this explanation a constitutive one. This is important because Wendt's overly demanding supposition that there needs to be a logical or conceptual independence between a cause and its effect unduly restricts what can count as a causal explanation; this, in turn, leads him to offer an over-expansive definition of constitutive explanation, accommodating within this category explanations which are really causal explanations, but in which there is a conceptual or logical link between the cause-event and the effect-event.

As this objection is quite a technical one, it is worth clarifying where we agree with Wendt and where we disagree. First, we agree with him that, in addition to causal relations, the social world features relations which it makes sense to characterize as 'constitutive', and that it is a proper part of social science to investigate both. Second, Wendt argues that many social kinds are

However, the statement 'but for X, Y would not have occurred' appears to refer to two specific events, X and Y, located in space and time.

constituted, in part, by shared ideas: for example, '"treaty violations" are constituted by a discourse that defines promises, "war" by a discourse that legitimates state violence, "terrorism" by a discourse that delegitimates non-state violence' (1999: 84). We agree. Third, Wendt argues that shared ideas have often played an important role in bringing about the outcomes social scientists seek to explain and that neglecting this can have important consequences. For example, he criticizes social scientists for failing adequately 'to problematize the role of … ideas in generating' the Cold War; he suggests, moreover, that this may have contributed to its 'naturalization' and hence its perpetuation. As he points out, it took 'the "New Political Thinking" of the Gorbachev regime' to make an end to the Cold War seem possible. To the extent that social scientists failed to recognize the role of shared ideas in generating and sustaining the Cold War, he argues, they also failed to appreciate how change might be possible; rather than 'helping to empower policymakers to end' the Cold War, he argues, many social scientists were therefore simply helping policymakers 'to manage it' (1998: 109). Once again, we agree.

Where we believe Wendt goes wrong is in his belief that an explanation which points to the role of shared ideas in generating (or sustaining, or ending) the Cold War must be constitutive, rather than causal. He appears to be thinking that such an explanation must be constitutive because the shared ideas which helped to bring about the Cold War are not conceptually independent from their effect: this is because Cold Wars are defined, in part, by the presence of just these kinds of shared ideas. As we have indicated, however, this is a mistake. Causes are often *not* conceptually independent of their effects. Hence, while we are sympathetic to Wendt's emphasis on the role that shared ideas play in international politics, we consider it important to distinguish more clearly than he does between two quite different kinds of claims: claims that shared ideas help *constitute* something as a certain kind of thing and claims that shared ideas *caused*, or helped to bring about, particular outcomes. Incorporating the role of shared ideas does not automatically transform a causal explanation into a constitutive explanation.[46] Ideas, as well

[46] As we read his work, Wendt focused on the role of shared ideas in the Cold War, and in international politics more broadly, because of what we would describe as a *pragmatic* concern that their role had been neglected in existing accounts, with potentially serious consequences for our collective understanding of the dynamics which generate, sustain, and end such conflicts. However, instead of offering a *causal* explanation of the occurrence of the Cold War, in which the protagonists' shared ideas were highlighted as a much neglected and hence especially noteworthy cause, Wendt claimed to be offering a distinctively 'constitutive' explanation, apparently seeing the neglect of shared ideas in existing explanations as an instance of a broader neglect in IR of the vital role of 'constitutive' inquiry in science. This was a mistaken and misleading move.

as norms, meanings, and other distinctively *social* features of international politics, can feature in *causal* explanations (see Kurki 2008).[47]

In our view, all explanations which are concerned with how specific events, located in space and time, were brought about are *causal* explanations. Once this is recognized, it should also become apparent that the significance which is often attached to the distinction between causal and constitutive explanations is overstated. We do not wish to diminish the value or scientific credentials of constitutive explanations; our point is simply that, most of the time, there will be no particular difficulty in distinguishing between explanations of what kind of thing something is and explanations of how something was brought about.[48] In short, it will usually be possible to get on with the business of developing both causal and constitutive explanations without worrying unduly about the relationship between them.

Conclusion

In this chapter, we have argued that a causal explanation explains how specific events (or sets of events), located in space and time, were brought about, and that it does this by identifying one or more of the causes of such events. We suspect that, stated in this way, the position we have laid out will seem uncontroversial to many readers. As we have shown, however, causal explanation is a rather neglected topic in IR, and when it is discussed a series of misunderstandings rear their heads.

Some of these concern what a causal explanation is and how it relates to other kinds of knowledge claims. We considered three such misunderstandings in Section III of this chapter: we showed that it is crucial to maintain the distinction between causal explanations, which relate to specific events, and causal theories, which abstract from them; we showed that causal explanation does not require knowledge of regularities; and we argued that *all* explanations of how specific events were brought about are causal explanations.

[47] In this respect, the categorical distinction that is often drawn between interpretive understanding and causal explanation (see, for example, Hollis and Smith 1991a) is quite misleading. It is possible that Wendt came to define causal explanations which refer to the role of shared ideas as 'constitutive' in part because he was influenced by such distinctions, which were a prominent part of the intellectual milieu in which he operated at the time (see Wendt 1998).

[48] It is sometimes suggested that what we are terming 'constitutive' explanations (that is, explanations of what kind of thing something is) are still causal in the sense that what an object is constituted of shapes its causal powers (see, for example, Jackson 2011: 107). However, we consider this misleading. It may be that constitutive inquiry helps us to understand what kind of thing something is and that this, in turn, helps us to appreciate what its causal effects may be (if the conditions are right), but it remains important to distinguish the (constitutive) question of what something is from the (causal) question of what its effects may be.

Other common misunderstandings relate to the pragmatic component of causal explanation, which is often neglected, but which it is crucial to appreciate when thinking about what makes causal explanations satisfactory. We considered several such misunderstandings in Section II of this chapter: we showed that the notion of a 'complete' causal explanation makes no sense, that causal explanations can take a wide variety of forms, and that the oft-heard call for causal explanations to identify causal mechanisms is best understood as a call for more detailed information than has previously been provided about how a particular causal process of interest unfolded.

Our chief aim has been to correct these misunderstandings and to highlight the importance of the pragmatic component of causal explanation. However, it is also important to appreciate how these misunderstandings connect to the misunderstandings of the contents and implications of causal statements and of the logic of causal inquiry which we considered in Chapters 4 and 5.

The common thread is the widespread, but misleading, supposition, associated with the culture of generalization, that the kind of knowledge generated through processes of 'causal inference' must, in some sense, be 'general'. As we explained in Chapter 5, this overlooks the foundational importance, in causal inquiry, of efforts to provide empirical support for explanatory statements relating to specific events located in space and time. Because this is overlooked, the significance of causal explanation as a topic is underestimated; this, in turn, has probably contributed to the neglect of the pragmatic component of causal explanation. In this chapter, we have sought to correct this. In Chapter 7 we will consider another topic in relation to which misleading suppositions associated with the culture of generalization also point in the wrong direction: causal reasoning.

7
Causal Reasoning

This is the second of two chapters which develop key themes arising out of our presentation, in Chapter 5, of the logic of causal inquiry. Having discussed causal explanation in Chapter 6, our focus in this chapter is on causal reasoning. Our aims are twofold. First, we will demonstrate more fully how our account of the logic of causal inquiry connects to existing discussions, especially in the research methods literature, of 'causal inference'. Second, we will explore how, in practice, one might reason about whether particular causal knowledge claims should be accepted. As we showed in Chapter 5, there is an underlying logic to causal inquiry, but it is important to recognize that, in practice, judgements about the acceptability of particular causal knowledge claims always have a degree of uncertainty attached to them. In this chapter, we will consider, in more detail, the kind of reasoning required to support *practical judgements* in the context of causal inquiry, within the broad contours of the logic we have already laid out.

We argued, in Chapter 5, that the only kind of causal statement which can be directly supported by empirical evidence is an explanatory statement, and that the kind of evidence which will provide persuasive support for such a statement is evidence which rules out competing explanatory statements. In the research methods literature, the question of how causal knowledge claims can be generated from empirical evidence is typically framed as a question of 'causal inference' (specifically, the 'internal validity' of causal inferences). Existing discussions offer valuable guidance as to how particular methods for gathering and analysing empirical evidence can be applied to generate causal inferences. As we shall see, however, the research methods literature gives a somewhat misleading impression of the kind of reasoning on which causal inferences rest, in part because of the widespread presumption, linked to the culture of generalization, that the kind of knowledge produced by a causal inference must, in some sense, be 'general'. In Section I, we will illustrate this problem and then seek to develop a more satisfactory account of the kind of reasoning which can support practical judgements about the acceptability of particular explanatory statements.

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0008

We also argued, in Chapter 5, that a well-supported causal theory can be applied to generate causal explanations as well as causally informed predictions and policy prescriptions, though we pointed out that this is less straightforward than is commonly supposed because it requires knowledge of the conditions under which the propensity described by the theory unfolds fully. In the research methods literature, the question of how and when existing causal knowledge can be applied in new situations is typically framed as a question about the 'external validity' of causal inferences. In our view, however, existing discussions provide an incomplete account of what is involved in applying causal theories in new situations, in part because of the presumption, linked to the culture of generalization, that it must somehow be possible to 'generalize' the conclusion of a reliable causal inference. In Section II, we will examine this problem and provide our own account of the kind of reasoning which can support practical judgements about how, and how confidently, existing causal knowledge can be applied in new situations. By the end of the chapter, we aim to have provided the first account of causal reasoning free of the influence of the culture of generalization.

I. 'Causal inference'

'Inference' is arguably *the* central concept in mainstream discussions of causal inquiry in IR. Indeed, King, Keohane, and Verba argue that inference 'is the ultimate goal of all good social science': the goal of making 'descriptive or explanatory *inferences* on the basis of empirical information' is what 'sets social science apart from casual observation' (1994: 34, 6–7; italics in original). According to Brady, Collier, and Seawright (2010a: 153–4), this emphasis on the importance of 'inference' is 'widely shared in contemporary social science': they note that even King, Keohane, and Verba's critics mostly share the 'overarching goal' of seeking 'valid descriptive and causal inferences about important phenomena in the political and social world'.

Despite its obvious importance, the term 'causal inference' is rarely defined in detail. There is very little discussion of what kind of inference a causal inference is and, in particular, of the kind of reasoning it involves—whether this reasoning is inductive, deductive, abductive, or of some other kind. Instead of exploring these underlying issues, discussions typically focus on showing how certain, favoured methods for gathering and analysing empirical evidence can be used to generate causal inferences. This kind of practical guidance is, of course, extremely valuable. However, we believe that it is *also*

important to consider what kind of reasoning will, in practice, be required to support a persuasive causal inference. As we shall see, one benefit of considering this issue is that it generates insight into the features that make a causal inference persuasive, *whichever* method for gathering and analysing empirical evidence has been used.

We argued, in Chapter 5, that the only kind of causal statement for which empirical evidence can provide direct support is an explanatory statement—that is, a concrete causal statement, paradigmatically '*a* caused *b*', relating to two or more specific events which have already occurred. It follows from this that the conclusion of what is termed a 'causal inference'—the generation of a causal knowledge claim through the study of empirical evidence—must also be an explanatory statement.[1] This goes largely unnoticed in existing discussions, which are strongly shaped by the culture of generalization and hence tend to presume, implicitly or explicitly, that causal inferences produce 'general' knowledge. As we shall see, this paints a misleading picture of the kind of reasoning such inferences involve.

We will begin our analysis by examining King, Keohane, and Verba's (1994) influential account of causal inference, showing how their commitment to the culture of generalization leads them to suggest, implausibly, that a causal inference is closely akin to an inductive generalization.

Having rejected the idea that a causal inference might be *inductive*, we will turn to the work of Nancy Cartwright (2007), who argues that at least some methods for generating causal inferences rely on *deductive* reasoning. We will contend that, in practice, judgements about whether or not to accept a particular explanatory statement cannot possibly rest on deductive reasoning, given the uncertainty attached to them. We will, however, find in Cartwright's account some clues which indicate what kind of reasoning *is* required.

We will argue that the reasoning which, in practice, can support a persuasive causal inference is *neither inductive nor deductive, but rather 'abductive'*: it is reasoning to the effect that a given explanatory statement should be accepted if it appears to be the 'best' of a set of plausible competing explanatory statements. Having laid out the basic ingredients of abductive reasoning and shown how they apply to causal inference, we will conclude by considering what makes such an inference more, or less, persuasive.

[1] We pointed out in Chapter 4 that an explanatory statement, '*a* caused *b*', implies a corresponding causal theory, '*A* causes *B*'. However, we will reserve the term 'causal inference' to describe an inference, based on empirical evidence, that an explanatory statement should be accepted. The further inference that *if* '*a* caused *b*', *then* '*A* causes *B*' is a deductive inference: it does not require empirical evidence.

King, Keohane, and Verba on 'causal inference'

As we saw in Chapter 4, King, Keohane, and Verba (1994) are committed proponents of the culture of generalization. This profoundly shapes their account of causal inference.

King, Keohane, and Verba do not explicitly discuss the kind of reasoning which this sort of inference involves. However, they do insist that the purpose of any inference is to produce 'general knowledge' and that the data which is collected in order to generate an inference must be organized so as to allow analysts to 'go beyond the particular to the general' (1994: 46, 48). They also draw a close link between 'descriptive inference' and 'causal inference'. They argue that social science research involves 'describing and explaining', and that both of these tasks 'depend upon rules of scientific inference' (1994: 34). Hence, they envisage themselves as elaborating 'rules of inference' which apply whether the aim is description (involving 'descriptive inference') or explanation (involving 'causal inference'). As we shall see, this generates a confusing picture of causal inference and of the kind of reasoning it involves.

Let us start by briefly considering King, Keohane, and Verba's account of 'descriptive inference', which they distinguish from 'description—the collection of facts'. They argue that '[s]ystematically collecting facts is a very important endeavour without which science would not be possible', but that it 'does not by itself constitute science' (1994: 34). Properly 'scientific description', they argue, 'involves inference: part of the descriptive task is to infer information about unobserved facts from the facts we have observed' (1994: 34; see also 55). Typically, these 'unobserved facts' will be parameter estimates—that is, estimates of parameters such as the mean or variance of a variable of interest in a population (see 1994: 63–73). Hence, King, Keohane, and Verba presume that when 'facts' are 'collected', these relate to a sample drawn from a broader population, and that what researchers are really interested in is this broader population.

It should be clear that, on this account, the kind of reasoning which underpins *descriptive* inference is inductive generalization: descriptive inferences generalize from a sample of observed facts to unobserved facts about a broader population, taking careful account of whether the sample observations 'reflect ... typical phenomena or [are] outliers' (King, Keohane, and Verba 1994: 56). Bearing this in mind, and also bearing in mind the close connection that King, Keohane, and Verba draw between descriptive inference and causal inference, alongside their underlying presumption that the aim of any kind of inference must be to produce 'general knowledge', let us now consider their account of *causal* inference.

Early on in their book, King, Keohane and Verba offer a broad definition of 'causal inference' as 'learning about causal effects from the data observed' (1994: 8).[2] They also suggest that social scientists seek to generate causal inferences in order to 'devise causal explanations of the phenomena they study' (1994: 75). However, rather than considering the range of ways in which this might be done, and the kind of reasoning it might involve, their focus quickly narrows to the question of how a counterfactual 'causal effect' may be defined and estimated (see McKeown 2004: 146). In outline, their account follows Holland's (1986) elaboration of 'Rubin's causal model', or what is also called the 'potential outcomes framework': they first define a counterfactual causal effect (roughly: as the difference between the value on a dependent variable that was actually observed and the value that would have been observed had the value on some independent variable been different) and then consider how, given certain quite demanding assumptions, statistical analysis can be used to derive estimates of such effects from comparative data (see King, Keohane, and Verba 1994: 76–85, 91–5).[3]

The potential outcomes framework provides a clear conceptualization of how it is that certain kinds of experimental and observational studies can generate estimates of unobservable counterfactuals which may be of interest in causal inquiry. We do not wish to raise any substantive objection to this enterprise, though it is worth noting that, in practice, it is closely associated with efforts to derive causal inferences through analysis of what Collier, Brady, and Seawright (2010b: 184–8) term 'data-set observations'; consequently, it fails to illuminate *other* ways in which causal inferences may be derived: for example, through process tracing. Rather than pursuing this point, we will focus on the picture which King, Keohane, and Verba's discussion appears to generate of the kind of reasoning involved in causal inference.

At some points, King, Keohane, and Verba indicate that a causal inference *depends on* a prior descriptive inference. They argue, for example, that 'causal inference is impossible without good descriptive inference' (1994: 75, see also 34). We find this confusing. To be fair to King, Keohane, and Verba, having introduced this idea, they do not pursue it any further:

[2] To see why we say their definition is 'broad', note that *any* postulated causal relationship can be described as involving a cause having an effect and, hence, that *any* approach which draws on empirical evidence to generate knowledge of causal relationships may be said to involve 'learning about causal effects from the data observed' (see Collier, Brady, and Seawright 2010a: 129). This is notable because, as many critics have pointed out, King, Keohane, and Verba end up adopting a very restrictive stance in relation to how causal inferences can be derived (see Brady and Collier 2004, 2010).

[3] The fact that King, Keohane, and Verba rely on a counterfactual definition of causation and attempt to show how statistical techniques can be used to control for potential sources of confounding makes it inaccurate to regard them as 'regularity theorists' in any of the senses we discussed in Chapter 1, Kurki's criticisms notwithstanding (2008: 101–8).

at no point do they seek to show that the data which underpins a causal inference must itself have been generated through a descriptive inference. What they do insist on is that estimating a causal effect is 'directly analogous' to estimating population parameters such as 'means and variances' (1994: 80–1, see also 97). However, this, too, strikes us as confusing. To see the problem, let us briefly consider how a causal inference might be derived from an RCT in line with King, Keohane, and Verba's preferred potential outcomes framework.

Suppose that the aim is to identify what the effect of a certain treatment would be on an individual unit (see Holland 1986: 946–7). Ideally, we would like to be able to observe the outcome for that unit both with and without the treatment. Yet, at any one time we can observe only one of these potential outcomes, not both.[4] What we *can* do, though, is observe these two potential outcomes on *different* units, applying the treatment to some units, and not others, and comparing its effect.[5] In an RCT, this is done in the fashion we have described previously: the test subjects are randomly assigned to two groups, one of which is given the treatment and one of which functions as a control group; in an ideal experiment, the difference, on average, in the outcome observed in the treatment group, as compared to the control group, constitutes an unbiased estimate of the average causal effect of the treatment on the test subjects. If it is further assumed that this effect is constant across all units, then the experiment will have generated an unbiased estimate of the effect of the treatment on each individual unit, even though the two 'potential outcomes' for each unit (the outcome with and without the treatment) could not both be observed (see Holland 1986: 949).

This account should, we hope, be reasonably familiar and uncontroversial. Yet, two important points flow from it. First, although the 'units', or 'test subjects', in an RCT may be conceptualized as a sample drawn from a broader population, it is clearly *not* necessary to estimate any population parameters in order to generate a causal inference. In other words, contrary to what King, Keohane, and Verba suggest, a causal inference does *not* depend on a prior descriptive inference. Second, and again contrary to what King, Keohane, and Verba suggest, there is a significant *disanalogy* between estimating a population parameter and generating a causal inference. Whereas estimating population parameters from a sample involves generalizing from

[4] King, Keohane, and Verba (1994: 79), following Holland (1986: 947), describe this is as the 'fundamental problem of causal inference' (italics removed from original).

[5] As Holland (1986: 947) points out, in a laboratory experiment we may be able to observe both outcomes, for a single unit, *at different times*. However, this will usually not be possible in the kinds of observational studies with which King, Keohane, and Verba are principally concerned.

a sample to a population, causal inference does not.[6] What an RCT generates is an estimate of the average effect of the treatment 'in the circumstances of the experiment' (Cartwright 2007: 39)—that is, *on the trial participants, in the experimental context*; this is why the 'external validity' of such experiments is sometimes questioned (see Lucas 2003; McDermott 2011: 35; Gerring 2012: 272; Shadish, Cook, and Campbell 2002: 18). We will discuss this in more detail later.

It is crucial to recognize that causal inferences and descriptive inferences draw on 'facts' in very different ways. A descriptive inference starts with a set of descriptive facts about a study sample and generates further facts of the *same kind*—descriptive facts—but at a *broader level of generality*: at the population level rather than at the sample level. By contrast, a causal inference starts with a set of descriptive facts about a study sample and generates facts of a *different kind*—causal facts—but at the *same level of generality*: still at the level of the study sample. This is summarized in Table 7.1.

Hence, the close link which King, Keohane, and Verba draw between 'causal inference' and 'descriptive inference' gives a misleading impression of the kind of reasoning required to generate a persuasive causal inference. Whereas descriptive inference involves inductive generalization, causal inference clearly does not.

Although some features of King, Keohane, and Verba's discussion are idiosyncratic, their thinking rests on an underlying presumption which is widely shared: that what a causal inference must produce is 'general'

Table 7.1 How descriptive and causal inference draw on descriptive facts

	Type of knowledge produced	Generality of knowledge produced
Descriptive inference	Same kind (descriptive)	Broader than the facts observed
Causal inference	Different kind (causal)	Same as the facts observed

[6] King, Keohane, and Verba muddy the waters by insisting that causal inference, like descriptive inference, requires researchers to 'partition the world into systematic and nonsystematic components' (1994: 79, see also 34). This strikes us as confusing. An RCT is powerful precisely because it can generate an unbiased estimate of the effect of a treatment even when very little is known about potential confounders, whether systematic or nonsystematic. King, Keohane, and Verba's acknowledgement (1994: 79) that, on this issue, they depart from Holland (1986), on whom they otherwise draw closely, is quite revealing. Whereas they emphasize the commonalities between causal and descriptive inference, one of Holland's key aims was to 'show why the statistical models used to draw causal inferences are distinctly different from those used to draw associational inferences' (1986: 945).

knowledge. It is this presumption, we believe, that encourages them to draw a misleading connection between descriptive inference, which involves inductive generalization, and causal inference, which does not. We believe that this presumption is shared, at least implicitly, across most of the research methods literature. Yet, if the aim is to understand what kind of reasoning will support a persuasive causal inference, this presumption points in the wrong direction. Let us therefore consider some other possibilities.

Cartwright: Causal reasoning as 'deductive'?

Nancy Cartwright is one of the few scholars we are aware of, in either IR or philosophy, to explicitly consider the question we are posing, viz. 'what kind of reasoning is required to support a persuasive causal inference?'. She argues that '[m]ethods for warranting causal claims fall into two broad categories … those that clinch the conclusion … and those that merely vouch for [it]'. On her account, the kind of reasoning relied on by methods in the first category is 'deductive' (2007: 25), which is why such methods 'clinch' their conclusions: if the premises of a valid deductive argument are true, then the conclusion *must* be true. By contrast, the reasoning involved when methods in the second category are applied is *not* deductive, which is why such methods merely vouch for their conclusions, rather than clinching them. As will become apparent, we do not find Cartwright's analysis wholly satisfactory, but we do consider its limitations to be quite instructive.

Cartwright's contention that some methods for warranting causal claims involve deductive reasoning is best elaborated through an example. Consider, once again, the design of an RCT. As Cartwright explains (2007: 31), the aim, in randomly assigning test subjects to treatment and control groups, is to ensure that factors other than the treatment 'are distributed equally' across the two groups. The aim of other aspects of the experimental setup, such as double blinding, and controlling the environment in which the trial takes place, is to control for other potentially confounding factors 'that randomization misses'. This suggests that it might, in principle, be possible to derive a deductively valid causal inference from an RCT: if some difference in the average value of the outcome variable in the treatment and control groups was observed, if it were assumed that this difference was produced by some causal factor(s) (see Cartwright 2007: 30), and if the design of the RCT had ensured that all possible sources of confounding were prevented (see Cartwright 2007: 31), thereby ensuring that the treatment was the only causal factor which differed across the two groups, then it would follow, deductively, that the observed difference in outcome across the treatment

and control groups must have been caused by the treatment. This reasoning is summarized in Table 7.2.

Although we have illustrated this pattern of reasoning by reference to an RCT, it can be expressed in a generic form so as to encompass *any* method for generating and analysing evidence in support of a causal inference, whether this method involves gathering evidence relating to a correlation or probabilistic dependency or evidence relating to the occurrence of a specific event. These generic forms of the argument are summarized in Tables 7.3 and 7.4:

Table 7.2 Deductive reasoning supporting the derivation of a causal inference from an RCT (Randomized Controlled Trial)

Premise 1	There is a difference in the average value of the outcome variable in the treatment and control groups
Premise 2	This difference was produced by *some* causal factor(s)
Premise 3	All possible sources of confounding have been prevented *from which it follows logically that*
Conclusion	The observed difference in the average value of the outcome variable in the treatment and control groups was produced by the treatment

Table 7.3 Generic deductive argument in support of a causal inference relating to a correlation or probabilistic dependency

Premise 1	A correlation (or probabilistic dependency) has been identified between the occurrence of *A*-type events and of *B*-type events
Premise 2	The observed *B*-type events were produced by *some* causal factor(s)
Premise 3	The influence of all possible causal factors other than the occurrence of the *A*-type events has been adequately controlled for *from which it follows logically that*
Conclusion	The observed *A*-type events contributed to the bringing about of the observed *B*-type events

Table 7.4 Generic deductive argument in support of a causal inference relating to the occurrence of a specific event

Premise 1	An event, *a*, was followed by another event, *b*
Premise 2	Event *b* was produced by *some* causal factor(s)
Premise 3	All possible accounts of how *b* was produced in which *a* does not appear as a cause can be ruled out *from which it follows logically that*
Conclusion	Event *a* caused event *b*

This may appear to suggest a clear answer to our question: the kind of reasoning required to support a persuasive causal inference is deductive. As Cartwright acknowledges, however, there is a serious problem with the idea that, in practice, a causal inference might be derived in this way. This relates to the nature of deductive arguments. As Cartwright explains, the power of such arguments is that they 'clinch' their conclusions: if the premises are true, then the conclusion must also be true. The flip side of this, however, is that in a deductive argument 'the conclusion can only be as certain as the premises': deductions 'can take us from truths to truths but once there is one false premise, they cannot do anything at all' (Cartwright 2007: 31, 34). Consider, for example, the following argument:

Premise 1	One of the authors of this book is called Adam
Premise 2	Everyone called Adam comes from Mars
Conclusion	One of the authors of this book comes from Mars

This argument is deductively valid: *if* the premises are true, then so is the conclusion. However, it is not sound: Premise 2 is not true, and hence the argument provides *no reason whatsoever* to accept the conclusion.

With this in mind, consider once again the deductive arguments laid out above. As Cartwright points out, the problem with such arguments is that the premises are 'extremely restrictive' (2007: 25). Even if we accept Premise 2, Premise 3 requires that 'every possible source of variation of every kind must be controlled' (2007: 29). Only if it is controlled, and if it is known to be controlled, can a valid conclusion be drawn. But, as Cartwright points out, '[o]ften we do not know how to' establish that every possibly confounding factor has been controlled for (or that every competing account of how an outcome of interest was brought about can be ruled out) and, worse, 'frequently we know' that this has not been achieved (2007: 33).

To illustrate this point, Cartwright asks how we might 'know, for an RCT … that *all* other [causal factors] … have been … controlled for' (2007: 33). It might be supposed that random assignment is sufficient to guarantee this, but that is not so. What random assignment guarantees is merely that an estimate of the average causal effect of the treatment in an RCT is *unbiased*, and not that all possible confounders have in fact been controlled for (such that an estimate of the average causal effect of the treatment fully accounts for their potential influence) (Rubin 1974: 693–4). In order to establish whether *all possible* confounders were *in fact* evenly distributed across the two groups, it would be necessary to check these one by one (Rubin 1974: 695–7). As Rubin

notes, such 'manual' checks are well advised (see also Cartwright 2007: 31), yet they can never be conclusive, for how could it be shown that the distribution of *all possible* confounders has been checked? How could all the *possible* confounders even be identified?

To summarize: the problem with the idea that, in practice, causal inferences employ deductive reasoning is that the conclusion of a deductive inference can be known to be sound only if the premises are known to be true, and in the case of a causal inference it is not possible to be certain that the premises required to support a deductive argument are true.

Cartwright explicitly acknowledges this problem and therefore suggests that it is necessary 'to take seriously' what she describes as 'non-deductive methods' (2007: 34)—that is, methods which 'vouch' for their conclusions but do not 'clinch' them. In practice, she observes, scientific conclusions are often accepted based on a considered evaluation of the 'breadth, variety, precision and novelty of evidence' (2007: 37), rather than on the basis of a deductive argument. We agree, but the question this raises is what kind of reasoning can be used to support a judgement that the evidence is *sufficiently* broad, varied, precise, novel, or whatever, that an inference in support of a particular causal statement should be accepted. Cartwright does not provide an answer. In fact, she doubts that 'there are any good 'logics' of non-deductive confirmation, especially ones that make sense for the great variety of methods' that are employed for causal inference (2007: 25)!

This creates a double bind. On the one hand, Cartwright shows how deductive reasoning could, in principle, be used to derive a causal inference, but acknowledges that, in practice, it will never be possible to establish that the required premises hold. This suggests that, in practice, even causal inferences derived through methods such as RCTs are not deductively derived. On the other hand, Cartwright dismisses the attempt to spell out systematically what form non-deductive reasoning in support of a causal inference might take. In effect, her analysis reveals why it is implausible to suppose that, in practice, causal inferences involve deductive reasoning, but fails to identify a clear alternative.

We accept Cartwright's contention that there are no good logics of non-deductive *confirmation*, if what this means is that there are no non-deductive forms of reasoning which can establish a conclusion *with certainty*. In the context of causal inquiry, however, certainty is probably an illusory goal. If so, then the important question is whether there is some non-deductive form of reasoning which can make a *persuasive case* for accepting a particular causal statement, given the available evidence, despite the residual uncertainty. In other words, the question is whether it is possible to provide what Stephen

Toulmin terms a 'justificatory argument' (2003: 12) which provides persuasive grounds for accepting a particular causal statement, given the available evidence. We believe that this is possible.

Toulmin argues that whatever their substantive content, 'justificatory arguments' share an underlying structure. In simplified form (see Toulmin 2003: 89–100), this goes as follows:

- Evidence is offered in support of some conclusion;
- There is a clear warrant for believing that evidence of this kind justifies acceptance of the conclusion;
- The conclusion is accepted subject to a qualification concerning the degree of uncertainty that remains.

The question, therefore, is what kind of reasoning would provide a clear warrant for accepting a particular explanatory statement, given the available evidence, even though the kind of certainty offered by a deductive argument cannot be secured. We believe that the kind of reasoning that is required is 'abductive'.

'Abductive' reasoning in support of causal inferences

Probably the most familiar distinction among kinds of inference is that between deductive and inductive inferences. As we have already pointed out, in a deductive inference the premises contain all the information required to generate the conclusion. Such inferences are hence sometimes termed 'necessary' inferences (Douven 2021: §1.1), for if the premises are true, then the conclusion must also be true. By contrast, inductive inferences are 'ampliative' or 'non-necessary': the conclusion goes beyond, or amplifies, what is contained in the premises, and hence is not *necessarily* true even if the premises are true (Douven 2021: §1.1). Often, ampliative inferences are further subdivided: the term 'inductive' is reserved for inferences 'based purely on statistical data'—for example, when a property of a population is inferred from information about a sample drawn from that population, as it is in what King, Keohane and Verba term 'descriptive inferences'; meanwhile, the term 'abductive' is used for inferences which involve 'explanatory considerations' (Douven 2021: §1.1). This generates a tripartite schema: inferences can be deductive, inductive, or abductive.[7]

[7] More loosely, one might also 'extrapolate', say, that the future will be 'like' the past, or that one case will be 'like' another. In terms of our tripartite schema, this is probably best described as an 'inductive'

We have already categorically rejected the idea that causal inferences might be 'inductive' in the sense that they involve generalization from samples to populations. We have also cast serious doubt on the idea that, in practice, the reasoning which supports causal inferences can be deductive. It is therefore worth considering the idea that, in practice, causal inferences may rely on abductive reasoning.

The basic idea behind 'abductive inference' is that we can have a good reason to accept a proposition based on how well it explains certain observed facts. Peter Lipton provides an everyday example: 'Faced with tracks in the snow of a certain peculiar shape, I infer that a person on snowshoes has recently passed this way' (2004: 1; see also Lipton 2004: 56; Van Fraassen 1980: 19–20; Lipton 2007a: 421; Douven 2021: §1). Lipton argues that we have good reason to believe that 'a person on snowshoes has recently passed this way' if this is 'the best of the competing explanations we can generate' for the peculiarly shaped tracks in the snow (Lipton 2004: 56). 'Abductive inference' is hence also sometimes called 'inference to the best explanation', the two terms being used largely interchangeably in the contemporary philosophical literature (Douven 2021; see also Bird 2005: 5).[8]

In Lipton's example, what is supported by an inference of this kind is a factual proposition: 'a person on snowshoes has recently passed this way'. As Alexander Bird (2010: 347) points out, the idea that a factual proposition of this kind 'explains' an outcome of interest—in this case, the peculiarly shaped tracks in the snow—is 'shorthand' for the idea that the event described in this factual proposition *caused* the outcome of interest. What Lipton is arguing is hence that, in the absence of direct evidence to support it, we will have a good reason to accept the proposition 'a person on snowshoes has recently passed this way' only if we have a good reason to believe that a person on snowshoes recently passing this way is what *caused* the peculiarly shaped tracks in the snow.

It is worth spelling this point out more generically. Suppose that an event, *b*, is known to have occurred and that we wonder whether another event,

inference, though one which is rather less secure than a parameter estimate. We will discuss 'extrapolation' further in Section II of this chapter.

[8] 'Inference to the best explanation' is, in fact, Lipton's (2004) preferred term. One reason why it is sometimes preferred to the term 'abductive inference' is to avoid confusion with earlier uses of the term 'abduction' (see, for example, Harman 1965: 88–9). As Douven explains, in 'the historically first sense' of the term, which originated with C. S. Peirce, the term 'abduction' referred to the use of 'explanatory reasoning in *generating* hypotheses' for subsequent testing; in the modern philosophical literature, however, it refers to the use of 'explanatory reasoning in *justifying* hypotheses' (Douven 2021; italics original). In IR, the term is still sometimes used in the earlier sense: 'abductive' reasoning is described as generating hypotheses for testing (see, for example, Chernoff 2005: 82; Jackson 2011: 83). However, we will follow more recent philosophical practice.

a, has also occurred. Lipton is arguing that we have good reason to accept that *a* occurred if '*a* caused *b*' is the best of the competing explanations for *b*. To spell this out even further: his contention is that if the statement '*a* caused *b*' is the best of the competing explanations for *b*, then this statement should be accepted, and that this implies, in turn, that the factual proposition '*a* occurred' should also be accepted. For our purposes, in this book, what we are interested in is the first part of this contention: that the statement '*a* caused *b*' should be accepted if it is the best of the competing explanations for *b*. In other words, we are interested in what reasoning might enable us persuasively to identify an explanatory statement, paradigmatically '*a* caused *b*', as the best of the competing explanatory statements that can be generated for *b*.

In order to understand what such reasoning might involve, it is crucial to consider what it means for a given explanatory statement to be the 'best of the competing explanations'. On Lipton's account, 'competing' explanations are 'incompatible' with one another: they 'cannot all be true' (2004: 57). To assert that an explanatory statement, '*a* caused *b*', is the 'best of the competing explanations' is hence to assert that all incompatible explanatory statements—all statements which assert or otherwise imply that *a* was *not* among the causes of *b*—should be rejected.

This is consistent with how we defined 'competing' explanatory statements in Chapter 5. However, it is important to note a potential source of confusion. We showed, in Chapter 6, that it is one question whether the information provided in an explanatory statement is *reliable* and another question whether this information is *relevant* for some specific explanatory purpose. When Lipton speaks of the 'best of the competing explanations', he is talking about the explanation that is the most reliable. To say that '*a* caused *b*' is the 'best of the competing explanations' is to say that, based on the evidence, there is good reason to believe that *a* did indeed contribute to the bringing about of *b* and that those explanations which imply that *a* was *not* among the cause of *b* are unreliable. There is no implication that this explanation is 'the best' in a pragmatic sense: Lipton *does not* mean to say that whenever an explanation of *b* is sought, it will be sufficient to provide the information '*a* caused *b*'. Whether this information satisfies the curiosity of a given audience will depend on what they already know and what they are puzzled by.

In order to retain the focus on which of a set of competing explanatory statements is the most reliable, and to avoid conflating this with the very different question of whether the information this statement provides is the most relevant for a specific explanatory purpose, we will, from hereon,

avoid the term 'best explanation'.[9] We will instead talk of whether a given *explanatory statement* is the *best of the competing explanatory statements*. In related fashion, we will also describe reasoning in support of the contention that a given explanatory statement is the best of the competing explanatory statements as 'abductive reasoning', thereby avoiding the potential confusion caused by the language of 'inference to the best explanation'.[10] By 'abductive reasoning', therefore, what we mean is reasoning which supports an explanatory statement by indicating why competing explanatory statements can be rejected.

Having clarified this, let us consider how it is possible to reason persuasively in support of the contention that a given explanatory statement is the best of the competing explanatory statements.

Simplistically, one might suppose that the statement '*a* caused *b*' is 'best' if it is 'better supported' by the evidence than competing explanatory statements. However, this is not sufficiently precise. For convenience, assume that just two statements are being evaluated, '*a* caused *b*' and '*c* caused *b*', and that '*c* caused *b*' implies '*a* did *not* cause *b*', making the two statements incompatible. The question, now, is whether the evidence indicates which of these two statements should be rejected. In practice, any judgement reached on this issue cannot be entirely certain (we will return to this point shortly), but there are, nevertheless, just three basic possibilities: first, the evidence is judged to be consistent with both statements, in which case neither is 'better supported' than the other; second, the evidence is judged to be consistent with one statement and inconsistent with the other, in which case the point is not so much that the former statement is 'better supported', but rather that the latter statement is not supported at all and should hence be categorically rejected; third, the evidence is judged to be inconsistent with both statements, in which case neither is supported and they should both be categorically rejected.

This may appear to suggest that abductive reasoning involves the *elimination* of all but one of a set of competing explanatory statements. This is what Alexander Bird has in mind when he argues that, in some cases, abductive

[9] Lipton appears to acknowledge the confusion that this term might generate when he notes that the 'suggestion of uniqueness that the word "best" carries' could potentially mislead. He emphasizes that the phrase 'best explanation' should not be taken to imply that there is a single 'best' explanation of any specific phenomenon (that is, one that is pragmatically the 'best' when offered in response to any explanation request), but only that there is always a 'best' of the *competing* explanations (2004: 62–3).

[10] We also have a further reason for wishing to avoid this language. In IR, 'inference to the best explanation' is most frequently discussed in a very specific context, viz. the debate about whether the 'best' explanation for the success of science is that the unobservable theoretical entities posited in our best scientific theories are real and, if it is, whether this, in turn, constitutes a compelling consideration in favour of scientific realism. We touched on this issue briefly in Chapter 3. It is a very different issue from the one we are interested in here.

inference involves 'eliminative induction' or what he also terms 'inference to the only explanation' (see Bird 2005, 2007, 2010). He imagines a scenario in which it is known that 'one of ten hypotheses is true' and evidence is obtained 'that is inconsistent with nine of the hypotheses'. He argues that in this scenario it can be established, deductively, that the one remaining hypothesis must be true (Bird 2005: 11; see also Bird 2007: 431). There are, however, two problems with this. First, how were the ten hypotheses selected for evaluation, and how is it possible to know that one of them is true (see Van Fraassen 1989: 143; Psillos 1996: 36–43; Ladyman, Douven, Horsten, and Van Fraassen 1997: 306–9; Lipton 2004: 152; Lipton 2007b: 452)? It will always be possible to imagine other competing hypotheses, even if these are somewhat contrived. Second, no hypothesis, taken in isolation, can be definitively falsified, for how is it possible to be sure that it is the hypothesis that is false, rather than the putative facts (Lipton 2007b: 452–3)?[11] These two problems suggest that it is unrealistic to hope that all but one of a set of competing explanatory statements can be *decisively* eliminated.[12] What abductive reasoning can generate is a good, or persuasive, reason for accepting a particular explanatory statement, not a rationally *compelling* reason.

In order to understand how this good, or persuasive, reason for accepting a particular explanatory statement can be arrived at, it is important to appreciate that, on Lipton's account, abductive inference is a 'two-stage process': the first stage 'selects the plausible candidates'—that is, the set of competing explanatory statements worth considering; the second stage 'selects from among them' (Lipton 2004: 148–9, 59; see also Bird 2007: 424; cf. Toulmin 2003: 15–21). Let us elaborate.

Whenever an effort is made to show that a particular explanatory statement should be accepted, the pool of *possible* competing statements is 'very large, including all sorts of crazy explanations nobody would seriously consider' (Lipton 2004: 59). These cannot all be individually assessed against the evidence. Consequently, the first stage of abductive reasoning involves drawing on 'background beliefs' to select from the pool of *possible* competing explanatory statements those which are 'reasonably plausible' (Lipton 2004: 150).

[11] The background problem here is the underdetermination of theory by evidence (see Hollis and Smith 1991a: 55–6; Chernoff 2007a: 112–15; Stanford 2023). Bird attempts to get around this by arguing that hypotheses should be eliminated not only by reference to evidence, but also by reference to the broader corpus of scientific knowledge (Bird 2005, 2007). However, this raises the question of how this broader corpus of knowledge was itself established (Lipton 2007b: 452–3).

[12] These problems with Bird's argument are similar in kind to the problems we identified with Cartwright's contention that causal inferences are deductive, viz. it is impossible to be sure that the assumptions required to derive a causal claim deductively from evidence in fact hold.

In the second stage, those explanatory statements which have been selected through this initial filter are individually assessed against the evidence. However, it will often be possible to 'save' an explanatory statement that appears inconsistent with the evidence by rejecting the evidence as mistaken or by qualifying the explanatory statement in a way that reconciles it with the apparently inconsistent evidence. In order to determine which of a set of competing explanatory statements to accept, it is therefore often necessary to consider their 'explanatory virtues', or what Lipton also terms their 'loveliness' (2004: 59–62, 122). As Lipton puts it: better, or more 'lovely', statements 'explain more types of phenomena, explain them with greater precision, provide more information about underlying mechanisms, unify apparently disparate phenomena, or simplify our overall picture of the world' (2007a: 423; see also 2004: 122). He contends, in short, that when the evidence is not decisive an explanatory statement's 'loveliness' can still provide a good reason to accept it (and to reject 'unlovely' competitors).[13]

As this idea may be hard to grasp, it is worth briefly illustrating the point in relation to an example we used in Chapter 5. Suppose that burn patterns indicate that a house fire originated behind an electrical socket in the kitchen, supporting the explanatory statement 'a short circuit caused the fire'. The evidence supports this statement to the extent that it rules out competing explanatory statements—for example, that the fire was caused by an unattended candle in the bedroom. However, the evidence may not *decisively* rule out these competing statements. It is rationally possible to maintain, say, that the fire was caused by an unattended candle in the bedroom and that the burn patterns which appear to suggest otherwise have been misinterpreted or that they were generated through some as yet unknown process. Lipton's point is that although such possibilities cannot be definitively ruled out on evidential grounds, they can be rejected as 'unlovely'.[14]

Importantly, the judgements made in both these two stages have a degree of uncertainty attached to them. In the first stage, many *possible* explanations are never explicitly considered at all (Lipton 2004: 149). Consequently, if we assert that a given explanatory statement is the best of the competing explanatory statements, we can never entirely discount the possibility that

[13] Lipton (2004: 150–1) suggests that the background beliefs which are invoked, in the first stage, to help identify a small pool of 'plausible' explanatory statements worth considering, may themselves have been accepted, in part, on account of their 'loveliness'. If so, then considerations of 'loveliness' are drawn upon in both stages.

[14] Similarly, Lipton suggests that competing explanations for the peculiarly shaped tracks in the snow—for example, that they were 'caused by a trained monkey on snowshoes, or by the elaborate etchings of an environmental artist'—can be rejected on the grounds of their unloveliness (2004: 56). Ultimately, he is claiming that loveliness 'is a guide to likeliness' (2004: 61).

there is a superior alternative lurking in the pool of competing statements which were dismissed as implausible and hence not individually evaluated. In the second stage, the fact that the evidence is never decisive means that it will often be necessary to rely on judgements of 'loveliness'.[15] This introduces an additional element of uncertainty, for the possibility that an explanatory statement which has been rejected as 'unlovely' was in fact reliable can never be wholly discounted.

Keeping this in mind, we are now in a position to summarize what is involved in abductive reasoning:

- an explanatory statement, '*a* caused *b*', is judged to be consistent with the evidence;
- all the plausible explanatory statements which compete with '*a* caused *b*' are judged to be either inconsistent with the evidence or relatively 'unlovely';
- hence, the statement '*a* caused *b*' is judged to be reliable and is accepted as the best of the competing explanatory statements that we can generate concerning the occurrence of *b*.

This implies, in turn, that the degree of confidence it is sensible to have in a conclusion generated through abductive reasoning will depend on how confident it is possible to be in the following judgements:

i. The explanatory statement in question is consistent with the evidence;
ii. All plausible competing explanatory statements have been considered;
iii. All these competing statements are either inconsistent with the evidence or relatively 'unlovely'.

None of these judgements can be entirely certain. As Lipton (2004: 65) recognizes, an abductive inference will be overturned if a better explanation is uncovered, whether because new possibilities emerge (say, because the background beliefs drawn upon to identify the set of plausible competing explanatory statements are revised), because new evidence is discovered, or because judgements about relative 'loveliness' are revised (cf. Toulmin 2003: 94–5). However, the more confident it is possible to be in making these judgements, the more confident it is possible to be in the conclusion that is reached.

[15] A different way of thinking about this is that considerations of 'loveliness' may enter into our judgements as to how well a given statement is supported by the evidence. If a statement does not fit well with what we already believe about the world, then we may be more willing to reject it in the face of prima facie incompatible evidence, rather than wondering whether the evidence itself should be trusted.

The fact that abductive reasoning can provide a persuasive reason for accepting an explanatory statement, even though it cannot provide absolute certainty, is consistent with what we regard as a reasonable aspiration for a causal inference. Achieving certainty that '*a* caused *b*' (and hence that '*A* causes *B*') might seem desirable, but, as we pointed out when we considered Cartwright's suggestion that the reasoning which underpins some causal inferences might be deductive, this would require certainty that all competing accounts of how *b* was brought about should be rejected. Given the challenges we have identified in this section—that more alternative possibilities can always be imagined and that rejection of a statement on evidential grounds can never be decisive—such certainty is almost certainly illusory.[16] Hence, one important benefit of identifying causal inferences as involving abductive reasoning is that this facilitates systematic evaluation of just how persuasive specific causal inferences in fact are, recognizing that they can never be entirely certain.

Abductive reasoning and specific research methods

In concluding our discussion of causal inference and the reasoning which supports it, let us briefly consider how what we have said in this section relates to the application of specific research methods. The most important point, in this regard, is that *all* causal inferences rely, at least implicitly, on abductive reasoning, regardless of what method is used to identify and analyse the empirical evidence. Consequently, our contention that causal inference involves abductive reasoning does not favour any particular method for gathering and analysing empirical evidence. Whether researchers are engaged in process tracing or in analysing the results of an RCT, our confidence that the evidence supports a particular explanatory statement will depend on our confidence that all plausible competing explanatory statements can be rejected.

However, it is worth noting that when individual methods are applied, this may or may not explicitly involve Lipton's two-stage process: first, identification of a set of plausible competing explanatory statements worth considering; second, evaluation of these statements against the evidence. As it is typically practised in IR, process tracing is an example of a method which *does* explicitly proceed via these two stages: first, a shortlist of plausible explanatory statements is generated; second, the statements on this

[16] Dewey famously argued that, '[i]n scientific inquiry, the criterion of what is taken to be settled, or to be knowledge, is being so settled that it is available as a resource in further inquiry; not being settled in such a way as not to be subject to revision in further inquiry' (cited in Isacoff 2002: 620). Abductive reasoning might, at best, allow us to be this confident in its conclusions, but no more.

shortlist are evaluated against the evidence (see, for example, Tannenwald 1999; Bennett 2010; Goddard 2015).[17] By contrast, an RCT does *not* explicitly involve these two stages. We contend, nevertheless, that it facilitates abductive reasoning in support of a causal inference. It is worth briefly indicating why this is.

We argued, in Chapter 5, that an RCT is properly understood as providing support for an explanatory statement, the paradigmatic form of which is 'the treatment caused the observed difference in outcome across the treatment and control groups (if there was one)', by generating evidence which indicates that competing explanatory statements—statements to the effect that the treatment did *not* contribute to the observed difference in outcome across the two groups—can be rejected. It does this by using random assignment to try to ensure that there is no difference between the two groups other than whether the treatment was applied, the point being that if there is no other discernible difference between the two groups, then it is hard to generate a plausible competing explanatory statement: if there is no difference between the two groups other than whether the treatment was applied, then what, other than the treatment, could possibly have caused the observed difference in outcome across the two groups?

On this account, an RCT does not involve a *preliminary* effort to identify a plausible set of competing explanatory statements, *followed by* evaluation of these statements against the evidence. Rather, the setup is designed to rule out *any* competing explanatory statement; there is no need individually to identify and assess the plausible candidates. It is important to appreciate, however, that an RCT provides a 'good reason' to accept the statement 'the treatment caused the observed difference in outcome across the treatment and control groups' only if this statement is consistent with the evidence (if the treatment was applied to just one group and there was an observed difference in outcome across the two groups) and if all plausible competing explanatory statements are either inconsistent with the evidence (say, because they posit a difference between the two groups which was not in fact present) or relatively 'unlovely' (say, because they attribute the observed outcome to an unspecified or unknown mechanism).

Furthermore, if, as Rubin and Cartwright recommend, the robustness of the conclusion derived from an RCT is tested by manually checking whether known confounders were in fact evenly distributed across the treatment and

[17] As we noted in Chapter 5, however, we have some reservations about whether the 'alternative' explanations which are evaluated in some examples of process tracing in IR really *compete* with one another. See also Chapter 8.

control groups, then this *will* involve Lipton's two-stage process: first, plausible competing explanatory statements will need to be identified, based on background beliefs about the kinds of factors which might have made a difference to the outcome were they not evenly distributed across the two groups; second, these competing explanatory statements will need to be individually assessed by checking how these factors were in fact distributed across the two groups. In short, a successful RCT *facilitates* abductive reasoning in support of a causal inference, even if its design does not follow Lipton's two-stage process.

We do not have space to repeat this discussion in relation to every possible method that might be used to generate and evaluate empirical evidence in the course of causal inquiry, but we contend, nonetheless, that *all causal inferences rely on abductive reasoning, regardless of whether they explicitly follow Lipton's two-stage process*. To the best of our knowledge, this is not spelled out in any of the existing research methods literature, despite its obvious importance.

II. 'External validity', theory testing, and extrapolation to new settings

Our discussion so far has been concerned with what is sometimes termed the 'internal validity' of a causal inference—that is, its reliability. In effect, we have argued that the kind of reasoning required to generate a reliable causal inference—one with high 'internal validity'—is abductive. There is, however, another dimension of causal inference, as it is typically discussed in the research methods literature, namely 'external validity', or the extent to which the knowledge generated through a causal inference can be applied to other settings. This is important because, in many cases, the ultimate aim of causal inquiry is not to generate a reliable account of how a specific outcome was brought about, but rather to generate a reliable, causally informed prediction or policy prescription, perhaps in relation to a setting that has not yet been studied.

We pointed out, in Chapter 4, that an explanatory statement, paradigmatically '*a* caused *b*', implies a corresponding theory, '*A* causes *B*', the implication of which is that an *A*-type event will produce a *B*-type event if the conditions are right. We also pointed out, in Chapter 5, that once such a theory is well supported (as the result of a reliable causal inference), it can be applied to generate further explanations, predictions, and policy prescriptions. We warned, however, that this is more difficult than is commonly supposed, for

in order to be able to predict, say, that an *A*-type event will cause a *B*-type event, it is necessary to know *not only* that there is a propensity for *A*-type events to produce *B*-type events, *but also* that this propensity unfolds fully under the target conditions. It is therefore important to consider how one might reason about 'external validity'—about the applicability of causal findings to settings other than the setting in which they were initially discovered. As we shall see, ideas associated with the culture of generalization once again prove unhelpful.

We will begin by considering what the existing literature has to say about the 'external validity' of causal relationships. We will show that establishing, by means of an inference from a single study, that the causal finding(s) generated in this study can confidently be applied in other settings is extremely difficult. In practice, knowledge of what is called 'external validity' is most likely to arise through the systematic and laborious testing of established causal theories in diverse settings. As we will point out, however, this process necessarily involves further rounds of abductive reasoning. Moreover, any attempt to extrapolate causal findings to settings in which they have not yet been tested will necessarily rely on informed judgements and must therefore be tentative at best.

External validity

As the term is typically used, a causal inference is said to be *internally* valid, or to possess high (or strong) internal validity, if the reasoning which supports it is robust, affording a high degree of confidence that it is reliable (see, for example, Shadish, Cook, and Campbell 2002: 37–8; Cartwright 2007: 220; Seawright and Collier 2010: 334; McDermott 2011: 28; Cartwright and Hardie 2012: 46; Gerring 2012: 84; Johnson and Reynolds 2012: 175). Suppose it is inferred that, in an RCT, the treatment caused the observed difference in outcomes across the treatment and control groups. This inference has high internal validity if the evidence about these outcomes and about how the experiment was set up and run provides clear grounds for ruling out competing explanatory statements—that is, competing claims about what caused the observed difference in outcomes across the two groups (see Shadish, Cook, and Campbell 2002: 54).

By contrast, a causal inference is said to be *externally* valid if its conclusion can be 'generalized' beyond the context of the study from which it was generated (see, for example, Bueno de Mesquita 1985: 133; Shadish, Cook, and Campbell 2002: 37–8; Lucas 2003: 237; Seawright and Collier

2010: 330; McDermott 2011: 34; Dunning 2012: 293–4; Gerring 2012: 84; Goertz and Mahoney 2012: 216; Johnson and Reynolds 2012: 177–8; Beach and Pederson 2016: 246; Imai 2017: 50; Findley, Kikuta, and Denly 2021: 368; Egami and Hartman 2022: 1). The idea that causal inquiry should be designed so as to maximize external validity is widely accepted in IR and political science, yet in many leading research methods texts the concept is barely discussed.[18] Moreover, Findley, Kikuta, and Denly (2021: 367) found that only an 'exceptional few' articles across a range of leading social science journals 'contained a dedicated external validity discussion' (see also Egami and Hartman 2022: 1). The idea that causal inferences can possess 'external validity' therefore warrants further discussion, especially in light of our criticisms, in previous chapters, of the culture of generalization.

One immediate puzzle concerns what, precisely, is involved in the 'generalization' of the conclusion of a causal inference. We have argued that a causal inference provides support for an explanatory statement, paradigmatically '*a* caused *b*', which further implies a corresponding theory, '*A* causes *B*'. As we have explained, causal theories abstract from specific events occurring in specific times and places; they are, in this sense, already 'general' (see Lucas 2003: 238). It therefore makes no sense to think of 'external validity' as involving the 'generalization' of a causal theory. However, it is also unclear what it would mean to 'generalize' an explanatory statement, given that it is tied to specific events, occurring in specific times and places. A generalization of '*a* caused *b*' might be 'all *A*-type events caused *B*-type events', but it is hard to see what would justify such a generalization: certainly, knowledge that, in one specific time and place, an *A*-type event caused a *B*-type event would not, on its own, justify an assertion that all other *A*-type events have also caused *B*-type events. In any case, what is typically of interest, when external validity is discussed, is not what has happened in the past, but what will happen in the future: the question is whether other *A*-type events, occurring outside of the initial study context, will also produce *B*-type events. Yet, even if we allow that an explanatory statement, which relates to past events, might be 'generalized' to produce a statement about future events, it is again hard to see what would justify such a generalization.

[18] There is no index entry for 'external validity' (or similar) in any of the following well-known research methods texts: King, Keohane, and Verba (1994); Van Evera (1997); Ragin (2000, 2008, 2014); and Beach and Pedersen (2013). In numerous other such texts, the concept is mentioned only in passing (see, for example, George and Bennett (2005); Brady and Collier (2010); Goertz and Mahoney (2012); Johnson and Reynolds (2012); Beach and Pederson (2016); Imai (2017)). Shadish, Cook, and Campbell (2002), Dunning (2012), and Gerring (2012) and (2017) are exceptions: they each offer a dedicated discussion of external validity.

This suggests that the close association that is typically drawn between external validity and 'generalization' is somewhat misleading: whatever it is that judgements of external validity involve, they do not literally 'generalize' the conclusion of a causal inference. It also suggests that the tendency to present 'internal validity' and 'external validity' as 'twin' concepts obscures a crucial difference between them. Internal validity is an intrinsic property of a causal inference: the degree of internal validity possessed by a causal inference can be determined by inspecting the evidence and reasoning which supports it. To claim that a causal inference has high internal validity is to offer approbation for the reasoning which supports it and to endorse its conclusion. However, 'external validity' is *not* a property that a causal inference, *considered in isolation*, can possess. Assessing the qualities of a causal inference—of the reasoning and evidence which supports it—reveals nothing about its external validity. What is relevant for external validity is how representative the study context was of the contexts in which *A*-type events occur. Let us unpack this a little further.

When external validity is discussed, the presumption tends to be that causal inferences are derived from the comparison of outcomes within a study sample, paradigmatically in an experiment, and that the ultimate aim is to prescribe policy interventions for a target population (see, for example, Shadish, Cook, and Campbell 2002; Lucas 2003; Cartwright and Hardie 2012; Findley, Kikuta, and Denly 2021; Egami and Hartman 2022). This offers a restricted picture of how causal inferences can be generated; for example, it fails to encompass process tracing, in which causal inferences are arrived at through study of the events leading up to a specific event of interest, not through comparison of outcomes within a study sample. It is also important to bear in mind that the aim of causal inquiry is not always to prescribe policy interventions, or even to offer predictions; the aim may, instead, be to explain a specific outcome, or to test a proposed causal theory. Let us focus, nonetheless, on cases in which a causal inference is generated by studying the effect of a treatment in a study sample, as it is in an RCT.[19]

If a treatment is found to have had a certain effect, on average, within a study sample, then by definition it must have had this effect within that part of the population which constituted the sample! As Cartwright points out, however, it does not follow from this that the treatment would have had the same effect, on average, if it had been applied to the *whole* population, for

[19] Although RCTs are paradigmatically associated with 'harder' sciences, especially medicine and psychology, they are also used to generate causal inferences in IR. We will consider an example in Chapter 8. In any case, the logic of what we are saying here about RCTs carries across to any causal inference, regardless of what method is used to generate and analyse the supporting evidence.

while it had this effect in *part* of the population, in another part of the population it may have had a bigger or smaller effect, or even no effect or the opposite effect, due to interference by other factors (2007: 31; see also 1983: 23–5).[20] Consequently, the average effect of a treatment when applied to a whole population may be different in magnitude, or even in direction, to its effect in a sample drawn from that population.

What Cartwright is drawing attention to is that even if a causal inference has high internal validity, the estimate it provides of the average effect of a treatment in a study sample may not constitute a reliable estimate of the average effect this treatment would have (or would have had) if applied across the target population (see also Findley, Kikuta, and Denly 2021: 374). In order to address this, it is often suggested that a sample should be 'representative' of the target population (see, for example, Lucas 2003: 244; McDermott 2011: 37; Gerring 2012: 86; Findley, Kikuta, and Denly 2021: 374–6, 379–80). The idea goes roughly as follows:

- *if* an experiment (say) has generated an unbiased estimate of the average effect of a treatment within a study sample
- *and if* this sample was representative of the target population
- *then* the experiment will *also* have generated an unbiased estimate of the average effect the treatment would have (or would have had) if applied across the population as a whole.[21]

This suggests that whereas 'internal validity' is an intrinsic property of an initial causal inference, 'external validity' is a property of a *further* inference that the average effect of a treatment in a study sample (as revealed through the initial causal inference) will be replicated in a target population if the treatment is applied across that population. This *further* inference is 'externally' valid if, or to the extent that, the study sample is representative of the target population, affording confidence that what happened in the study sample constitutes a reliable guide to what would happen in the population at large. In order to be confident in advancing a prediction about the likely effect of a treatment on a target population, it is therefore necessary both that the initial causal inference concerning the effect of the treatment in the study

[20] This relates to the point we made in Chapter 4 that even if there is a propensity for *A*-type events to produce *B*-type events, in an open system there may be no observable correlation between the occurrence of events of these types, because of the interference of other factors.

[21] Note, however, that if the ultimate aim is to inform policy, information about the expected *average* effect of the treatment in the target population may not be sufficient: as King, Keohane, and Verba point out, before taking a risk on implementing a particular policy, policymakers might also 'wish to understand what the maximum and minimum causal effects, or at least the *variance* of the causal effects, might be' (1994: 84).

sample has high 'internal validity' and that the further inference concerning the relationship between the study sample and the target population has high 'external validity'.

In our view, this way of thinking clarifies what is typically being discussed under the heading 'external validity' and how it relates to 'internal validity'. There is, however, a significant problem with it. This concerns what it means for a sample to be 'representative' of a given population.

According to Findley, Kikuta, and Denly (2021: 379–80), a 'representative' sample 'unbiasedly represents a population'; this is most straightforwardly achieved through random sampling (see also Gerring 2012: 87).[22] This raises an immediate practical problem: in a discipline such as IR, even when causal inferences are derived from the comparison of outcomes across an identified study sample, this sample will *not* typically have been randomly selected from the target population (see Shadish, Cook, and Campbell 2002: 91–3; Cartwright and Hardie 2012: 128; Dunning 2012: 294–5; Gerring 2012: 87, 272).[23] However, there is also a deeper problem: even if a study sample is randomly selected from a target population, this is not sufficient to justify an inference that the average causal effect measured in the study sample will be replicated across the target population: the conditions which prevail in the settings in which a treatment is applied also matter.

Consider a hypothetical experiment which finds that, on average, the application of sunscreen increased, by a certain percentage, the time it took for the tests subjects' skin to start to burn. The question is whether it is reasonable to expect sunscreen to have the same effect, on average, that it is estimated to have had in the experiment, when it is used across a target population.[24] To see why random sampling is important, suppose that the average age and level of skin pigmentation of the test subjects was very different from that of the target population. If so, then it would clearly *not* be justified to interpret an estimate of the average effect of sunscreen in the experiment as an unbiased estimate of its effect within the target population, even if the experiment had high internal validity. However, it is also important to consider whether

[22] In essence, a sample will represent a population 'unbiasedly' if, on any given variable, the average value measured in the sample is no more likely to deviate from the value measured in the population in a positive direction than in a negative direction (and vice versa).

[23] This practical problem becomes even more serious if the population of interest is (partly) a future population, for future populations cannot even be sampled, let alone randomly sampled (Rubin 1974: 698).

[24] In practice, the focus will often be on the less demanding question of whether the effect will be in the same direction in a target population as it was in a study sample, rather than on the more demanding question of whether the effect will be of the same magnitude (see, for example, Shadish, Cook, and Campbell 2002: 90–1; Egami and Hartman 2022).

the contextual conditions under which sunscreen was applied in the experiment differ from those under which sunscreen will be applied in the target population. If, in the target population, sunscreen is often used when people are at the beach, exposing them to the potentially confounding effects of sea, sand, wind, and the like, but in the experiment such confounding was prevented, then the effectiveness of sunscreen when used across the target population is likely to differ from its effectiveness in the experimental setting. Simply ensuring that the test subjects were representative of the target population, through random sampling, would not control for this important contextual difference.[25]

In this hypothetical example, exposure to sea, sand, wind, and the like constitutes what Egami and Hartman term a 'context moderator' (2022: 6): a feature of the context which interferes with how a particular causal relationship unfolds in that context, generating variation in the outcome observed in this context as compared to the outcome observed in another context.[26] Crucially, as Egami and Hartman (2022: 6) acknowledge, 'there is no general randomization design' that can control for the influence of such factors.[27]

Egami and Hartman point out that the counterfactual effect of a treatment on an individual unit 'will be the same regardless of contexts, as long as the values of the context moderators are the same' across these contexts (2022: 6). In other words, if we could know that there are no causally relevant contextual differences between a setting in which a treatment has some desired effect and another setting in which we wish to produce the same effect, then we could be confident that the change of setting will not alter the effectiveness of the treatment (see Reiss 2009: 30–1).[28] But how could this be known? As Cartwright and Hardie point out, it is not enough that the two settings are 'similar'; what matters is that they are similar *in the relevant respect* (see 2012:

[25] Cartwright and Hardie spell out what this implies for policy prescription: what an experiment, such as an RCT, reveals is that 'a policy worked there, where the trial was carried out', but in seeking to manipulate the world what is needed is information about what will work *here*, in the situation of interest (2012: ix). Their point is that the inference that a treatment worked *in the RCT* is not, on its own, sufficient to justify a policy prescription; it is necessary also to consider whether the treatment will have a similar effect in the situation of interest.

[26] Shadish, Cook, and Campbell (2002: 87, 89–90) term this 'context-dependent mediation'.

[27] It is sometimes suggested that it might, in principle, be possible to study a sample which is randomly selected across settings, thereby arriving at an unbiased estimate of the effect of a treatment across the population of settings, though the practical obstacles are severe (see, for example, Shadish, Cook, and Campbell 2002: 32, 91–2; Findley, Kikuta, and Denly 2021). Yet, even if a population of settings could be defined and randomly sampled, this would still only reveal the average effect of a treatment across settings, whereas what we will typically want to know is the effect that the treatment will have in a *particular* setting where we are considering applying it.

[28] In addition, it will be important for external validity that the treatment and outcome measure applied in the study sample are the same as those applied in the target population (see Shadish, Cook, and Campbell 2002: 87–9; Findley, Kikuta, and Denly 2021: 371–2; Egami and Hartman 2022: 4–5). We set these issues to one side.

46–9)—that is, in respect of the presence or absence of context moderators, or what Cartwright and Hardie term 'support factors'.[29] But how can we even know what the relevant context moderators are unless we have studied what difference various changes of context actually make to the effectiveness of the treatment?

Let us summarize. We have suggested that existing discussions of 'external validity' are best understood as relating to the validity of a *further* inference, which may follow on from an initial causal inference, concerning the relationship between the study sample used to generate this causal inference and a target population. However, the difficulty of establishing what the context moderators are, for any given treatment, in any given setting, poses a serious challenge to any attempt to generate knowledge of 'external validity' through such an inference. As Rose McDermott puts it: 'A single study, regardless of how many subjects it encompasses or how realistic the environment, cannot alone justify generalization outside the population and domain in which it was conducted' (2011: 37; see also Cartwright 2007: 38–40; Cartwright and Hardie 2012: 40–1; Dunning 2012: 297).[30]

This does not mean that the idea behind 'external validity' should be abandoned. It will often be extremely valuable to consider whether a causal relationship which unfolded fully in one setting is also likely to unfold fully in another setting. It will be especially important if the aim is to generate a causally informed prediction or policy prescription for, as we explained in Chapter 5, a causal theory can be applied to generate such further knowledge claims only if we are quite confident that the conditions which prevail in the scenario to which these further claims relate are conditions under which the propensity described by the theory unfolds fully. The question is not whether knowledge of such matters is valuable, but rather how it can be produced.

Our analysis leads us to believe that it will be extremely challenging to establish, by means of an inference from a single study, conducted under a specific set of conditions, that a causal relationship which unfolded fully under these conditions will also unfold fully in some other setting. We agree with McDermott that, in practice, knowledge of the conditions under which a causal relationship of interest unfolds fully—knowledge about its 'external

[29] What Cartwright and Hardie (2012: 61–75) term 'support factors'—that is, factors that would have to be present for a treatment to have its intended effect in the target setting—are closely related to what Egami and Hartman term 'context moderators'.

[30] It is sometimes suggested that causal inferences derived from the comparison of outcomes within a relatively large or diverse sample *automatically* have higher external validity (see, for example, Bueno de Mesquita 1985: 133; Gerring 2011: 1144; Gerring 2017: 235). However, this is misleading, for the question is whether what happened in the study sample (however large or diverse this is) will happen elsewhere—that is, outside this study sample, in a target population (see McDermott 2011: 34).

validity'—is most likely to emerge out of systematic and laborious efforts to test this 'across diverse populations and different settings' (2011: 34; see also Lucas 2003: 239; Dunning 2012: 297; Gerring 2017: 236). In other words, knowledge of what is often termed 'external validity' is likely, in practice, to be the product of extensive and methodical theory testing. To appropriate a famous phrase from another context, developing knowledge of what is termed 'external validity' can be achieved only through the 'strong and slow boring of hard boards' (Weber 1948: 128): there are no easy shortcuts.

This suggests that a key component of causal inquiries will often be to 'test' well-supported causal theories across diverse settings, with the aim of expanding knowledge of the contexts in which the propensities described by such theories unfold fully.[31] This is in keeping with the logic of causal inquiry, as we laid it out in Chapter 5, for the more that is known about the various sets of conditions under which the propensity stated in some theory unfolds fully, the more confident it is possible to be in applying this theory in support of an explanation, prediction, or policy prescription. However, this raises two further questions. First, what kind of reasoning is required to 'test' a causal theory in diverse settings? Second, when is it possible confidently to 'extrapolate' to settings that have not yet been studied? Let us conclude our discussion by briefly considering each of these questions in turn.

Causal reasoning and theory testing

When thinking about the reasoning required to 'test' causal theories in diverse settings, it is worth being aware of an influential 'red herring' which often appears when the subject of 'theory testing' is discussed, and which is closely linked to the culture of generalization: the idea that causal theories should be 'falsifiable'. Let us give just one example. Having developed their account of 'causal inference', King, Keohane, and Verba follow this up by presenting a series of 'rules for constructing causal theories' (1994: 99–114). Their first rule is '*Construct Falsifiable Theories*'; they argue that 'a "theory" incapable of being wrong is not a theory' (1994: 100, emphasis original). They therefore advise researchers to 'design theories so that they can be shown to be wrong as easily and quickly as possible', adding that 'even an incorrect theory is better than a statement that is neither wrong nor right' and that

[31] This may, of course, require trade-offs between internal and external validity (see, for example, Shadish, Cook, and Campbell 2002: 96–102; Cartwright 2007: 220–1; McDermott 2011: 38; Gerring 2012: 271–3; Imai 2017: 50; Findley, Kikuta, and Denly 2021: 372): the conditions under which it is easiest to test whether a causal propensity of interest unfolds fully may not be the conditions under which it is most interesting to discover whether it unfolds fully.

one 'question that should be asked about any theory … is simple: what evidence would falsify it?' (1994: 100; see also Van Evera 1997: 20; Johnson and Reynolds 2012: 43).[32]

It is perfectly sensible to insist that a *descriptive* theory (or hypothesis) should be stated in a fashion that makes clear how it might be falsified. As we pointed out in Chapter 5, however, this does *not* carry across to *causal* theories. A causal theory, '*A* causes *B*', is not a generalization, but rather a propensity statement. Such a statement will not be 'falsified' if the propensity it describes fails to unfold fully in one specific situation, or even in many such situations. If a short circuit fails to produce a fire, this does not 'falsify' the theory 'short circuits cause fires'; it simply reveals a set of conditions under which the propensity for short circuits to produce fires does not unfold fully.

It is important to appreciate, moreover, that if a propensity fails to unfold fully under one set of conditions, this does not imply that it will also fail to unfold fully under another set of conditions in which one or more causally relevant features of the context has been altered in some way. In order to find out what difference contextual factors make for the unfolding of a causal propensity of interest, it is necessary to test this by varying these contextual factors systematically (or by systematically studying conditions in which they vary), identifying, on a case-by-case basis, under what conditions this propensity unfolds fully.

This provides an important clue about the reasoning required to 'test' a causal theory: it is abductive! The reason for this is quite simple.

Suppose that a theory, '*A* causes *B*', is well supported: a reliable causal inference has been generated from an initial study which indicates that, under conditions of the kind which prevailed in this initial study, a specific *A*-type event caused a *B*-type event. The question of interest is whether the known propensity for *A*-type events to produce *B*-type events will also unfold fully under conditions of some other kind. Suppose that under a different set of conditions from those which prevailed in the initial study an *A*-type event occurs but that, after an appropriate period of time has elapsed, a *B*-type event has not occurred. This indicates that under these conditions the propensity for *A*-type events to produce *B*-type events does *not* unfold fully. Now suppose, instead, that this *A*-type event was followed by a *B*-type event. Is this sufficient to show that under these conditions the propensity for *A*-type events to produce *B*-type events *does* unfold fully? Clearly, the answer must

[32] Some critics have argued that King, Keohane, and Verba overstate the importance of theory testing as compared, for example, to *developing interesting theories* (see McKeown 2004: 162; Ragin 2004: 127; Collier, Brady, and Seawright 2010b: 170). These critics do not, however, challenge King, Keohane, and Verba's basic claims about the nature and purpose of theory testing.

be 'no'. In order to be confident that this propensity unfolded fully, it will be necessary to show that the *A*-type event *caused* the *B*-type event. To do this, it will be necessary to generate a causal inference to this effect and, as we explained in Section I, this inference will involve abductive reasoning: reasoning which persuasively supports an explanatory statement to the effect that this *A*-type event caused the *B*-type event by indicating why competing explanatory statements can be rejected.

It follows from this that abductive reasoning is absolutely central to causal inquiry: it underpins not only the generation of causal knowledge from empirical evidence in any individual study, but also systematic efforts to test how broadly this knowledge can be applied. As far as we are aware, this crucial point is missed in existing discussions of 'causal inference' and of 'external validity'.

Extrapolation of causal findings to new settings

However extensively a causal theory is tested, in practice it will never be possible to know everything there is to know about the various sets of conditions under which the propensity this theory describes unfolds fully, or about the context-specific factors which support or interfere with its unfolding (see Cartwright and Hardie 2012: 69). After all, how could it be known that the effect of *all* causally relevant changes in context have been examined? How can it even be known which changes in context are causally relevant, except through trial and error (informed, of course, by existing knowledge)?

This creates a practical problem which arises whenever existing causal knowledge is applied in a new setting: however much is known about the various sets of conditions under which a propensity of interest unfolds fully, we can never be *certain* that the conditions which prevail in the new setting match, in every causally relevant detail, a set of conditions under which this propensity has been shown to unfold fully. In practice, therefore, applying existing causal knowledge in new situations will always involve an element of extrapolation. This extrapolation involves an explicit or implicit judgement to the effect that the setting of interest appears to be *sufficiently similar* to a setting which has been studied—similar in terms of the presence or absence of known or plausible context moderators—that the full unfolding of a propensity of interest in this new setting can be predicted with reasonable confidence.[33]

[33] Egami and Hartman (2022: 12) express a desire to avoid this kind of extrapolation, but we do not see how this is possible when considering settings which have not yet been studied.

To point out that *certainty* can never be achieved is not, of course, to deny that it might be possible to apply causal knowledge in new settings with some confidence. In general terms, it is possible to be more confident in judging that a new setting is relevantly similar to a setting that has previously been studied when more is known about the variety of contexts in which the propensity of interest does and does not unfold fully and when background knowledge about the kinds of contextual factors which might plausibly interfere with its unfolding is more advanced (see Cartwright and Hardie 2012: 127, 146). It is important to recognize, however, that any extrapolation of this kind involves an element of judgement, and that this requirement for judgement cannot be avoided through a statistical inference concerning the representativeness of a study sample, despite what the literature on 'external validity' may sometimes appear to imply.

Conclusion

In this chapter, we have done two things: we have explored how our account of the logic of causal inquiry relates to and informs existing debates, especially in the research methods literature, about 'causal inference' and 'external validity'; we have also explored the kind of reasoning required, in practice, to generate reliable causal inferences and to apply existing causal knowledge in new settings. By way of conclusion, we wish to emphasize three key takeaways from our discussion.

First, if the aim is to understand the nature and limits of causal reasoning, then language associated with the culture of generalization is quite unhelpful; in fact, it frequently points in precisely the wrong direction. This applies both to discussions of how causal inferences may be generated and to discussions of their 'external validity'. It is crucial to recognize that a causal inference is not, and does not involve, any kind of inductive generalization. It is also important to appreciate that it will typically be possible to apply a causal finding in a new setting only when much has been established about the influence that various contextual factors have on what effect the causal factor of interest produces (if any) in different settings; even when much is known, this kind of extrapolation remains intrinsically uncertain.

In saying this, our point is not to take issue with existing guidance on how particular methods for gathering and analysing empirical evidence can be put to work to generate reliable causal inferences. Our point is that the widespread presumption that causal inferences should produce some kind of 'general' knowledge, either directly or via a further inference about 'external

validity', distracts attention from the true nature of the reasoning that will, in practice, be required to support a persuasive causal inference and from how, in practice, it will be possible to develop knowledge of the settings in which causal propensities of interest do and do not unfold fully. As we will see in Chapter 8, the continued prevalence of language associated with the culture of generalization can also make it hard to express clearly and precisely what has (and, just as importantly, what has not) been established in any given causal study.

Second, abductive reasoning is absolutely central to causal inquiry. It is required to generate a persuasive causal inference and to explicate how and why it is persuasive. It is also required to expand knowledge of the various kinds of conditions under which given causal propensities unfold fully. This is not adequately recognized in the existing literature. In addition, it is important to appreciate that a wide range of methods for gathering and analysing empirical evidence in the service of causal inquiry can underpin abductive reasoning in support of reliable causal inferences. We will elaborate on this point in Chapter 8, where we will show how a range of different methods, from RCTs to historical inquiry, have been successfully applied in IR to generate causal knowledge.

Third, and finally, there is no escaping the need for considered judgement in causal inquiry. If the aim is to generate a causal inference, judgement will be required about whether all the plausible competing explanatory statements have been identified and about whether they can confidently be rejected. If the aim is to apply existing causal knowledge in new settings, judgement will be required about whether these settings are sufficiently similar to settings that have been studied, in relation to the contextual factors present within them, that the outcomes observed in these new settings will be similar to those observed in the previously studied settings. It is also important to emphasize that the requirement for judgement in causal inquiry holds regardless of what method for gathering and analysing empirical evidence is being used.

In Chapter 8 we will evaluate four examples of causal inquiry, in each of which a different method has been used to generate causal knowledge from empirical evidence. In doing so, we will show how the arguments advanced in this chapter can inform the practice of causal inquiry.

8
Examples

In the previous four chapters, we provided a new analysis of the main kinds of causal statements and how they relate to one another (see Chapter 4), of the basic logic of causal inquiry (see Chapter 5), of causal explanation (see Chapter 6), and of causal reasoning (see Chapter 7). This analysis has departed from the conventional wisdom in various important ways; most notably, we have pointed out repeatedly how language associated with the 'culture of generalization' often seriously impedes accurate understanding. However, our arguments have, necessarily, proceeded at quite a high level of abstraction and the examples we have provided have often been hypothetical, in order to ensure that they clearly illustrate the point being made. In this chapter we will therefore round out our analysis by showing how the arguments we have developed over the previous four chapters can illuminate some concrete examples of high-quality causal inquiries in IR. We aim to show how paying explicit attention to the key ideas contained within our analysis can help to make sense of how reliable conclusions are arrived at in high-quality examples of causal inquiry, while also bringing out important themes that may often be neglected and highlighting where further work might be done to render the conclusions reached even more secure.

To this end, we will consider four examples of high-quality causal inquiry in IR. These have been chosen to illustrate how a range of methods for gathering and analysing empirical evidence can be applied to generate useful causal knowledge and also to illustrate the range of goals that may be pursued through causal inquiry, from developing and testing causal theories to explaining particular outcomes and generating causally informed predictions and policy prescriptions. We make no claim that these studies are 'representative' of contemporary causal inquiries in IR, but collectively they enable us to discuss a range of key issues to which we have drawn attention in previous chapters and which we consider to be of crucial significance for causal inquiry in IR. Reviewing these four very different studies together also allows us to draw out some common themes and challenges which arise in the course of causal inquiry, regardless of which methods for gathering and analysing empirical evidence are used.

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0009

In relation to each example, we will consider three broad issues. First, we will assess how secure their conclusions are, paying attention both to the 'internal validity' of causal inferences generated from the analysis of empirical evidence and, where relevant, to questions of 'external validity'. Second, we will scrutinize the language and reasoning used to reach these conclusions, considering how carefully and accurately causal language is used and how explicit the reasoning is that supports particular conclusions. Third, in light of this, we will seek to identify ways in which the studies we examine might be further developed.

Our discussion will unfold in two parts. In Section I, we will consider two examples of causal inquiry in which conclusions are drawn from the analysis of what Collier, Brady, and Seawright (2010b: 184–8) term 'data-set observations'—that is, from the comparison of outcomes associated with different values of an independent (or treatment) variable. In one of these studies (Fair, Littman, Malhotra, and Shapiro 2018), causal inference is facilitated by random assignment of test subjects to treatment and control groups in the context of an RCT; in the other study (Lyall 2009), causal inference is facilitated by the assumption of as-if random assignment in the context of a natural experiment.[1] In Section II, we will consider two examples of causal inquiry in which conclusions are drawn from the analysis of what Collier, Brady, and Seawright (2010b: 184–91) term 'causal-process observations'—that is, from the analysis of evidence about the causal history of some specific event that is being explained.[2] One of these studies (Goddard 2015) is by a scholar who explicitly describes herself as employing process tracing; the other (Trachtenberg 1991a) is by a well-known historian. In both sections, we will consider the strengths and weaknesses of the individual studies in detail, while also seeking to derive lessons about the application of the methods they employ.

[1] There are also many examples of causal inquiries in IR which involve analysis of data-set observations that were not generated through random assignment and in relation to which an assumption of as-if random assignment is implausible. Typically, such studies rely on statistical controls to justify an assumption that 'assignment' and outcome were conditionally independent (see King, Keohane, and Verba 1994: 94–5; Collier, Seawright, and Munck 2010: 44–9; Brady 2010: 74–6; Collier, Brady, and Seawright 2010b: 172–7). However, as we noted briefly in Chapter 2, this approach has increasingly been superseded by what Dunning (2010, 2012) terms 'design-based' approaches, which include experiments and natural experiments of the kind we consider here.

[2] There are various other ways of demarcating the distinction between approaches which draw on 'data-set' and 'causal-process' observations. In a broad sense, the former are typically 'quantitative', while the latter are typically 'qualitative', but this need not necessarily be the case. Sometimes 'cross-case' analysis or methods are contrasted to 'within-case' analysis or methods (see, for example, Collier, Mahoney, and Seawright 2004: 86), but 'data-set observations' need not necessarily arise from multiple distinct cases. Ragin (2014) uses the suggestive terms 'case-oriented' and 'variable-oriented' to describe research using causal-process and data-set observations, respectively.

I. Examples of causal inquiry in IR drawing on 'data-set observations'

RCTs are widely seen as constituting 'the gold standard' for causal inquiry in the social sciences (Cartwright 2010: 59; Imai 2017: 48–9). As we have explained previously, they can underpin highly persuasive causal inferences. Random assignment of test subjects to treatment and control groups aims to ensure that there is no difference between these two groups other than the fact that the treatment is applied to one but not the other.[3] If this is achieved, and if the outcome, on average, is different in the treatment group from the control group, then this constitutes powerful evidence in favour of the claim that the treatment caused this difference in outcomes, for it is difficult to see what else could have caused it. Competing explanatory statements, to the effect that this difference in outcomes was caused *not* by the treatment, but rather by something else, can hence be ruled out with considerable confidence.

RCTs are widely used in applied sciences in order to test the effectiveness of a particular policy intervention as a means to a desired end (see Cartwright and Hardie 2012). Due to practical and ethical constraints, it has been relatively rare to see RCT-based empirical research in IR (see Berman, Felter, and Shapiro 2018: 47–9). However, there are some exceptions to this rule. In this section, we consider one such exception: Fair et al.'s (2018) use of an RCT to study determinants of support for militant groups in Pakistan.[4]

In practice, there are many causal questions of interest in IR in relation to which it is impractical or unethical to run an RCT. In part for this reason, there has also been increasing interest in research designs which do not involve random assignment by the investigator, but rather draw on observational data which, it is claimed, have naturally and fortuitously been generated as-if, or as good as, randomly. Such research designs are aptly called 'natural experiments' and we also consider an example of this kind of study: Jason Lyall's (2009) analysis of the effect of indiscriminate violence on insurgent activity in Chechnya.[5]

[3] As we pointed out in Chapter 7, random assignment guarantees that an RCT produces an *unbiased* estimate of the effect of the treatment; however, it is necessary to check manually whether potential confounders are in fact evenly distributed across the treatment and control groups.

[4] Some other exceptions which are worthy of note include Beath, Christina, and Enikolopov (2012, 2013); Lyall, Blair, and Imai (2013); Crost, Felter, and Johnston (2016).

[5] Thad Dunning (2012: 41–102) distinguishes what he terms '*standard* natural experiments', such as Lyall's, from certain other methods of data analysis which are also sometimes described as 'natural experiments': in particular, so-called Regression-Discontinuity (RD) and Instrumental-Variables (IV) approaches. Unlike 'standard' natural experiments, these involve efforts by the investigator to secure an as-if random assumption by recourse to ingenious data-processing methods. In our view, the range of noteworthy issues arising from the application of such research designs does not differ significantly from

In considering these two studies, we will develop three key themes. First, we find the causal inferences which are derived in each study to be quite persuasive, though we note how much harder it is to develop secure conclusions in a natural experiment as compared with an RCT. Second, we show how important it is to be precise in the use of language when describing causal findings. Language associated with the culture of generalization, where it appears, is consistently misleading; however, there are some interesting passages in both studies which appear to display an implicit awareness, on the part of the authors, that causal theories state propensities, not generalizations. Third, noting that the aim of RCTs and natural experiments is usually to provide support for a causal theory, in the hope, ultimately, of being able to provide support for some kind of intervention or policy prescription, we pay careful attention to what the two studies say about matters of external validity.

Using an RCT to study determinants of support for militant groups in Pakistan

Fair et al.'s study (2018) built on other recent academic literature, especially a survey by Blair, Fair, Malhotra, and Shapiro (2013), which had found, contrary to the received wisdom, that poverty and exposure to violence were *negatively* correlated with support for militant groups in Pakistan (Fair et al. 2018: 58; see also Berman, Felter, and Shapiro 2018: 195, 230). The conventional wisdom, Fair et al. explain (2018: 58, see also 75), is that *income* is 'negatively correlated with support for violent political organizations. Exclusion from economic progress and feelings of injustice induce grievances, and therefore those who feel excluded from and disadvantaged by the existing political hierarchy are more likely to support non-state actors trying to disrupt it'.

In other words, the conventional wisdom has it that the poor, who are disadvantaged by the status quo, are *more* supportive of militant groups (poverty is *positively* correlated with support), while the better off are

those which arise in relation to RCTs and 'standard natural experiments', so in the interests of brevity we set such designs to one side in our discussion. Examples of these methods being applied in IR include Crost, Felter, and Johnston (2014) (using RD) and Wright, Condra, Shapiro, and Shaver (2017) (using IV). These methods are also discussed extensively by Dunning (2012). Outside of IR, see also Thistlewaite and Campbell (1960) on the RD approach and Angrist (1990) on the IV approach. Angrist won a Nobel prize in economics for his work: https://www.aljazeera.com/news/2021/10/11/nobel-prize-in-economics-goes-to-natural-experiments-pioneers (accessed 28 July 2023).

less supportive (income is *negatively* correlated with support). By contrast, Blair et al. (2013) had found that poverty was *negatively* correlated with support for militant groups in Pakistan: the poor were *less* likely to be supportive.

Fair et al. (2018: 72) identified a similar correlation to that observed by Blair et al. (2013): they found that 'the poorest individuals in Pakistan held the most negative attitudes toward militant groups'. They then went on to investigate what 'mechanism' links poverty to anti-insurgency attitudes in Pakistan (see Fair et al. 2018: 57–9). In particular, they were interested in the 'psychological effects of *perceptions* of poverty and violence on attitudes' (2018: 58). They explored whether those who perceive themselves to be relatively poor and those who believe that their country is relatively more violent take a more negative view of militant groups. They hypothesized that 'both relative poverty and higher perceived levels of violence in one's country will decrease support for militancy' (2018: 59). They appear to be thinking that those who *perceive* themselves as relatively poor, or who *believe* their country already to be relatively violent, will fear that they are more likely to be harmed by insurgent violence and hence take a more negative view of militant groups who propagate this violence.

In order to explore these issues, Fair et al. conducted very carefully designed RCTs, thereby providing 'the first experimental evidence for the impact of perceptions of poverty and political violence on attitudes toward militant organizations' (2018: 59). Their ultimate aim was to provide a more robust basis for policymaking in regard to the provision of aid in conflict environments. As they explain (2018: 59):

> If relative perceptions of poverty are important determinants of support for militant groups, then economic development *per se* may not yield full benefits if people still feel relatively poor or if these interventions increase aspiration levels. In addition, if perceptions influence attitudes, then policy interventions informing people about how militants introduce violence and economic hardship into communities may be especially effective in reducing support.

This exemplifies the desire of much of the quantitative literature in conflict studies to contribute to policymaking. It is also interesting that Fair et al. are trying to show how, contrary to the supposition of some critics (see, for example, Dessler 1991; Kurki 2006; Mearsheimer and Walt 2013), quantitative approaches can address questions about the causal mechanisms that underpin observed correlations.

In their RCTs, Fair et al. compared the support for various militant organizations among test subjects randomly assigned to different treatments.[6] Those in one group were given the impression that they belonged, within Pakistan, to the 'relatively poor' class, and those in the other group were given the contrary impression. Additionally, one group was told that there was more insurgent-related violence in Pakistan than in another country in the region, and the other group was given contrary information (Fair et al. 2018: 67–70). The investigators then compared these groups' attitudes towards the three militant organizations.[7] To cut a complex story short, this is what they found:

> [R]espondents experimentally induced to feel poorer and to perceive Pakistan as more violent than neighboring countries exhibit decreased support for militant organizations, demonstrating that *perceptions of relative poverty and violence can influence attitudes*. Increasing perceptions of Pakistan as a violent country had an even greater effect than inducing individuals to feel poor. (2018: 59–60, emphasis added; see also 71–6)

Assessing Fair et al.'s study: Internal validity

In our view, the fact that Fair et al. employed a carefully designed RCT gives their study high internal validity. They not only assigned their test subjects randomly to treatment and control groups, but also actively checked the distribution of possible confounders across these groups and conducted further robustness checks (Fair et al. 2018: 70, 75). Thus, for example, they checked that the treatment and control groups were in fact balanced in relation to variables such as the gender and age of the respondents and their level of education and wealth (Fair et al. 2018: online appendix). They also checked that the results were replicated in key subgroups: for example, respondents living in the areas that had experienced the most insurgent violence in recent years (Fair et al. 2018: 75). Expressed in the terms we have used previously: Fair et al. generated a highly persuasive causal inference because they demonstrated that plausible competing explanatory statements—those suggesting that priming the test subjects with information about their relative poverty

[6] The test subjects were selected not only from the four main provinces but also from six of the seven 'agencies' in the Federally Administered Tribal Area (Fair et al. 2018: 63). These subjects' attitudes towards three militant organizations were examined: Sipah-e-Sahaba Pakistan, the Pakistan Taliban, and the Afghan Taliban. Al Qaeda was excluded because of heightened sensitivity about this group's actions at the time the survey was conducted, in 2012, given the recent (2011) assassination of Osama Bin Laden (Fair et al. 2018: 60–1).

[7] Technically, they combined an endorsement experiment, designed to ascertain indirectly the test subjects' level of support for each militant organization (see Fair et al. 2018: 64–7), with an RCT, testing how this level of support was affected by priming the test subjects with information about their relative poverty and/or about the relative level of violence in Pakistan.

and vulnerability to violence did *not* causally influence their level of support for the three militant organizations—can be rejected. The investigators showed quite persuasively that it is hard to account for the experimental results unless priming the test subjects with this information *did* causally influence their level of support for the three militant organizations.

Although we endorse their findings, the language in which Fair et al. describe their results illustrates a point we have made previously about the importance of being precise in how causal language is used. In some places, they expressed their findings very carefully and accurately; in other places, their language is looser and hence potentially misleading. Let us elaborate.

What Fair et al.'s experiment revealed is that *the test subjects*' perceptions of their relative poverty and vulnerability to violence (as manipulated in the experiment) causally contributed to *their* (the test subjects') negative attitudes towards the three militant groups, as measured in the experiment conducted in 2012. It is important to appreciate that this explanatory statement concerns what caused the outcome observed *in this experiment*. To say that the experimental findings had high internal validity is hence to say that the investigators' causal interpretation of what happened *in this experiment* was justified.

Fair et al. were therefore right to use the past tense when, in the summary of their findings quoted above, they reported: 'Increasing perceptions of Pakistan as a violent country *had* an even greater effect than inducing individuals to feel poor' (2018: 60, emphasis added). This use of the past tense emphasizes that their findings relate to what *happened in the experiment*. However, in the same passage they also wrote, using the present tense: 'respondents experimentally induced to feel poorer and to perceive Pakistan as more violent than neighboring countries *exhibit* decreased support for militant organizations' (2018: 60, emphasis added). It would have been more precise to say 'respondents experimentally induced to feel poorer and to perceive Pakistan as more violent than neighboring countries *exhibited* decreased support for militant organizations'.[8] Asserting, in the present tense, that appropriately primed respondents 'exhibit' decreased support for militant organizations makes it sound as if 'respondents' *generally*—that is, any and all respondents who are appropriately primed—will (on average) exhibit decreased support for militant organizations. This is not justified on the basis of Fair et al.'s study.

[8] Elsewhere, similarly, they describe their results as suggesting 'that part of the observed relationship between poverty and opposition to militant groups *stems* [rather than *stemmed*] from psychological feelings of relative deprivation' (2018: 73, emphasis added').

When we presented Fair et al.'s summary of their results (above), we italicized what we consider to be another important passage, in which they claimed that their findings demonstrate that '*perceptions of relative poverty and violence can influence attitudes*' (2018: 60). Although Fair et al. did not draw attention to its importance, what we consider significant here is the word 'can'. What their experiment demonstrated is not only that priming the test subjects to believe that they were poorer, or more vulnerable to insurgent violence, *caused* them, on average, to have more negative attitudes towards militant organizations, but also, consequently, that there is a *propensity* for those who believe that they are poorer, or more vulnerable to insurgent violence, to have more negative attitudes towards militant organizations. We therefore consider their use of the word 'can' to be quite apt: to say that perceptions of relative poverty and violence *can* influence attitudes is to say that it *sometimes* does, but also to acknowledge, at least implicitly, that it sometimes may not.

This naturally leads us to consider questions of external validity, but before we do this it is important to point out that Fair et al. were able to field an RCT because they were asking a distinctive kind of question: one about the influence of test subjects' perceptions on their attitudes. To address this question, it was possible for the investigators to conduct randomized tests, in which the treatment and control groups were 'primed' in contrasting ways. As Fair et al. acknowledge (2018: 60), such tests are common in political psychology research. However, they are possible only because they involve a treatment (a psychological prime) that can, practically speaking, be randomly assigned and because applying or withholding this treatment does not raise any deep ethical challenges. This will not be the case in relation to many of the kinds of causal questions that are of interest in a discipline such as IR. Hence, although Fair et al.'s study shows that RCTs *can sometimes* be used very successfully to address causal questions in the study of world politics, it does not show that RCTs can *readily* be used in relation to a wide range of causal questions in IR.[9]

Assessing Fair et al.'s study: External validity

Fair et al. do not explicitly discuss the external validity of their findings. In some passages, they do, nevertheless, reveal awareness of the difficulties of extrapolating to settings other than the one they studied. For example,

[9] When it is not practically possible for academic researchers to conduct experiments on the subjects directly, they may sometimes be able to make use of experimental data gathered by researchers working for the relevant governments or international organizations. See, for example, Crost, Felter, and Johnston's study of the effectiveness of 'conditional cash transfer programmes' in the Philippines (2014).

they note that '[i]f individual support for militancy can be moved by subtle primes manipulating reference points for poverty and violence, then perceptions *may* play an important role in determining support in the real world' (2018: 59; emphasis added). Yet, in other passages, they appear, somewhat carelessly, to slip into unjustified generalization. For example, they claim that their 'study reports the first large-scale experimental evidence that there is a causal relationship between relative poverty, perceived violence, and level of support for violent organizations *in the developing world*' (2018: 75, emphasis added). Here, their reference to 'the developing world', and also their use of the present tense ('there *is* a causal relationship'), makes it appear as if their results hold beyond the confines of their experiment. They do acknowledge that 'the purpose of a survey experiment is to isolate a hypothesized relationship in a controlled setting in hopes of understanding how this phenomenon plays out under external conditions' (2018: 60), but they do not elaborate on this point, writing, at times, as if the 'controlled setting' permeated not only the specific context of their experiment, but also 'Pakistan', and even the entire 'developing world'.

In this regard, a particularly problematic claim made by Fair et al. is that their experimental findings 'explain' outcomes observed in other settings. For example, they state: 'demonstrating that support for militant groups can be shifted via the adjustment of reference points provides evidence that at least some of the effects of poverty identified in previous research are indirect and the result of perceptions' (2018: 68). It is worth reiterating once again that the term 'can' here is quite apt: Fair et al. have indeed demonstrated that support for militant groups *can* be shifted via the adjustment of reference points. However, it is a long way from demonstrating this to showing that *in some other case* where there was a negative correlation between poverty and support for militant groups, that was due, in part, to the effect of perceptions. Fair et al. do not even show that, in these other cases, the relevant perceptions were present, let alone that they causally contributed to the level of support for militant groups in these cases. Nevertheless, they repeat this kind of claim about how their findings justify particular interpretations of other findings, in other studies, several times (see, for example, Fair et al. 2018: 74).[10]

What, then, can justifiably be claimed about the policy implications of Fair et al.'s study? They suggest, on the basis of their findings, that a media campaign to influence citizens' perceptions '*could be* a promising policy intervention' (2018: 60; emphasis added). This strikes us as sensible: the

[10] By contrast, speculating about whether the results found in the experiment *might* also be found elsewhere is reasonable, so long as it is clearly identified as speculation (see, for example, Fair et al. 2018: 77).

word 'could' indicates an implicit acknowledgement that a media campaign is not quite the same thing as the application of a psychological prime in an experiment and also that context matters: this kind of campaign might work under some conditions, but may not be very effective under other conditions, for example, if it is initiated by a government which has lost trust among its citizenry. It therefore seems reasonable to suggest that policymakers should *consider the possibility* that, contrary to the received wisdom, improving the material well-being of individuals may not necessarily reduce support for violent political organizations. However, more work would have to be done to expand knowledge of the range of settings under which the causal relationship Fair et al. have discovered is reproduced, probably through further studies of the kind they conducted, but in different settings.[11]

Using a natural experiment to study the effect of indiscriminate violence on levels of insurgent activity in Chechnya

The classic example of a 'natural' experiment is John Snow's epidemiological study, in which he presented strong evidence that cholera is transmitted by contaminated water (1965). Table 8.1 shows household death rates during London's cholera outbreak of 1853–4 in the sixteen subdistricts where water was supplied competitively by two companies:

Straight away, a significant difference in death rates hits the eye: 315/10,000 for those unfortunate households supplied by the Southwark and Vauxhall Water Company, compared with 37/10,000 for those lucky ones receiving

Table 8.1 Death rates from cholera by water-supply source in the London epidemic of 1853–4, adapted from Snow (1965)

	Houses supplied	Deaths from cholera	Deaths per 10,000 houses
Southwark & Vauxhall Water Company	40,046	1,263	315
Lambeth Water Company	26,107	98	37

[11] Berman, Felter, and Shapiro (2018) offer something of a meta-analysis of existing studies. However, they tend to assume that if a result has been found to hold in several different countries, it must therefore hold 'generally'; this strikes us as unjustified.

water from the Lambeth Company. Importantly, Snow found that both these companies collected water from the River Thames, but that while the Southwark and Vauxhall Company had kept its intake pipe downstream, the Lambeth Company had moved its intake pipe further upstream the year before the outbreak, thereby obtaining 'a supply of water quite free from the sewage of London' (Snow 1965: 68).

However, *from the viewpoint of the research design*, what mattered was not the difference in the death rates as such, stark though it was, but the fact that the two groups of houses—one receiving water from one company, the other from the other—were generated *as good as randomly*. Snow noted that it was unlikely that the residents consciously chose their water provider; they may not even have been aware who their provider was (1965: 75). He explains this situation very clearly:

> The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time where the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. *Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies* … As *there is no difference whatever in the houses or the people receiving the supply of the two Water Companies, or in any of the physical conditions with which they are surrounded*, it is obvious that no experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this, which circumstances placed ready made before the observer. (Snow 1965: 74–5; emphasis added)

In IR, there are few clear examples of such a pure natural experiment (Hyde 2007: 46), but Lyall's study (2009) of the effect of Russian attacks on Chechen villages provides a valuable illustration.[12]

Lyall (2009: 331) describes himself as 'testing' the 'proposition' (what we would call a causal theory) that 'a state's use of indiscriminate violence incite[s] insurgent attacks'. As he explains, it is commonly presumed that use of indiscriminate violence *does* incite insurgent attacks; indeed, he cites a review (Kalyvas 2006) which identified 'no fewer than one hundred studies and forty-five historical cases in which a state's reliance on collective targeting of the noncombatant population provoked greater insurgent violence'

[12] Hyde (2007) is another helpful illustration. From the Policy Science perspective, Galiani and Schargrodsky (2010) is noteworthy. Another illustration, from Comparative Politics, is Posner (2004), but see also Dunning's critique (2012: 300–2).

(Lyall 2009: 331). However, other studies have found that 'a state's use of indiscriminate violence can actually suppress insurgency, at least under certain conditions'. Lyall also points out that state militaries have used indiscriminate violence 'with alarming regularity' (2009: 332).[13] This makes it practically and ethically important to identify the causal dynamics in play, but it also reinforces the puzzle: why is this tactic used so often if, as the conventional wisdom suggests, it is counterproductive?

Lyall tests the 'presumed relationship between a state's indiscriminate violence and insurgent attacks' in relation to a new case: 'Russian artillery strikes on populated settlements in Chechnya (2000 to 2005)' (2009: 332). This case was chosen in part because, in Lyall's view, it constitutes a 'most likely' case for the link, postulated by the conventional wisdom, 'between indiscriminate repression and increased insurgent attacks' (2009: 339), but also, and more importantly, because the Russian artillery strikes were 'uncorrelated with key spatial and demographic variables thought to drive insurgent-attack propensity' (2009: 338): in short, they appear to have been randomly targeted. This enabled Lyall to consider the case under study as providing a 'natural experiment'.

To cut a complex story short once again, what Lyall found is that indiscriminate violence did *not* incite insurgent attacks. In fact, those Chechnyan villages which were subjected to indiscriminate shelling by Russian forces experienced a 24.2% *reduction* in insurgent attacks relative to comparable villages which were spared (2009: 349). This might seem somewhat underwhelming when contrasted to the dramatic difference in deaths per 10,000 houses which Snow discovered. However, what was decisive in Snow's study was the fact that the allocation of the households to the two groups was as-if, or as good as, random. How secure, by comparison, is Lyall's finding?

How Lyall sought to secure the internal validity of his causal inference

As Lyall explains, the Russian practice of shelling, known as 'harassment and interdiction', was haphazard. Not only was it 'designed to consist of barrages at random intervals and of varying duration on random days *without evidence of enemy movement*' but, Lyall noted, '[p]erhaps most importantly, *these artillery strikes are uncorrelated with many of the variables commonly cited as explaining insurgent violence*' (2009: 343; emphasis added). In effect, Lyall argues, Russian decisions to target some villages

[13] He defines 'indiscriminate violence' as 'the collective targeting of a population without credible efforts to distinguish between combatants and civilians' (2009: 358).

and not others approximated 'a lottery assignment mechanism' (2009: 343). To reinforce this interpretation, he noted that the shelling was at times a result of the Russian soldiers being drunk or high on drugs, such that they even *shelled themselves* accidentally (2009: 345)! This haphazardness is thought by Lyall to have contributed towards ensuring that the villages with those features which are more, or less, strongly associated with insurgent violence did not, in the phraseology of the research methods literature, 'self-select' either into the 'treatment' (shelled) or the 'control' (not shelled) group; this, of course, is a key consideration in as-if random natural experiments, for such 'self-selection' would bias an estimate of the treatment effect and hence undermine the internal validity of the causal inference.

But how secure is this as-if random assumption in this case, and how persuasive is Lyall's inference that the use of indiscriminate violence by Russian forces *caused* a reduction in insurgent attacks? While the Russian strategy may have been haphazard with respect to the duration of shelling, the intervals between shelling cases, and the days on which shelling took place, the supposition that the shelled villages became targets as-if randomly—'as if' because the claim here is not that the Russians were *literally* using a randomization procedure to select the targets—is based only on a few pieces of *suggestive* evidence. When compared with Snow's study, the reasoning that sustains the claim of as-if randomness in the generation of the data seems somewhat less secure.

Lyall was fully aware of this. Quoting Susan Hyde, he rightly noted that 'the burden rests on the researcher to demonstrate that the treatment—whose assignment was clearly not supervised by this author—can be treated as if its assignment were indeed random' (2009: 345; see Hyde 2007: 46). Lyall therefore checked whether the 'treatment' (shelled) and 'control' (non-shelled) groups of villages were balanced in relation to a number of potentially confounding features, in particular, eight demographic, spatial, and conflict-related variables commonly used to explain patterns of insurgent violence, such as the population size of the village, the altitude of the village, and so forth (see 2009: 341–2, 345–7).[14] He concluded that there was, on average, no concerning difference between the shelled and non-shelled villages which might account for the difference in level of insurgent attacks they experienced (2009: 345–7).

[14] The background assumptions here are that the larger the population, the less difficulty the insurgents would face in recruiting fighters, making their suppression that much harder; that a higher altitude suggests a rough terrain, advantaging the insurgents and hence increasing the probability of insurgent violence; and so on (Lyall 2009: 341–2).

However, Lyall's concern was not entirely assuaged. He worried that 'the most important determinants of insurgent violence may be war-induced dynamics that arise out of the interaction of Russian and rebel strategy' (2009: 338). As he explained: checking that the two groups of villages were balanced in terms of plausible confounders still did not 'eliminate the possibility that target selection reflects private Russian information about specific villages' conflict propensities' (2009: 347).[15] In order to rule out this possibility, he supplemented his natural experiment with a 'matching' strategy (2009: 347). This is how he proceeded.

A total of 147 villages or population centres, falling within range of at least one artillery base, were studied. Of these, 73 villages were attacked at least once and 74 were spared (Lyall 2009: 342–3). During the period under study (2000–5), there were 159 *cases* of indiscriminate shelling from two Russian bases in Chechnya.[16] Those 159 indiscriminate shelling cases in the victimized villages were then *matched* for comparison with the equal number of no-shelling cases in spared villages. Lyall explains that the non-shelled villages were matched with shelled ones on (i) the eight potentially confounding variables mentioned above, and (ii) two additional measures of wartime practices, the latter of which he claimed would take care of *conflict dynamics* potentially skewing his estimate of the causal effect of the shelling. 'In brief', Lyall reported, 'our sample now consists of two similar populations across both preshelling fixed covariates as well as important war-induced factors' (2009: 349). Given the care he took in setting up his comparative analysis, it is unsurprising to find that there were no concerning differences, on average, between the two groups (see 2009: 346).

In this way, Lyall not only sought to show that he was justified in treating Russian shelling as if it had been randomly assigned to Chechen villages, but also attempted to reinforce the conclusion suggested by his natural experiment: that Russian indiscriminate violence caused a reduction in insurgent attacks. By checking systematically that there were no concerning differences between shelled and non-shelled villages which might account for the reduced level of insurgent attacks experienced by the shelled villages he, in effect, ruled out competing explanations for this reduction—that is, explanations which compete with the claim that it was *caused* by Russian

[15] Lyall seems to be thinking that Russians gathering information about specific villages' conflict propensities in the course of the conflict constitutes an example of what he rather vaguely calls 'war-induced dynamics'.

[16] It appears that *every* case of Russian shelling was indiscriminate, such that the contrast is between shelled villages, which were each subjected to indiscriminate violence every time they were shelled, and villages which were spared of *any* shelling!

indiscriminate violence.[17] Hence, we consider his causal inference to have high internal validity. Moreover, his extensive efforts to secure the internal validity of this inference illustrate an important point which Lyall himself acknowledges: as we move away from randomized experiments to reliance on the assumption of as-if random assignment, much more work needs to be done to develop a persuasive inference in support of a particular account of what produced some outcome of interest.

However, as in Fair et al.'s study, there is some imprecision in how Lyall describes his findings, which is worth briefly noting. Like Fair et al., Lyall sometimes uses words such as 'can', suggesting an implicit awareness that what is being described in a proposition such as 'a state's use of indiscriminate violence incites insurgent attacks' is a causal propensity. For example, he describes his study as showing that indiscriminate violence '*can* suppress insurgent violence' (2009: 351, emphasis added; see also 2009: 332, 336, 337, 338). He also notes the likelihood that the effects of indiscriminate violence 'are not uniform but are *conditional on the nature of the insurgent organization itself*' (2009: 338, emphasis added), again indicating an implicit awareness that he is studying a causal propensity which will unfold fully only under the right conditions. Yet, he also talks of isolating '*the* causal effect of indiscriminate violence' (2009: 338, emphasis added) as if it has, or should have, a consistent effect across different settings. Moreover, Lyall, like Fair et al., sometimes uses the present tense without sufficient care where really the past tense is required. For example, when he reports his findings, he writes: '*Does* indiscriminate violence incite insurgent attacks? In brief, no' (2009: 349, emphasis added), whereas it would have been more precise to ask: '*Did* indiscriminate violence incite insurgent attacks in these villages in Chechnya in 2000–5? In brief, no'. Clearly, his formulation has a greater rhetorical flourish, but this is because it seems to point—*though illicitly*—towards a more 'general' finding.[18] This leads us to the issue of external validity.

External validity in Lyall's study

In the case of Fair et al.'s study, we noted that the authors demonstrated some awareness of the context-bound nature of their findings, but that they did not fully recognize that what can be inferred from an RCT is an abstract

[17] We say 'in effect' because he did not explicitly describe himself as ruling out competing explanations, but rather as considering 'confounding variables' (2009: 341, 350). As we explained in Chapter 5, however, the efforts pursued in quantitative analyses to control for confounders constitute one way in which competing explanations can be ruled out.

[18] In criticizing Lyall in this way, we are departing from Lebow's view (2014: 153; 2020: 13, 14, 224, 286, 310) that causal claims are *only* 'rhetorical'.

causal theory, stating a causal propensity, not a 'general' causal proposition. Lyall, for his part, was also clearly aware that what he developed was a concrete causal explanation *of the outcome observed in Chechnya from 2000 to 2005*: that shelled villages experienced, on average, lower levels of insurgent attacks. Yet, instead of explicitly recognizing that this indicates a *propensity* for indiscriminate violence to reduce insurgent attacks—a propensity the full unfolding of which is likely to depend on a range of contextual factors—he focused on the question of whether his findings were 'generalizable' (2009: 353).[19]

In this respect, Lyall's suggestion that his finding 'helps explain the otherwise puzzling persistence' of the use of indiscriminate violence 'among the world's militaries' (2009: 357) goes far beyond what can legitimately be claimed on the basis of his study alone; he did not present *any* evidence that what was going on in such cases was that military leaders understood that indiscriminate violence can, under the right conditions, have a suppressive effect, let alone that they employed it for this reason. This parallels Fair et al.'s unwarranted claim, noted earlier, that their experiment showed that there is a causal relationship between relative poverty, perceived violence, and level of support for violent organizations *in the developing world*, and that their findings explained the relationship between poverty and support for violent organizations found *in other studies*.

On balance, however, Lyall's attitude towards the question of whether his findings are generalizable is cautious and reasonable. He is quite right to acknowledge that 'there are … limits to the insights derived from a single case, and only replication in different subnational and cross-national contexts will reveal whether these findings are broadly generalizable' (2009: 355; see also 358). This accords with the account of external validity which we provided in Chapter 7. It is also reasonable for him to claim that his study raises 'new questions about the scope conditions under which indiscriminate violence suppresses insurgent violence' and about 'how state violence may have a variable effect across different types of insurgent organizations' (2009: 357–8). This suggests a clear awareness of the importance of contextual factors for how causal propensities unfold.

Lyall ends his paper by speculating on whether what he has uncovered in Chechnya is likely to be representative of a broader population of cases (2009: 355–7). There is nothing untoward in this, but his somewhat hastily arrived

[19] It is interesting that Lyall presents the question of whether his findings are generalizable as an 'inferential threat', as if there is something wrong with a causal inference if its results fail to generalize to other cases (2009: 353). As we explained in Chapter 7, this misrepresents the relationship between internal validity and external validity.

at view that the Chechnyan case is 'sufficiently representative' (2009: 357) can only be considered as a tentative judgement which needs to be defended more carefully. His study draws conclusions about the impact of a particular form of indiscriminate violence (apparently random shelling) on attacks by particular insurgent groups in a particular context—that is, villages within Chechnya between 2000 and 2005. As Lyall appears to acknowledge, changing the context in any of a number of ways might well influence whether a similar result would be observed in another setting. Any extrapolation to another setting must therefore be tentative at best.

II. Examples of causal inquiry in IR drawing on causal-process observations

In Section I, we considered two examples of causal inquiry in which inferences were generated through the study of so-called 'data-set observations', aided by the assumption of randomness or as-if randomness in the way the data has been generated. In this section, we will consider two examples of causal inquiry in which inferences are generated through the study of so-called 'causal-process observations'—of evidence about the process which brought about a specific event of interest. As we explained in Chapter 7, all causal inferences rely, explicitly or implicitly, on abductive reasoning, and any method which can be drawn upon to help rule out plausible competing explanatory statements can therefore contribute to the generation of such inferences. Our aim in this section is to demonstrate how evidence about the history of specific events can underpin persuasive causal reasoning and to identify some of the challenges this raises.

Use of 'causal process observations' in causal inquiry is most commonly associated with applications of the method known as 'process tracing' (see George 1979; George and Bennett 2005: 205–32; Collier 2011; Beach and Pedersen 2013; Bennett and Checkel 2015a) and we will consider one study by a scholar who explicitly describes herself as employing this method: Stacie Goddard's (2015) investigation into what caused the sudden change in British policy towards Germany in 1938–9. However, we will also consider a study by someone who views himself as a historian, rather than a process-tracer: Marc Trachtenberg's (1991a) analysis of the origins of the First World War.

In evaluating these two studies, our principal focus will be on how evidence about the history of a specific event can be drawn upon to generate causal inferences with high internal validity. However, we will also touch on questions of external validity, for while scholars who draw on causal-process

observations are typically seeking to develop reliable explanations of the events whose history they are studying, they are often also aiming to generate broader lessons.

By juxtaposing a study by a self-identified process-tracer with a study by a historian, we also wish to draw out the close relationship between these two modes of inquiry and to emphasize the potential value of 'historical analysis' in causal inquiry. Perhaps because of fears about the inferiority of methods which draw on causal-process observations as compared with methods which draw on data-set observations, 'process tracing' is often described in quite scientific-sounding language. For example, Bennett and Checkel (2015b: 7) define it as 'the analysis of evidence on processes, sequences, and conjunctures of events within a case for the purposes of either developing or testing hypotheses about causal mechanisms that might causally explain the case' (see also George and Bennett 2005: 206; Beach and Pedersen 2013).[20]

Most historians, we surmise, would not use this kind of language to describe what they are doing when they draw on evidence about the history of some event of interest to generate an explanation of how it was brought about.[21] Moreover, many process tracers, being methodologically self-conscious, may wish to avoid being associated with 'history', not least because of the widespread, though in our view misleading, idea that 'historical interpretation' is antithetical to rigorous 'social science'.[22] In our view, however, there is very little practical difference between 'process-tracing' and 'history', at least when the purpose of the latter is to develop an account of what caused a specific event of interest; what matters, for the purpose of causal inquiry, is not what labels are used, but whether evidence is presented which supports a reliable causal inference.

[20] It is important to bear in mind that process tracers' methodological self-consciousness may not be unambiguously virtuous: it may contain misleading ideas about the logic of causal inquiry. For example, George and Bennett (2005: 225–30) are right to consider the relationship between process tracing and historical explanation, but their account of historical explanation as involving the discovery of (micro) covering laws is misleading: as we pointed out in Chapter 6, the covering-law model of explanation has long since been discredited. Beach and Pedersen are right to stress that process tracing does not merely generate case-specific explanations, but also provides support for causal theories, yet their suggestion that the kind of theory for which support is provided is or should be 'generalizable' (2013: 3, 9, 11, 12, 16) is misleading: as we have repeatedly pointed out, causal theories are abstract propensity statements, not generalizations.

[21] Of course, many historians have aims other than to explain the occurrence of specific events in this way. Our interest is in how 'historical analysis' is used when the aim *is* to develop a causal explanation of an event of interest.

[22] For example, Bennett and Checkel (2015b: 9) insist that process tracing 'is not simply glorified historiography'. On the relationship between 'history' and 'social science' (or 'political science') in IR see, for example, Elman and Elman (2001a).

Using process tracing to explain changes in British policy towards Germany in 1938–9

As one might expect from a scholar drawing on causal process observations, Goddard (2015) engages in a detailed empirical fashion with a particular case of interest: the sudden change in British policy towards Germany in the period 1938–9. However, as is typical of someone who is self-consciously applying 'process tracing' as a method (see 2015: 111), she explains the value of this work partly in terms of its theoretical payoff: she argues that the case she examines provides support for what she calls 'legitimation theory'. She sets the scene by noting the following shift in British policy, which occurred in a relatively short interval between November 1938 and March 1939:

> Whereas prior to October 1938 Britain had done its best to avoid entangling alliances, after Munich it strengthened its commitment to defend France and even sought out an alliance with the Soviet Union. And while before Munich the British public seemed reticent if not downright hostile to a grand strategy of confrontation, after Munich the public declared itself determined to oppose any further German expansion. (2015: 96–7)

So, what caused this radical shift from appeasement to confrontation? To paraphrase, Goddard's answer is that Hitler's pre-Munich legitimation rhetoric, which invoked the internationally accepted norms of equality and national self-determination, resonated with the British government and public, and helped persuade it to adopt a policy of appeasement, but the sudden post-Munich shift in Hitler's rhetoric to one based on German might and needs was dissonant and caused Britain to turn its policy to confrontation. In short, the change in British policy was caused, among other things, by a change in the rhetoric with which Hitler justified German actions; Hitler's abandonment of his previous 'legitimation rhetoric' *caused* the change in British policy.

In our view, this claim is broadly persuasive: Goddard largely succeeds in developing a persuasive case in support of the following explanatory statement, which we will label (A): 'the change in German rhetoric caused the change in British policy'. However, in order to see *how* Goddard establishes this claim it is necessary to engage closely with her argumentation and to appreciate how her twin aims of explaining this specific outcome and testing legitimation theory are intertwined.

The internal validity of Goddard's causal inference

How persuasive is the reasoning which supports Goddard's favoured causal explanation, (A), of the sudden shift in Britain's German policy? Goddard begins, as process tracers in IR often do, by identifying a series of 'alternative' explanations which she will reject in favour of her preferred explanation. In her case, she considers and rejects three prominent explanations of the shift in Britain's German policy that she finds in the existing literature (2015: 101). These are:

> (B) appeasement had been improvised as a necessary means to buy time until the strategic balance became more advantageous to Britain, at which point Britain switched to confrontation;
> (C) new information about Hitler's intentions supported a switch to confrontation; and
> (D) changes in British domestic politics and public opinion supported a switch to confrontation.

Of these, Goddard first rejects (B) and (C) as contradicted by the order of events: in the crucial five-month period in which the shift in the German rhetoric was followed by a u-turn in Britain's German policy, there simply was no shift in the balance of power, nor was there any new information that Britain gained regarding Hitler's intentions (2015: 102, 104); and so neither (B) nor (C) could possibly be sustained.[23] As for (D), Goddard contends that 'it is not enough … to say that domestic politics now supported a strategy of confrontation': it is necessary to explain 'why the public, so prone to appeasement before Munich, suddenly converged around confrontation' (2015: 106).

By rejecting (B)–(D), Goddard prepares the ground to make the case for (A). It is important to appreciate, however, that neither (B), nor (C), nor (D) *competes with*—is inconsistent with—(A). For example, there is no prima facie reason why a shift in the strategic balance (B) *and* a change in Germany's legitimation rhetoric (A) could not *both* have contributed to the change in British strategy. To express the point more formally: it is conceivable that Britain shifted its German policy because of the factors referred to in (B) *and* (A), (C) *and* (A), *or* (B), (C), *and* (A). As for (D), to say of an explanation that it 'is not enough', as Goddard does, is to point to its *pragmatic* inadequacy: it is to say that if an audience were presented with (D) as an explanation of

[23] Goddard also points out that the British supporters of appeasement did not treat it *just* as a delaying tactic to buy time but believed that it was an appropriate policy to adopt, given Germany's apparent willingness to compromise (2015: 102). This, for her, is another reason for rejecting (B).

the shift in Britain's German policy, this would fail to satisfy their curiosity. Moreover, (D), like (B) and (C), is not incompatible with (A); indeed supplementing (D) with (A) may, as Goddard indicates, be what is required to make (D) more satisfactory.

The crucial point is that in order to make the case for (A) it is not *necessary* to reject (B), (C), or (D). Moreover, rejecting them is not *sufficient* to make the case for (A), either. To make a compelling case for (A), Goddard will have to show that it is implausible to hold that 'the shift in Britain's German policy was *wholly uninfluenced* by the shift in Hitler's legitimation rhetoric'. In order to generate a persuasive causal inference in favour of (A), it is this contrary thesis—call it (E)—that must be rejected.[24]

Goddard does not spell this out. Yet, having rejected (B)–(D), she clearly did *not* suppose that this was sufficient to make the case for (A). In fact, she went to considerable further lengths to provide empirical evidence in support of (A) (see 2015: 111–29). This indicates an implicit appreciation that merely arguing against some prominent alternative explanations (as distinct from ruling out genuinely competing explanations), is not sufficient to generate a persuasive causal inference.[25] However, Goddard did not make explicit how and why the further evidence she presented was intended to facilitate a persuasive causal inference. It is worth elaborating on this. In order to do so, let us return to a point we have already noted: that Goddard's aims are both to *explain* the change in British policy and also to *test* legitimation theory.

Goddard presents her favoured explanation of the change in British policy as involving an application of 'legitimation theory'. As she makes clear in her later book (2018), this theory, expressed in broad terms, concerns the domestic and international causes and consequences of the language a state uses to legitimate its policies. However, the specific version of this theory with which Goddard is concerned in the article we are considering here (2015) is more tightly focused on power transitions in interstate relations: it asserts that the rhetoric a rising power uses to legitimate its revisionist foreign policy influences the policy which status quo states adopt in response.[26]

[24] As we have consistently reiterated, the demand we are making, here, of a process tracing study, is no different from the demand we would make of a study using RCTs, natural experiments, historical analysis, or any other method.

[25] By contrast, some process tracers in IR appear to fall into the trap of supposing that rejecting one or two prominent alternative explanations is sufficient to make the case for a favoured explanation (see, for example, Bennett 2010).

[26] In her book (2018: 16), Goddard elaborates on this a little further: 'How a rising power legitimates its claims—how it justifies its demands to an international audience—significantly shapes how great powers understand its intentions, and thus affects whether great powers will accommodate or confront its increasing might'. However, the versions of 'legitimation theory' which Goddard explores in this book appear to vary somewhat across the cases she considers. We set this issue to one side and focus on the argument we find in her article (2015).

It is important to appreciate straightaway that Goddard does not arrive at her favoured explanation, (A), by *applying* legitimation theory in the fashion we described in Section III of Chapter 5. In order for her to have done so, it would have been necessary not only for the specific version of legitimation theory being applied already to be well supported, establishing that there is a propensity for revisionist states' legitimation rhetoric to shape status quo states' responses, but also for Goddard already to have established that this propensity unfolds fully under conditions of the kind which prevailed in the particular case she is examining—that is, British policy towards Germany in 1938–9. Goddard rightly makes no claim to this effect. Rather than *inferring* that because this propensity is known to have unfolded fully in some other setting, it must also have unfolded fully in the British case, she draws on detailed historical evidence to assess *whether or not* it unfolded fully in this case. In order to show persuasively that it did, she needed to generate a reliable causal inference to the effect that the change in German rhetoric *caused* the change in British policy. This required engaging in 'archival research' (Goddard 2015: 112), much as a historian would, in order to rule out competing explanations for this change.

As we shall see, this is just what Goddard does. However, this is somewhat obscured by the fact that, having rejected explanations (B)–(D), her next step is to explicate legitimation theory in more detail (2015: 107–11) by identifying the 'causal mechanisms' (2015: 109) through which a revisionist state's rhetoric may influence a status quo state's actions, viz. 'through signalling constraints, through rhetorical coercion, and through demands on ontological security' (2005: 108). She then frames her historical research as a 'test' (2015: 111) of legitimation theory. This gives the appearance that what she is doing is very different from what a historian would do if challenged to explain the change in British policy towards Germany in 1938–9, but this appearance, we believe, is deceptive.

Goddard argues that in the British case all three of the 'mechanisms' she associates with legitimation theory operated (see 2015: 111–29). In brief, she contends that:

(i) Hitler's pre-Munich legitimation rhetoric invoked the internationally accepted rules and norms of equality and national self-determination, as embodied in the Versailles Treaty and the League of Nations and accepted Europe-wide; this indicated Germany's self-restraint in its pursuit of revisionism, and thus Germany's aims were accepted by status quo powers as being reasonable, warranting accommodation;

(ii) this had the effect of imposing a rhetorical trap on domestic opponents of appeasement in Britain, for anti-appeasers faced the charge of hypocrisy in failing to apply to Germany the norms and rules that they endorsed elsewhere; and
(iii) German legitimation rhetoric was consonant with Britain's identity, or understanding of itself, as standing for liberal values such as equality and national self-determination, reinforcing the case for accommodation.

Conversely, when, after Munich, Hitler radically altered his rhetoric and claimed that Germany had a right to expand based on its might and needs, rather than on internationally accepted norms and rules, Britain responded immediately by making a u-turn from appeasement to confrontation. This was because Britain now saw Germany as lacking self-restraint, because the anti-appeasers in Britain were now freed from the possible accusation of hypocrisy, and because Britain now perceived an existential threat to its identity.

As we have indicated, we are broadly persuaded by Goddard's line of argument. However, it is important to be clear about *why* it is persuasive. It is not persuasive simply because Goddard has presented evidence which is *consistent* with legitimation theory and with the mechanisms which putatively underpin it: showing that her favoured explanation, (A), is consistent with the evidence is necessary but not sufficient. Goddard's line of argument is persuasive because, or to the extent that, she succeeds in showing that plausible competing explanations can be ruled out—that it is not plausible to believe (E): that the shift in British policy was *wholly uninfluenced* by the shift in Hitler's legitimation rhetoric. In our view, Goddard does succeed reasonably well in showing this: this is what enables her persuasively to claim *both* that the change in British policy was caused by the change in German rhetoric *and* that this case is 'explained' by legitimation theory in the sense that the propensity described by this theory—the propensity for revisionist states' legitimation rhetoric to shape status quo states' responses—unfolded fully, contributing to the change in British policy.

In our view, the evidence Goddard presents broadly succeeds in showing: (a) that Germany's expression of restraint within the League of Nations framework was one of the reasons why the key actors in Britain felt they could and should appease Germany (2015: 118, 119); (b) that anti-appeasers in Britain felt constrained and 'trapped' by their own rhetoric with which Hitler's legitimation strategy struck chords (2015: 122–3); (c) that the norms

and values in terms of which Germany initially formulated its legitimation rhetoric resonated with Britain's political self-identity as a liberal state, which thereby contributed to its pursuit of appeasement (2015: 125–6); and (d) that the shift in German rhetoric to one based on the assertion of German might and needs made it necessary for the British appeasers to change their stance (2015: 120–1) and made it possible for anti-appeasers to justify Britain's adoption of a policy of confrontation (2015: 124–5) without appearing hypocritical. Goddard also claims (e) that Britain felt an 'existential threat', or a threat against its very identity, from Hitler's new rhetoric, and provides some grounds for thinking that this may have provided an important incentive for Britain to abandon appeasement (2015: 118, 128–9); and, citing some recent historical literature, she reasons (f) that it was Britain's deep commitment to the institution of the Versailles Treaty and the League of Nations system that made the German rhetoric exert a strong impact on Britain (2015: 119–20, 123–4, 126–7).

In the apparent absence of any contrary evidence, the evidence Goddard provides appears to support her interpretation of the British case quite strongly.[27] When combined with her brief observation that, in the crucial five-month period, 'Hitler really did not do much of anything' to the status quo powers other than to radically alter his legitimation rhetoric (2015: 105), it becomes difficult to believe (E): that the shift in Britain's German strategy was *wholly uninfluenced* by the shift in Hitler's legitimation rhetoric.[28] Since (E) appears implausible, the contrary claim (A) appears reasonably persuasive. Our judgement, therefore, is that Goddard has, in effect, employed abductive reasoning to generate a persuasive causal inference in support of (A), even though she did not frame her analysis in this way: she did not spell out the contrary claim, (E), which would need to be rejected in order to make the case for (A). However, viewing her argument as an argument against (E) serves, in our view, to clarify *why* it is persuasive.

The issue of external validity in Goddard's study

Unlike Fair et al. and Lyall, Goddard does not seem especially concerned with how far her account of the British case can be 'generalized'. She does, however, seek to learn lessons: in particular, she warns, on the basis of her analysis of the British case, that dismissing legitimating rhetoric as 'cheap talk' would be 'foolish' (2015: 129–30). She argues that although '[t]alk may be cheap … it

[27] We say '[i]n the *apparent* absence of any contrary evidence' because Goddard did not *explicitly* state that, despite her concerted effort, no contrary evidence could be found in her sources.

[28] Goddard points out: Hitler 'did not change his demands, he did not accelerate rearmament, he did not threaten to invade. What did change was not Germany's actions, but its talk' (2015: 105).

may also reveal much about a state's willingness to work within an existing order' (2015: 130).

There is nothing intrinsically wrong with this suggestion, which implies that the status quo states, when faced with rising powers, should keep appeasement as a possible foreign policy option. Yet, it is curious that *this* is the lesson that Goddard should draw from her study, for this lesson concerns *how status quo states should respond to revisionist states' legitimation rhetoric*, whereas her study of why Britain's policy toward Germany changed concerned the *effect of a revisionist state's change in rhetoric*.[29] What her study showed was that a revisionist state's rhetoric *can*—it has a propensity to—influence how status quo powers respond. It is a further question how status quo powers *should respond*. The more straightforward, though perhaps less palatable, lesson to draw *from her study of why British policy towards Germany changed in 1938–9* is that it would be worthwhile for rising powers with revisionist goals to consider whether the situation in which they find themselves is relevantly similar to the situation in which Hitler's Germany found itself in relation to the UK, and hence whether they, like Hitler's Germany, may be able to achieve at least part of their revisionist goals by seeking to legitimate these goals rhetorically, as Hitler's Germany did prior to Munich.

Finally, it is worth noting Goddard's judicious use of the word 'may' when she says, as noted above, that '[t]alk may be cheap, but it may also reveal much about a state's willingness to work within an existing order' (2015: 130). This indicates a clear awareness, on her part, that context matters for how causal propensities unfold. In showing, persuasively, that Germany's rhetoric shaped Britain's actions and that the change in Germany's rhetoric post-Munich contributed to the change in British policy towards Germany, she has also provided support for legitimation theory. Yet, in thinking about how legitimation theory may apply in *other settings*, it is crucial to recognize the influence that contextual factors have on whether known causal propensities will unfold fully in any specific setting. It is not part of Goddard's purpose, in exploring the British case, to address the question 'in what *other* settings will this propensity also unfold fully?' She does, however, appear to recognize the importance of due caution in speculating about such matters.

[29] In the conclusion to her book, Goddard also focuses on this question of how status quo powers *should respond* to changes in revisionist states' rhetoric, but this time in relation to US policy towards China: she considers whether the US should respond to China's increasingly aggressive rhetoric by shifting from a policy of 'hedging' to one of 'containment and confrontation' (2018: 190–4).

Using historical analysis to explain the origins of the First World War

Finally, we turn to an example of a historian using causal-process observations: Marc Trachtenberg's effort (1991a) to eliminate some commonly held misunderstandings about the origins of the First World War.[30] In the course of his study, he pointed towards a more persuasive explanation of how the war was brought about; he also undermined an influential theory of 'inadvertent war' which, he claimed, was almost entirely based on an erroneous understanding of how the First World War came about (see 1991a: 47, 95, 97–9).

Trachtenberg's approach is distinctive of a professional historian.[31] He does not hesitate to go into minute details of chronology and, unlike in the two examples we considered in Section I, his account of what his analysis reveals is consistently phrased in the past tense. However, he rejects the idea that 'the historian, obsessed with detail and factual accuracy', is guilty of 'mindless empiricism' (1991b: 261). In his view, 'the causal structure of the past is reflected in its chronological structure', making it essential to study 'in detail exactly what happened and when' (1991b: 262; see also 2006: 141). He claims, in short, that historians often (though not always) seek to reveal how events in the past were *causally* linked, even if they do not explicitly use causal or social-scientific-sounding language (see Trachtenberg 2006: 185–90; see also Dray 1978; Ringer 1989; Schroeder 1997; Carr 2002; Gaddis 2002; Suganami 2008; Humphreys 2011; Lawson 2012).

How Trachtenberg's analysis unfolds

To appreciate how Trachtenberg's analysis can be understood as developing causal knowledge, it is helpful to begin by considering the 'spectrum of possible interpretations' of the origins of World War I which he outlines (1991a: 55–6):

(1) is the extreme version of the well-known and controversial Fritz Fischer thesis that Germany 'set out deliberately to provoke a war with

[30] Parts of this essay originally appeared in Trachtenberg (1990–1).

[31] Trachtenberg is now Professor of Political Science, but he has a History PhD and spent nearly three decades as Professor of History (see http://www.sscnet.ucla.edu/polisci/faculty/trachtenberg/cv/cv.html; accessed 15 Aug. 2023). The fact that he was able to switch from History to Political Science reinforces our argument about the intimate connection between social scientific process tracing and the kind of detailed analysis provided by a historian seeking to develop a persuasive causal explanation of a particular event. On his commitment to historical analysis, see Trachtenberg (2006).

Russia and France', hoping to defeat these powers before they became too strong (1991a: 55, see also 49).

(2) is a 'more moderate view' according to which 'the Germans and the Austrians were intent on crushing Serbia' and 'Russia, supported by France, was determined to prevent this', the war being a product of this irreconcilable political conflict and of political leaders pursuing these aims 'regardless of the consequences' (1991a: 55). On this interpretation, neither side sought deliberately to provoke a great war; rather, a continental war was caused by the recklessness of the political leadership on both sides—by their turning a blind eye to the likely consequences of their actions.

(3) focuses on what happened during the July crisis, stressing the 'role of miscalculation, misperception, and misunderstanding' and emphasizing 'the sheer dynamics of the crisis situation'. On this account, 'the different parties to the conflict … ended up taking positions that none of them would have taken if they had been able at the outset to see how things were going to develop' (1991a: 55–6). In other words, the key actors were acting not recklessly but carelessly—that is, not turning a blind eye, but rather failing adequately to recognize what the outcome of their choices might be.[32]

(4) is the 'inadvertent war' thesis. It is important to note that Trachtenberg understands the concept of 'inadvertent war' quite narrowly. On his account, this interpretation is *not*, like (3), concerned with 'miscalculation, misapprehension, and misjudgement', but rather with political leaders' lack, or loss, of control over the military (1991a: 48), a process in which 'the rigidity of the military plans, the heavy emphasis on offense, and the pressure to mobilize first' are said to have played a crucial role (1991a: 56).

Let us now turn to Trachtenberg's evaluation of these interpretations.

He starts by dismissing (1) as having 'little basis' (1991a: 55) in the evidence. As he points out, Fischer and his followers provide 'no real evidence … of a decision by the German government', following the assassination of Archduke Franz Ferdinand, 'to engineer a war with Russia and France' (1991a: 51; see also 52, 53). In fact, Trachtenberg argues, Fischer's limited attempt to make his case involves 'a simple twisting of the evidence' (1991a: 51). For example, Trachtenberg argues that Fischer wrongly interpreted the

[32] On 'recklessness', 'carelessness', and related ideas, see Suganami (2002b).

German Emperor Wilhelm II's infamous phrase 'now or never' as referring to 'a great war with Russia and France', whereas in fact it referred to a showdown between Austria-Hungary and Serbia (1991a: 51–2).

Having dismissed (1), Trachtenberg focuses his attention on (4). He considers evidence about the rigidity of military plans (1991a: 57–64), about the so-called 'cult of the offensive' (1991a: 64–72), and about mobilization plans (1991a: 72–95). Let us focus on the last of these.

Trachtenberg stresses that, on the afternoon of 29 July 1914, hearing news of Russia's partial mobilization against Austria, the German political leadership dramatically revised its previous hardline policy and began trying to contain the crisis by restraining Austria towards a negotiated settlement with Serbia (see especially 1991a: 83–7).[33] As it proved, this change of tack was too late to prevent Russia from upgrading its military readiness to general mobilization the following afternoon. On receiving information that Russian general mobilization was imminent, Bethmann Hollweg, the German Chancellor, abandoned his efforts to manage the crisis, instead resigning himself to 'a leap in the dark'; he rationalized the war which this made inevitable as 'preventive' in the sense that, for Germany, fighting Russia in 1914 would be less hazardous than fighting a few years hence. As Trachtenberg emphasizes, however, Bethmann made this choice *not* because he was under irresistible pressure from the military to pre-empt Russia, but because of the information he was receiving about Russian actions and because of his political judgement that the international situation was fast getting out of hand, beyond Germany's capacity to control, and that there was hence no longer anything to be gained from pressuring Austria to compromise (see 1991a: 87–92).

In short, Trachtenberg argues, Germany entered the war not because its political leadership lacked or lost control over its military, as the 'inadvertent war' view suggests, but rather because it lost control over the development of the crisis.

In the case of Russia, too, Trachtenberg argues, the 'inadvertent war' view cannot be sustained. Russian actions reflected its political judgement that it could not abandon Serbia, hence war with Austria was inevitable, and therefore war with Germany, too, would be hard to avoid. If so, it would be better not to get caught unprepared; it was therefore necessary to begin general mobilization as swiftly as was politically feasible; nevertheless, aware of the magnitude of this step, Russia first declared a partial mobilization against Austria, despite the fact that this was militarily disadvantageous; the

[33] See Jack Levy's criticisms of this interpretation and Trachtenberg's robust response in Levy, Christensen, and Trachtenberg (1991).

subsequent decision to order a general mobilization was taken only because the political leadership judged that war was unavoidable (1991a: 93–5). Russia's military moves were therefore based on *political* judgements: they did *not* result from Russian statesmen *capitulating* to pressure from the military, a view which Trachtenberg characterizes as being unable to 'withstand the basic test of chronology' (1991a: 94). Trachtenberg does concede, however, that the Russian political leadership acted with 'astonishing irrationality', plunging Russia 'into a venture that it knew was beyond Russia's strength' (1991a: 70).

In sum, Trachtenberg argues that interpretation (4), like (1), 'is not supported by the evidence' (1991a: 49). War did not break out in 1914 because statesmen were 'overpowered by the workings of the military system' (1991a: 48). There was 'no 'capitulation' to the generals; the military had in no real sense taken control of policy' (1991a: 96).

Trachtenberg stresses that, in conducting this analysis, his aim was not to arrive at a definitive 'interpretation of the coming of the First World War', but rather to 'test … against the evidence' some claims about the origins of the war which have often been accepted rather uncritically, especially the 'inadvertent war' thesis (1991a: 95). Nevertheless, he does reveal that his 'own view falls about midway between (2) and (3)' (1991a: 56). In line with interpretation (2), summarized above, he considers a crucial cause of the war to have lain in the irreconcilable political conflict in Europe but, contrary to (2), he does not endorse the idea that statesmen, especially in Germany, acted recklessly, cognizant of the likely consequences of their decisions but turning a blind eye (he is, however, more willing to characterize the Russian leadership as reckless); rather, he emphasizes, consistently with (3), that statesmen (who did not lose control to the military!) were unable to control the dynamics of crisis escalation.

Assessing Trachtenberg's implicit causal inference

Trachtenberg not only stops short of propounding an authoritative 'interpretation' of the origins of the First World War, but also frequently refers to interpretations 'of the coming of the First World War', or to claims about its 'origins', rather than to what 'caused' it (see, for example, 1991a: 95). In our view, however, his inquiry *is* a causal one. He does, at times, question whether particular factors were 'causes' of the war (see, for example, 1991a: 73–4) and his argument focuses on ruling out some prominent explanations of how the war was brought about. His study thereby demonstrates how historical analysis might be used to generate a more plausible *causal* explanation. Let us elaborate.

It is important to note that, of the interpretations Trachtenberg considers, (1) and (4) are clearly inconsistent with one another: as he explains, 'if World War I were in essence a war of German aggression', as the Fischer thesis holds, then 'one could hardly claim that it came about because a political process, which might otherwise have brought about a peaceful settlement, had been swamped by forces from within the military sphere', as the 'inadvertent war' thesis holds (1991a: 49). Moreover, (1) and (4) are each clearly inconsistent with (2) and (3), and with the midway point between (2) and (3) which Trachtenberg favours. On the one hand, if the Fischer thesis is correct, and the war was a deliberate war of aggression, knowingly provoked by Germany's political leaders, then it makes little sense to contend that the war was caused by their recklessness or carelessness, for they were neither reckless (turning a blind eye to the likely consequence of their hardline support for Austria—a continental war) nor careless (failing to give due care and attention to this possible consequence) but rather deliberately provoked the war. On the other hand, if politicians on all sides really were 'overpowered by the workings of the military system', then *their* degree of apparent recklessness or carelessness was incidental, for they were overpowered.

One might go further and contend that (2) and (3) are also inconsistent with one another: that a statesmen can be either reckless or careless, but not both. However, let us set that question to one side.[34] The crucial point is that in comparing (2) and (3) (or some position in between) with (1) and (4), he is comparing *competing explanatory statements*. Moreover, insofar as he rejects both (1) and (4) as inconsistent with the evidence, he may be interpreted as *ruling out* competing explanatory statements, thereby providing support for his preferred interpretation. In so doing, he goes some way towards generating a causal inference along the following lines: given the deep and apparently irreconcilable political conflict in Europe, the political leadership on all sides was unable to control the dynamics of conflict escalation during the July crisis; this *caused* the First World War.

We say that Trachtenberg goes *some way* towards generating this causal inference because he does not explicitly claim to do so and because he does not explicitly consider whether there may be *other* competing explanatory claims, in addition to (1) and (4), which would need to be ruled out in order for this inference to have high internal validity. This reflects his stated aim, which was *not* to arrive at a definitive interpretation of the origins of World War I, but rather to argue *against* some views—(1) and (4)—which are often

[34] Trachtenberg appears to favour a position *between* (2) and (3) in part because he attributes different degrees of culpability to different actors: as we have pointed out, he seems to regard the German leadership more as careless but the Russian leadership more as reckless.

uncritically accepted. That said, we believe that his analysis does provide a clear roadmap for how historical analysis can be used to generate a reliable causal inference—that is, by ruling out competing explanatory statements. It is important to note that this is, in essence, indistinguishable from the way in which historical evidence is used in process tracing, as illustrated in our discussion of Stacey Goddard's work in the previous section.[35]

Before we move on, there is one important additional observation which is worth making regarding the position Trachtenberg comes to endorse. Recall, from Chapter 6, that in order for a causal explanation of a specific event to be *satisfactory* it must not only be reliable, in the sense that it is better supported by the evidence than competing explanations, but also relevant, in the sense that it tells an intended audience what they want to know. Insofar as what Trachtenberg advocates is a position somewhere between (2) and (3), as against (1) and (4), his particular focus is not on longer-term causes of the war, but on the intentions of the key decision-makers in the months and days, even hours, before the war's outbreak. He is particularly interested in the question of whether they actively provoked war, were reckless or careless, or were overpowered by the working of the military system. As for the longer-term causes of the political crisis that emerged in July 1914, he emphasizes the deeply rooted political conflict in Europe as a significant background factor, though without much elaboration.[36] This decision to focus primarily on relatively proximate causes of the war was a pragmatic choice.

This highlights an important feature of any work which draws on causal-process observations to arrive at an explanation of a particular event of interest, whether this work is presented as 'process-tracing' or as 'historical analysis': there is always a preliminary choice to make about *which* causes of the event to focus on. What this kind of work can demonstrate is that one or more specific factors were or were not causes of the event of interest. As we pointed out in Chapter 6, it makes no sense to suppose that what is generated is a 'complete' explanation of this event—complete in the sense that *all* its causes have been identified. Hence, when a given explanation is presented as the conclusion of an analysis based on causal-process observations, it is

[35] We do not mean to advance a broad assertion that 'history' or 'historical analysis' is indistinguishable from 'process tracing'. Clearly, 'history' may consist in many things other than using historical evidence to rule out competing explanations of what brought about an event of interest. Our point is that insofar as historians draw on historical evidence to support particular causal claims, they draw on this evidence in essentially the same way that process tracers do.

[36] A partial exception to this narrow focus is his detailed examination and rejection of the claim that the 'cult of the offensive' constituted a crucial longer-term cause of the war (on this see, especially, Van Evera 1984).

always necessary to consider not only whether the inference which generated this explanation has high internal validity, but also whether the explanation that is generated provides the kind of information that is sought.

The external validity of Trachtenberg's findings

We have suggested that Trachtenberg goes some way towards generating a causal inference to the effect that, given the deep and apparently irreconcilable political conflict in Europe, the political leadership on all sides was unable to control the dynamics of conflict escalation during the July crisis; this *caused* the First World War. As with any explanatory statement, this implies a corresponding causal theory: roughly, that 'a deep political conflict causes the governments involved to lose control of the development of a crisis stemming from it, causing war'. Of course, this is a broad statement; how and to what extent such a propensity unfolds is likely to depend enormously on the circumstances. This may be one reason why Trachtenberg, unlike Goddard, does not draw attention to the theory for which his case-specific findings implicitly provide support.

Logically, this theory could be used to derive policy recommendations. Trachtenberg can be said to have shown, without explicitly stating it, that a failure to address an underlying political conflict causes war *under conditions of the kind encountered prior to the outbreak of the First World War*. However, those wanting to infer from Trachtenberg's study a prescription for how to avoid war in future would need to pay careful attention to whether the situation in which this prescription is to be applied is relevantly similar to the one analysed by Trachtenberg. Clearly, much more work would be needed to identify other settings in which the propensity towards war which is implicit in Trachtenberg's preferred interpretation of the origins of the First World War unfolds fully.

It is vital to recognize, however, that in this respect Trachtenberg, as a historian, is in no worse a position than *anyone* who wishes to derive a lesson from one particular study, *whatever method they may have used to generate and analyse the evidence*. It is emphatically *not* because Trachtenberg's work is one of 'history', and not 'social science', that a generally applicable policy prescription cannot be derived from it; historians and social scientists are, in this respect, in the same boat. As we have emphasized throughout, causal theories are abstract, not 'general', and, as we emphasized in Chapter 7, applying them in new contexts is a necessarily tentative endeavour. If social scientists are, however, typically *more interested* than historians in whether the propensities described by their theories also unfold in the same way in other settings,

this does not indicate any difference in the logic of causal inquiry which they follow, but only in their aims.

Interestingly, what Trachtenberg emphasizes is not the abstract theory that is logically implied by his preferred interpretation of the origins of the First World War, but rather the fact that his analysis *undermines* the popular theory of inadvertent war—that is, the idea that 'political leaders' lack or loss of control over the military causes wars'. He points out that this theory is 'rooted, to a quite extraordinary degree, in a specific interpretation of a single historical episode': the coming of the First World War (1991a: 47). Clearly, if a theory is accepted predominantly based on an analysis of a single case, then the theory will be seriously undermined if this analysis fails to stand up to scrutiny. By refuting the suggestion that the First World War was caused by political leaders' lack or loss of control over the military, Trachtenberg undermines the principal reason for accepting a theory of 'inadvertent war', thereby illustrating how historical analysis can potentially be used to undermine a causal theory, even one that is widely believed to apply quite generally.

More broadly, his study points to the need for careful consideration of the basis on which particular causal theories are accepted. As well as asking whether conclusions travel from one case to another, it is also important to appreciate how specific analyses of specific cases may underpin the acceptance of particular theories and to consider what it might take to undermine these analyses.

Conclusion

The four examples of high-quality causal inquiry which we have discussed in this chapter collectively demonstrate that a wide range of methods for generating and analysing empirical evidence can be used to produce causal knowledge in IR. They also collectively illustrate the key contention we advanced in Chapter 5: that there is a deep logic to causal inquiry which holds regardless of which method is favoured.

As each of the four examples reveals, generating causal inferences by drawing on empirical evidence and existing knowledge to rule out competing explanatory statements is absolutely central to causal inquiry. It is not an easy task, whatever method for generating and analysing empirical evidence is used. It is hard to generalize about whether methods drawing on data-set observations or on causal-process observations are likely to be more effective in developing reliable causal inferences: which is to be preferred is likely to

depend primarily on the kind of evidence that is available in any specific case. However, we believe that the pattern we noted in our discussion of Fair et al.'s and Lyall's studies is likely to hold more broadly: when drawing on data-set observations, inferences with high internal validity will generally be harder to generate as we move away from an experiment, such as the RCT implemented by Fair et al. (2018), towards a 'natural experiment', such as the one which Lyall (2009) describes, and finally towards studies in which even an assumption of as-if randomness in the generation of data cannot be supported. As we also noted, however, the circumstances under which it will be possible to employ an experiment, or even a natural experiment, are quite limited. To point this out is not to diminish the value of these approaches, but merely to reinforce what should already be obvious: in a discipline such as IR, causal inquiry is hard!

As we have emphasized repeatedly during our discussions in this chapter, it is crucial to recognize that causal inferences are tied to the specific settings in which the events being studied have occurred. Whatever method for generating and analysing empirical evidence is used, it is possible to arrive at causal inferences with high internal validity only when much is known about particular features of these settings, for only then can competing explanatory statements confidently be ruled out. This is obviously true of studies drawing on causal-process observations, which require detailed historical investigations, but it is equally true of studies drawing on data-set observations, as illustrated by the painstaking checks which Lyall (2009) undertook in order to assure himself that shelled and non-shelled villages in Chechnya during the period under study really were comparable on a range of relevant attributes.

It follows from the fact that causal inferences are tied to specific settings that advancing confident claims about external validity is difficult. Although an explanatory statement which is supported by a persuasive causal inference logically implies a corresponding causal theory, it is always a further question whether a propensity found to have unfolded fully in one setting will also unfold fully in another setting. In the examples we have discussed, especially those in Section I, this fact is implicitly recognized. If we were being critical, however, we would argue that both Fair et al. (2018) and Lyall (2009) are, on balance, a little too sanguine about such matters. It would, we believe, be beneficial for the discipline of IR as a whole if the difficulties involved in advancing claims about external validity were more explicitly acknowledged.

Above all, what our discussion of these four examples reveals is the need for more explicit accounts of the reasoning employed in developing causal

knowledge claims. In general terms, we believe that knowledge of how particular methods can be applied to generate causal inferences is reasonably advanced; this is demonstrated by the four studies we have considered, each of which we consider quite persuasive in its analysis. In our view, what is required is a more explicit and more detailed accounting of how, in specific cases, competing explanatory statements are ruled out, of the kind of knowledge this is said to generate, and of how confident we can be in these findings.

Our discussions suggest that in relation to methods drawing on data-set observations, there is a need for improved precision in the use of causal language, especially in how causal knowledge claims are presented. Of particular importance here is the need for consistency in use of the past tense in describing case-specific findings and for a more explicit acknowledgement that causal theories state propensities which unfold fully only under the right conditions. This will benefit discussions of both internal and external validity.

In relation to methods drawing on causal-process observations, our discussions suggest that there is a need for more explicit analysis of which alternative explanations for a given event in fact compete with one another and hence which explanatory statements need to be rejected in order to make a persuasive case for a particular explanation (and for the causal theory which this explanation implies). This applies both to those studies characterized as drawing on 'process tracing' and to instances of historical analysis which aim to generate an explanation of how a specific event or set of events was brought about.

Conclusion

This book emerged out of two core convictions.

First, we, the authors, are convinced that causal inquiry (properly understood!) has a crucial role to play in the study of world politics. We recognize, of course, that scholars of international relations ask a vast range of questions about a wide array of substantive topics. Answering these questions quite rightly involves the application of various modes of inquiry. We do *not* contend that causal inquiry is more important than other forms of inquiry, nor that causal questions are more important than other kinds of questions. Rather, our insistence that causal inquiry should be recognized as a central component of the discipline of IR reflects three underlying beliefs: first, that questions about why things happened, about how things came to be as they are, and about how things might be made different *are causal questions*; second, that whereas 'causal inquiry' is often construed quite narrowly in IR, there are, in fact, few substantive research programmes which do not raise these kinds of questions at least to some extent; and third, that such questions often have an urgent practical and moral significance which cannot be ignored.

Our second conviction, however, is that, in the discipline of IR, important aspects of causal inquiry are widely misunderstood. One dimension of this problem has been exposed by Milja Kurki (2006, 2008): she points out that both proponents and critics of causal inquiry in IR have tended to adopt rather narrow understandings of what causation is and of how it manifests empirically, and have hence also offered unduly restrictive accounts of the kinds of questions causal inquiry can answer (see also Patomäki and Wight 2000; Patomäki 2002; Jackson 2011; Lebow 2014; Humphreys 2017a). Yet, there is also a second dimension to this problem, which has received less attention: that the contents and implications of causal statements are often quite seriously misinterpreted. This has contributed to the development and widespread acceptance of skewed and misleading accounts of the logic of causal inquiry and of its methodological implications.

With these two convictions in mind, the overarching aim of this book has been to provide a fuller and philosophically more adequate account of

Causal Inquiry in International Relations. Adam R. C. Humphreys & Hidemi Suganami, Oxford University Press.
 DOI: 10.1093/oso/9780198823582.003.0010

causal inquiry than has so far been developed in IR. We have sought to demonstrate what is problematic about established ways of thinking, provided a new account of the logic of causal inquiry and of what it implies for a range of methodological questions, and endeavoured, thereby, to enhance collective understanding of the contribution that causal inquiry can make to empirically rich and critically aware scholarship about world politics.

By way of conclusion, we will do two things. First, we will briefly review the arguments we have developed in the preceding chapters. Second, we will consider what these arguments indicate about the place of causal inquiry within IR; we will focus, in particular, on how causal inquiry relates to normative, interpretive, critical, and historical inquiry.

Our argument, reviewed

In Part I of the book—'Metaphysics'—we examined a range of arguments which speak to the contention, prominently advanced by critical realists, among others, that in order to improve and expand the scope of causal inquiries in IR it is necessary to embrace causal and scientific realism. In broad terms, we rejected this contention: we did not accept that an embrace of causal and scientific realism is *necessary*.

We pointed out that arguments for causal and scientific realism in IR often trade on the idea that those conducting causal inquiries face a dichotomous choice: they must embrace either causal realism or causal idealism. In the terms of this dichotomy, causal idealism is associated with the thinking of David Hume, and with the so-called 'regularity view' of causation, whereas causal realism is most often associated with the work of Roy Bhaskar. We showed, however, that despite some common misunderstandings, Hume was not a causal idealist and did not endorse the versions of the 'regularity view' which are most often attributed to him. In fact, his arguments point towards a way of thinking about causation which is neither 'idealist' nor 'realist' but rather 'agnostic': he refrained from judging whether causation is (or is not) part of the world as it exists independently of human thought, for this was not something that he believed he was in a position to know. We also showed that Bhaskar's arguments for causal realism are unconvincing. Although causal realism constitutes a coherent basis on which to conduct causal inquiries, and although Bhaskar makes some valuable conceptual and methodological observations (which we carried forward into our own analysis of causal inquiry in Part II of this book), he does not succeed in showing

that causal powers are a basic part of the world existing independently of human thought.

These findings—that Hume was not in fact a causal idealist and that Bhaskar's arguments for causal realism are unconvincing—undermine the realist–idealist dichotomy which has often been invoked in recent philosophically informed discussions of causation in IR. In order to reinforce our account of what is wrong with this dichotomous way of thinking, we also considered the work of Bas van Fraassen, who is the most prominent critic of causal and scientific realism in contemporary philosophy of science. As we showed, Van Fraassen explicitly defends agnosticism in relation to causation. His arguments also provide strong reasons to believe that adopting an agnostic stance will not hinder substantive causal inquiries. This seriously undermines the idea that in order to improve and expand the scope of causal inquiries in IR, it is necessary to embrace causal and scientific realism. This is not only because it seems hard to imagine a convincing argument *for* causal realism, but also because there is a viable alternative.

Our key take-away from all this is that the emphasis on the importance of *metaphysical* questions, which has been a feature of much recent work on causation in IR, is a red herring. In our view, it will be more profitable to focus on the important *methodological* questions which arise whenever causal inquiries are undertaken: it is by correcting the many misunderstandings which have arisen in relation to these methodological questions that it is possible to make the most far-reaching progress in enhancing our collective understanding of how to conduct causal inquiries in IR.

In Part II of this book, we therefore turned our attention to the methodology (broadly understood) of causal inquiry. Our motivating belief was that in IR, and in the social science research methods literature, numerous aspects of causal inquiry, including its basic logic, are systematically misdescribed and misinterpreted and that the widespread failure to recognize these mistakes is due, in part, to the pervasive influence of the culture of generalization—the tendency to prioritize the production of 'general' knowledge over knowledge of 'particular' facts and events as a key goal of social scientific research.

We pointed out that, as it is applied to causal inquiry, this culture of generalization frequently points us in the wrong direction. We highlighted a number of significant examples of this:

- The tendency to label statements such as 'revolutions cause wars' as 'general', and to contrast these statements with supposedly 'singular' statements such as 'the revolution caused the war', appears to

suggest, misleadingly, that statements such as 'revolutions cause wars' are 'generalizations', whereas we showed that they in fact describe causal propensities.

- It is widely supposed that causal statements imply regularity statements describing general patterns expected to be observable in the world around us. As we pointed out, however, a statement such as 'revolutions cause wars' does *not* imply that revolutions should always or often produce wars, but only that a revolution will produce a war *if the conditions are right*.
- It is also widely supposed that the priority, in causal inquiry, should be to produce 'general' knowledge, which is then 'applied' in explanations of how 'particular' events of interest were brought about. As we explained, however, only statements about specific events, located in space and time, can be directly supported by empirical evidence. Consequently, *all* causal inquiry depends upon the provision of support for statements about how specific events were brought about.
- It is sometimes suggested that the reasoning which supports 'causal inferences' is analogous to inductive generalization. We argued, however, that the kind of reasoning which is, in practice, required to derive causal knowledge from the study of empirical evidence is 'abductive': it shows that a statement about how one or more specific events were brought about can confidently be accepted when competing statements can be ruled out.
- It is frequently suggested that once a reasonably persuasive causal inference has been arrived at, a crucial question to ask is how far it can be 'generalized'. We pointed out that this question is framed in a misleading way: what it is getting at is whether the causal relation which has been found to unfold in one situation will also unfold in other kinds of situations, but this can only be established through the systematic study of these various kinds of situations.

Our aim, in Part II of the book, was to provide an account of causal inquiry which is free of the influence of this culture of generalization. We sought, thereby, both to correct predominant misunderstandings and to provide a more secure basis on which to conduct, interpret, and evaluate substantive causal inquiries.

We did this in a progressive sequence of steps.

We started by examining, in some detail, the distinction between two basic kinds of causal statements: those which refer to specific events, located in space and time (we termed statements of this kind 'concrete'), and those which abstract from such events to state causal propensities (we termed

statements of this kind 'abstract'). We also explored the implications of these two kinds of statement. Although correctly understanding causal statements is important in its own right, we also had another reason for doing this: as we showed, the logic of causal inquiry emerges out of the relationship between these two kinds of statement. This is because only concrete statements—those which refer to specific events, located in space and time—can be directly supported by empirical evidence. The only way to provide support for an abstract statement such as 'revolutions cause wars' is to provide empirical support for the claim that some specific revolution caused a war, from which it can be inferred that 'revolutions cause wars'. Only then can this abstract statement be applied to generate further causal knowledge in the form of causal explanations, causally informed predictions, and policy prescriptions. We demonstrated, in short, that *all* causal knowledge depends upon the provision of empirical support for statements about how specific events were brought about. This fundamental feature of the logic of causal inquiry has been missed in existing discussions in IR.

Having laid out the logic of causal inquiry, we went on to consider how the account we provided might inform understanding of causal explanation and of the kind of causal reasoning required, in practice, to generate persuasive causal knowledge claims. In doing this, we corrected numerous misunderstandings and, we hope, provided a clearer way of thinking that can support more accurate analyses and discussions in future. In Chapter 8, we sought to illustrate how the approach we have developed can, in practice, inform and aid substantive causal inquiries by exploring four high-quality examples of causal inquiry in IR. We showed how our approach can be drawn upon to make sense of the often implicit reasoning on which these studies rely, and to identify how they might be further tightened up and improved. We also shed light on how it is that a range of very different methods for gathering and analysing empirical evidence can all be drawn upon in similar ways to generate causal knowledge.

The place of causal inquiry in IR

In bringing this book to a close, we wish to reflect on what our analysis of causal inquiry suggests about its place within the discipline of International Relations.

Although we doubt that embracing causal realism constitutes any kind of panacea, we agree with Kurki (2008) that causal inquiry, properly understood, is much more pervasive than is typically recognized. Like her, we also

reject the idea that causal inquiry is categorically distinct from critical and interpretive inquiry. In our view, many research endeavours in IR which may not explicitly be described by their authors as 'causal' are nonetheless properly understood as involving at least some elements of causal inquiry.

In saying this, we do not seek to diminish the value of non-causal inquiry. To give just one example, recent work on race and racism has posed challenging and important questions about the origins of IR as a discipline and about the focus and assumptions of some of its most widely cited theories (see, for example, Anievas, Manchanda, and Shilliam 2015; Vitalis 2015; Sabaratnam 2020). This work does not, for the most part, involve causal inquiry, but that does not, in our view, make it any the less valuable. Our point is not to insist that causal inquiry should be given greater priority, but rather to issue a plea: the common supposition that the aims and methods of causal inquiry place it intrinsically at odds with various other forms of inquiry should not be accepted uncritically. In order to understand the place of causal inquiry within IR it is necessary to start with an accurate account of its basic logic and then to consider, carefully, how it stands in relation to other significant and widely used approaches.

In what follows, we sketch the outlines of such an analysis by commenting briefly on how we see the relationship between causal inquiry, as understood in this book, and four prominent forms of inquiry in IR which are widely construed as 'non-causal': normative, interpretive, critical, and historical inquiry. We argue that causal inquiry is categorically distinct from normative inquiry, but that they are complementary and are often intertwined: causal inquiry will often, quite appropriately, be informed by normative ideas. More controversially, we contend that, as a category, causal inquiry *overlaps* with interpretive, critical, and historical inquiry: some instances of causal inquiry are also interpretive, critical, or historical and some instances of interpretive, critical, and historical inquiry involve causal inquiry. While this challenges the predominant view of how the discipline of IR is internally demarcated, we consider this view to be unduly restrictive. Challenging it has the potential to open up productive new lines of inquiry in IR and hence further illustrates the value of the analysis of causal inquiry which we have provided in this book.

Causal and normative inquiry

In our view, causal inquiry is wholly *complementary* to normative inquiry: although they are typically taught and practised separately from one another,

there is no tension, let alone contradiction, between them. In fact, they can and do inform each other in important ways. Let us elaborate.

Causal inquiry is often driven by suppositions about what it is pressing or important to know or do, suppositions which show up both in choices about which topics to explore in the first place and also in pragmatic judgements about which causes of events of interest to focus on when seeking to derive causal inferences. These suppositions are often rooted in, or informed by, normative considerations (see Suganami 2011). For example, when introducing their study of support for insurgent groups in Pakistan, Fair et al. (2018: 57, 59) noted that 'support for militancy … is a critical outcome in its own right', and explicitly highlighted the importance of understanding which approaches to reducing support for militancy are most likely to be effective. In other words, they justified their study in part by appealing to the normative value of research on their topic. In his study of the effect of indiscriminate violence on insurgent attacks, Lyall (2009: 357) also invoked normative considerations, though in a slightly different way: he warned against interpreting his findings as 'endorsing the use of random violence against civilians', emphasizing that '[s]uch actions are morally abhorrent and are rightly regarded as war crimes'. In other words, he argued that it is appropriate for normative considerations to limit which policy implications of causal findings should be pursued.

Causal and normative arguments are also sometimes more intimately intertwined. An interesting example is found in debates about the morality and effectiveness of strategic bombing. Writing on this topic in the 1990s, Robert Pape (1996) observed that the study of military coercion (including strategic bombing) had stagnated since the 1960s, in part because coercion 'usually involves hurting civilians' and is hence 'seen as morally repugnant'. He argued, however, that 'social scientists have a responsibility to advance knowledge on subjects relevant to policy', especially when policy decisions 'have great moral consequences'. In fact, he contended, 'the concern that studying coercion will provide strategic tools for international aggressors is misplaced … [for] states tend to over-estimate the prospects of coercion' (1996: 3). He found, in brief, that 'strategic bombing does not work' (1996: 314). Strategic bombing typically involves the imposition of costs (or 'punishment') on civilian populations, in an attempt to undermine morale and hence bring down enemy regimes, but, as he concluded, this is 'not only immoral, but futile' (1996: 135).[1]

[1] He further clarifies that when coercion does succeed, it does so through 'denying the opponent the ability to achieve its goals on the battlefield', but 'even denial does not always work … and strategic bombing is rarely the best way to achieve denial' (1996: 314).

For our purposes, what is interesting about this is how Pape's causal findings intersect with normative arguments. Pape observed that, in the absence of convincing *causal* inquiries into the effectiveness of strategic bombing, 'bad ideas and bad policies' may come to prevail. For example, in August 1990 'strategic bombing advocates were able to lobby successfully for air strikes on Iraqi electric power plants in part because the American defense community did not remember the failure of such campaigns in Korea and Vietnam' (1996: 3–4). In short, he expressed concern that if policymakers believe (rightly or wrongly) that strategic bombing campaigns are likely to be effective, normative considerations alone may not be sufficient to constrain their use (see also Thomas 2001); what will convince policymakers is evidence of their ineffectiveness.

Neither we, nor Pape, should be understood as suggesting that causal inquiry can *resolve* normative questions (see Pape 1996: 3). The issue Pape is raising is about what combination of arguments will persuade policymakers. He is suggesting that causal inquiry into the effectiveness of strategic bombing is normatively desirable, *even though the policy itself is morally objectionable*, because of the great moral consequences of choices in this area and because normative arguments, on their own, may not be sufficient to sway policymakers. The clear implication of this is that if causal inquiry shows strategic bombing to be ineffective, then this will make the development of convincing normative arguments against it less urgent (from a policy perspective), for the sheer ineffectiveness of strategic bombing should be sufficient to constrain policymakers from employing it, regardless of how strong the normative arguments are.[2]

What these examples illustrate is that causal inquiry will often take place within contexts that are shaped in significant ways by normative considerations. This seems to us to be entirely appropriate. There is no contradiction, or even tension, between causal and normative inquiry; they can and should operate hand in hand.

Causal and interpretive inquiry

Whereas we have argued that causal inquiry *complements* normative inquiry, we advance a stronger claim about the relationship between causal and interpretive inquiry. We contend that there is a significant *overlap* between these

[2] On the other hand, if causal inquiry were to demonstrate that strategic bombing *is* effective (under some conditions), this would make it *more* important to develop convincing normative arguments against it. This kind of scenario is exemplified by Lyall's (2009) study of the effects of indiscriminate violence in Chechnya and by his warning, noted above, that his findings should not be understood as endorsing use of indiscriminate violence, given the overwhelming moral case against it.

categories: sometimes, causal inquiry will also be interpretive, while some instances of what would typically be described in IR as 'interpretive' inquiry are also properly understood as examples of causal inquiry. This runs counter to Hollis and Smith's (1991a) influential claim that 'explanation' (which is centred on causal inquiry) and 'understanding' (that is, interpretive inquiry) are categorically distinct and cannot easily be combined (see also Bevir and Blakely 2018: 14). However, the account of causal inquiry which we have developed in this book suggests that Hollis and Smith are mistaken.

Some forms of interpretive inquiry are clearly non-causal. For example, an attempt to identify the meaning of a particular action in a particular situation does not involve causal inquiry; in the context of some broader research programme, efforts to identify meanings may *complement* efforts to identify causes, but the two enterprises are conceptually distinct. However, interpretive inquiry also includes the provision of reason-giving accounts of social actions—that is, accounts of social actions which identify the reasons why they were performed (see, for example, Hollis and Smith 1991a: 215; Bevir and Rhodes 2018: 3). Here, we argue, there is a clear *overlap* between the categories of causal and interpretive inquiry. A persuasive account of why an action was performed, pointing to the actor's reasons for performing it, will require a context-sensitive investigation of those reasons—it will require interpretive inquiry—but it will generate causal knowledge—knowledge of what brought about the action.[3]

We cannot provide a full analysis here of why Hollis and Smith fail to reach this apparently natural conclusion. However, a key part of the problem is that Hollis and Smith's account of causal inquiry is strongly shaped by the culture of generalization, giving rise to a misleading picture of the relationship between causal and interpretive inquiry.

As we noted in the Introduction to this book, Hollis and Smith associate causal inquiry with the attempt to develop generalizations which can serve as a reliable source of explanations. Thus depicted, the purpose of causal inquiry appears to be at odds with that of detailed investigation into the reasons held by particular actors in particular contexts. As we have emphasized, however, this depiction of causal inquiry is badly misleading. In fact, as we have shown, *all* causal knowledge relies on investigation into the causes of specific outcomes of interest, and what such investigations generate is knowledge of causal propensities, not generalizations. There is therefore no

[3] It might be objected that identifying the reasons why an action was performed *explains* that action, but that it is not a *causal* explanation. We believe this objection to be mistaken. Causation is a relation of production: to say that '*a*' caused '*b*' is to say that '*a*' contributed to the production, or bringing about, of '*b*'. To cite a reason as (part of) an explanation of an action is to identify that reason as having contributed to the bringing about of that action. If the reason did not contribute to the bringing about of the action, it would be odd to describe the reason as explaining the action.

reason to consider investigation into the reasons held by particular actors in particular contexts to be incompatible with causal inquiry.[4] Once it is recognized that causal statements concern the unfolding of propensities and that all causal inquiry involves identifying what brought about specific events of interest, the case for a categorical distinction between causal inquiry and context-sensitive interpretive inquiry collapses.

There is also a further reason to suspect that inquiry into the reasons held by individual actors in specific contexts might sometimes give rise to causal knowledge: that the logic of how empirical evidence can support a context-sensitive reason-giving account precisely matches the logic (which we laid out in Chapter 5) of how empirical evidence can support what we have termed 'explanatory statements'. In each case, the most powerful evidence is that which rules out competing accounts or statements.

Suppose it is asserted that some actor possessed a particular reason, *r*, for acting in a certain way and that this accounts for their action. Suppose, moreover, that the evidence supporting this assertion is the actor's own testimony. An obvious concern may be that this evidence is quite weak, for actors may have incentives to misrepresent their motives (see Hollis and Smith 1991a: 176, 185). This raises the following question: what kind of evidence would make the assertion that *r* motivated this action more persuasive? The answer, surely, is that it is evidence which serves to rule out competing accounts of what motivated the actor to act as they did (see Bevir and Rhodes 2018: 11).[5] In short, the generation of a persuasive reason-giving account is likely to rely on abductive reasoning: a persuasive reason-giving account is likely to involve a reasoned argument about why competing accounts of what motivated the action of interest can be ruled out.

We wish to draw two key points out of this brief discussion. The first is that, contrary to what Hollis and Smith (1991a) suggest, causal and interpretive inquiry can and do overlap. It follows from this that causal inquiry can potentially answer a wider range of questions than their analysis allows and can generate more nuanced and context-sensitive conclusions than they realize. Second, a key limitation of Hollis and Smith's way of thinking is their

[4] Moreover, such investigations are wholly compatible with the idea that actors' reasons themselves emerge out of their social contexts. To identify an actor's reasons as causing their behaviour is not to identify these reasons as constituting some kind of originary, uncaused cause (see Kurki and Suganami 2012). Rather, it leaves open the question of how these reasons may themselves have been shaped by the actor's social context.

[5] Of course, it may not be possible convincingly to eliminate all competing accounts, but, as we explained in Chapter 7, abductive reasoning cannot provide absolute certainty: the question is always how persuasive the case is for accepting a particular explanatory claim. This question arises in exactly the same way for reason-giving accounts as it does for other kinds of causal explanations.

failure to develop a plausible account of the logic of causal inquiry. An important aim of our book has been to show where the kinds of assumptions on which they rely go wrong and thereby to facilitate a fuller appreciation of what causal inquiry can contribute to IR.

Causal and critical inquiry

Like interpretive inquiry, critical inquiry is typically described as being in tension with causal inquiry. Certainly, a common construal of the so-called 'third debate' in IR (Lapid 1989) is that it situates causal inquiry as part of 'mainstream', 'positivist' IR and depicts this kind of approach as being directly antithetical to the aims of 'critical' inquiry (see, for example, Campbell 1998b; see also Kurki 2008; Kurki and Suganami 2012: 415–18).

So far as we can tell, this understanding of causal inquiry as being intrinsically at odds with the aims of critical theory is quite widely shared within IR. In our view, however, it is misguided. Just as there is a significant overlap between the categories of causal and interpretive inquiry, there is also a significant overlap between the categories of causal and critical inquiry: sometimes causal inquiry serves critical ends, and sometimes critical inquiry involves causal inquiry.[6] What prevents wider recognition of this is, once again, the simplistic and misleading picture of causal inquiry suggested by the culture of generalization. Once it is appreciated that causal statements concern the unfolding of propensities, and are not generalizations or regularity statements, and once it is appreciated that *all* causal knowledge rests on the study of specific events, located in space and time, the presumed opposition between causal and critical inquiry largely dissolves.

To get a more detailed sense of why we believe the conventional story about the relationship between critical and causal inquiry to be misleading, consider the distinction which Robert Cox (1981), drawing on Horkheimer, proposed between 'problem-solving' and 'critical' theory.

Cox starts by noting that '[s]ocial and political theory is history-bound at its origin, since it is always traceable to an historically-conditioned awareness of certain problems and issues, a problematic'. In other words, all theories

[6] An important example of how causal inquiry can serve critical ends is found in the notion of 'explanatory critique' (Patomäki 2002: 151). The basic idea is that emancipatory goals can be served by showing how predominant understandings are not only false but are generated by oppressive social structures (see Bhaskar 1998: 25–79, esp. 52–3; Collier 1998; Patomäki 2002: 143–63).

'have a perspective'—they 'derive from a position in … social and political time and space'. Next, he describes 'two distinct purposes' which social and political theory can serve. Problem-solving theory 'takes the world as it finds it, with the prevailing social and power relationships and the institutions into which they are organised, as the given framework for action'. Its aim 'is to make these relationships and institutions work smoothly'. By contrast, critical theory 'stands apart from the prevailing order of the world and asks how that order came about'; it 'does not take institutions and social and power relations for granted but calls them into question by concerning itself with their origins and how and whether they might be in the process of changing'. It thereby opens up the possibility of 'creating an alternative world' (1981: 128–9).

We agree with Cox that all social and political theory must emerge from a historically conditioned awareness of certain problems and issues—how could it be otherwise? Certainly, this is true of causal inquiry: it will typically be motivated, at least in part, by historically conditioned ideas about what it is desirable to find out. We also agree with Cox that social and political inquiry can potentially serve a variety of ends, ranging from understanding how particular aspects of the existing order function to exploring how that order is being (or might be) transformed. It is crucial to recognize, however, that this is equally true of *causal* inquiry: one can investigate the causal processes which serve to maintain particular aspects of the existing order, but one can equally investigate the causal processes by which the existing order is being (or might be) transformed. In our view, therefore, the different possible *aims* of inquiry to which Cox draws attention do not correspond to two fundamentally different, and competing, *forms* of inquiry. As with Hollis and Smith's analysis, we believe that the problem with Cox's analysis is his reliance on a simplistic and misguided picture of causal inquiry. This point is worth spelling out more fully.

According to Cox, the 'strength of the problem-solving approach' is that it reduces 'the statement of a particular problem to a limited number of variables which are amenable to relatively close and precise examination', making it possible to 'arrive at statements of laws or regularities which appear to have general validity but which imply, of course, the institutional and relational parameters assumed in the problem-solving approach'. Its weakness, however, is that it 'posits a continuing present (the permanence of the institutions and power relations which constitute its parameters)'. Cox characterizes this as 'a false premise, since the social and political order is not fixed but (at least in a long-range perspective) is changing' (1981: 129). To us, what is notable about this argument is that, like many others, Cox appears to have been

misled by the culture of generalization.[7] We have shown that although causal inquiry may *appear* to generate 'statements of laws or regularities which … have general validity', such an appearance is deceptive. As we have been at pains to point out, a claim such as 'revolution causes war' is not a regularity statement, but rather a propensity statement. To advance a causal theory is *not* to suppose that 'the future will always be like the past' (Cox 1981: 131), but rather to advance a contingent and contextually sensitive claim about how the world works *under certain conditions*—conditions which *may or may not* extend significantly back into the past or forwards into the future.

Furthermore, it appears clear that what Cox describes as 'critical' theory is, at least in part, oriented towards the generation of knowledge which is, in our view, properly understood as *causal* knowledge! Let us give just two brief examples. First, Cox describes critical theory as 'a guide to strategic action for bringing about an alternative order'. Yet, as we have pointed out, causation is a relation of 'bringing about' and it is causal knowledge which provides a guide to strategic action.[8] Second, Cox argues that his historical materialism differs from neorealism insofar as it 'focuses upon the production process as a critical element in the explanation of the particular historical form taken by a state/society complex' (1981: 134). What he is discussing here appears to be a *causal* explanation: an explanation of how and why a specific state/society complex came to have a particular form at a particular time (see also Kurki 2008: 138). This suggests that it would be a mistake to suppose that neorealism, as an archetypal problem-solving theory, involves causal inquiry, but that historical materialism, as an archetypal critical theory, does not, despite the deep and important differences between these two approaches.

Our aim here is not to criticize the substance of Cox's influential work. In our view, he is right to point out that neorealism largely ignores an important dimension of world politics: the (economic) production process. Our aim is to question the widespread supposition, for which his work is often cited as support, that causal inquiry necessarily stands in opposition to critical theory. Of course, we do not claim that *all* critical inquiry is also causal, nor even that all of what Cox advocates by way of critical inquiry is causal. Moreover, we accept that some examples of causal inquiries in IR do slip into the kinds

[7] Kurki (2008: 131) makes a similar point, but she argues that Cox is taken in by a 'Humean' understanding of causation. We showed, in Part I of this book, that 'Humeanism' is a somewhat misleading category. We also rejected any binary choice between 'Humeanism' and 'causal realism'.

[8] This is not to deny that there is an important difference between, on the one hand, developing a guide to strategic action for *bringing about alternative orders* (which Cox describes as the task of critical theory) and 'developing a guide to tactical action which, intended or unintended, *sustains the existing order*' (which Cox describes as the task of problem-solving theory) (1981: 130; emphasis added). We accept that these tasks differ, but we believe that they will both necessarily involve causal inquiry.

of implicit assumptions (of a 'fixed order' and a 'continuing present') which Cox rightly criticizes. Our point is one of principle: there is nothing in the logic of causal inquiry, properly understood, which would place it in *intrinsic* opposition to critical inquiry; in fact, it seems not unreasonable to hope that one of the benefits of critical inquiry might be that it drives those conducting causal inquiries to ask a wider range of questions—not just about how existing structures and practices function and can be maintained, but also about how they were brought about and how they might be transformed.[9] We are not suggesting, of course, that this is the *only* function of critical theory, nor even that it should be the principal function. Rather, our point is to emphasize that critical theory can serve as a productive stimulus for causal inquiries in much the same way that normative theorizing can.

That said, our analysis also suggests the need for a more critical perspective *on* causal inquiry than many critical theorists have provided. As Kurki (2008) has pointed out, critical theorists in IR have, perhaps ironically, tended to be quite *uncritical* of the characterizations of the causal 'other' against which they contrast their own positions (see also Kurki and Suganami 2012). Too often, critical theorists have been content to present a fixed and simplistic picture of causal inquiry, rather than exploring how this simplistic picture came to dominate. This has closed down the space which might potentially have been made available for more careful articulations of what is involved in causal inquiry and for more nuanced appreciation of the ends it might possibly serve. Remedying that deficiency has been one of the key aims of this book.

Causal and historical inquiry

We will round off our discussion by considering the relationship between causal and historical inquiry.

The relationship between 'theory' and 'history' in IR has been much debated (see, for example, Elman et al. 1997; Smith 1999; Elman and Elman 2001a; Elman et al. 2008; Lawson 2012).[10] In our view, however, these debates have failed to make plain quite how deeply causal inquiry

[9] Kurki (2008) advances a similar line of argument. However, we believe that she is wrong to contend that in order for causal inquiry in IR to be capable of addressing a broader range of questions it is necessary for those conducting causal inquiries to embrace causal realism.

[10] One recurrent ambiguity in these debates has concerned whether what is being discussed is the relationship between *two approaches to the study of international relations*—that is, theoretical and historical approaches in IR—or rather the relationship between *two professionalized disciplines*—that is, IR and (Diplomatic or International) History. This reflects a deeper contestation over the relationship between 'history' and 'science' and the place of historical judgement in the study of world politics. This issue was

depends upon knowledge of specific, past events. Moreover, the recent, and welcome, resurgence of 'Historical International Relations' has not been accompanied by a sustained reflection on the contribution that 'Historical IR' might make to *causal* enquiries. Indeed, the tendency has been in the opposite direction: as with those doing 'interpretive' and 'critical' work in IR, those doing 'historical' work have tended to distance themselves from the very idea of causal inquiry. For example, a recent *Handbook of Historical International Relations* (de Carvalho, Lopez, and Leira 2021) lacks even a single index entry for either 'cause' (or the like) or 'explanation' (or the like).

Before we present our own observations about the relationship between causal and historical inquiry in IR, it is important to acknowledge that 'historical inquiry' is a very broad category: it encompasses work with a range of aims, on a vast array of substantive topics, and informed by a wide range of ontological, epistemological, and methodological assumptions (see, for example, Munslow 2006; Hobson and Lawson 2008). It is clear, however, that at least *some* historical work in IR does aim, among other things, to explain how and why specific, past events came about.[11] In our view, such work clearly constitutes a form of *causal*, as well as *historical*, inquiry, as exemplified by Trachtenberg's analysis of the origins of World War I which we examined in Chapter 8. This reveals that the categories 'causal inquiry' and 'historical inquiry' overlap: some examples of causal inquiry in IR are also examples of historical inquiry, and vice versa. Yet, our analysis of the logic of causal inquiry in Chapter 5 suggests that it is also important to recognize a further point: that *all* causal knowledge consists in or stems from knowledge of how specific, past events were brought about—that is, on precisely the kind of knowledge which many historians seek to generate.

One important reason why this crucial point may have been missed is that the culture of generalization has profoundly shaped understandings, in IR, of what 'causal inquiry' and 'historical inquiry' consist in, thereby contributing to a misleading picture of the relationship between them. Let us give two brief examples.

raised, provocatively, by Hedley Bull (1966) in IR's so-called 'second debate', but it arguably also underpinned the 'first debate' (see Guilhot 2011), while the idea that causal inquiry is inherently *ahistorical* was also central to the 'third debate' (see, for example, Cox 1981).

[11] Of course, much historical work in IR is *not* concerned with explaining how specific events were brought about. This includes, but is not limited to, work on the history of political thought (see, for example, Keene 2005; Owens and Rietzler 2021), on the history of the discipline (see, for example, Schmidt 1998; Ashworth 2014; Vitalis 2015), and much of the new global history (see, for example, Hunt 2014; Getachew 2019).

First, David Lake (2011) draws a categorical distinction between two approaches to the generation of causal knowledge: 'nomological' and 'narrative'. He argues that the experimental and statistical methods which are predominant in mainstream IR support a 'nomological approach' to causal inquiry which 'builds generalizable relationships from covering laws'; by contrast, he argues that in a 'narrative approach' causal inquiry 'takes the form of "just-so" accounts' for which the criteria are 'largely subjective' (2011: 474–5).[12] One problem with this view is that it seriously caricatures historical inquiry. As we demonstrated in Chapter 8, historical inquiry does *not* rely on subjective 'just-so' accounts: rather, as we showed, historical inquiry can underpin abductive reasoning in support of persuasive causal inferences in just the same way that experimental and statistical methods can—that is, by helping to rule out competing accounts of how specific events were brought about. A second problem is that, contrary to Lake's supposition, when experimental and statistical methods are employed in the service of causal inquiry, they do *not* provide support for 'covering laws'; rather, what they can (potentially) generate is knowledge of causal propensities—of causal relationships which unfold fully only under the right conditions. Because of his debt to the culture of generalization, Lake misses this crucial point. He therefore also overlooks the important commonality in how historical, experimental, and statistical approaches generate causal knowledge.

Our second example is Jack Levy's (1997) exploration of the relationship between 'history' and 'political science' in the study of international relations (see also Levy 2001). In contrast to Lake, Levy rejects the notion that there is a fundamental distinction between 'theory-based' and 'narrative-based' explanations (1997: 27). He argues that there *is* an important difference between historians and political scientists, but that it relates to *what* they are 'trying to explain', and not to *how* they explain (1997: 25): typically, he contends, '[h]istorians describe, explain, and interpret individual events or a temporally bounded series of events, whereas political scientists generalize about the relationships between variables and construct lawlike statements about social behaviour' (1997: 22; see also 2001: 41). His key point is that although historians tend to be more interested than social scientists in explaining the

[12] Similar distinctions have also sometimes been drawn by historians (see, for example, Gaddis 2001: 312–13; Schroeder 2001: 408). This reflects a long-standing tradition of historians and philosophers of history rejecting the covering-law model of explanation, often by appealing to the idea that there is something distinctive about the historian's 'narrative' (see, for example, Mink 1966; Dray 1971). In our view, rather than accepting mainstream suppositions about the close association between causal inquiry and generalization and emphasizing how their approach is different, historians in IR should be 'challenging—and replacing' the misleading suppositions about causal inquiry that often appear in mainstream IR (Lawson 2018: 80).

occurrence of specific events, this does not make historical inquiry 'atheoretical': explanations of the occurrence of specific events are not necessarily any less 'theory-based' than explanations of more general patterns (1997: 25). However, Levy's view, like Lake's, is shaped in problematic ways by the culture of generalization.

One manifestation of this is that Levy makes the case for construing some historical explanations as 'theoretical' by arguing that although the 'primary goal of historians is to explain the particular … they often do so with resort to the general' (1997: 26–7). The problem with this is that causal theories are *not* generalizations, but rather abstract propensity statements. As we explained in Chapter 5, propensity statements *can* potentially be invoked to explain particular events. For example, if it can be established that there is a propensity for revolutions to produce wars and that this propensity unfolds fully under conditions of a certain kind, then if, under conditions of this kind, a further revolution is followed by a war, the theory 'revolutions cause wars' can be invoked to justify an assertion that this further revolution *caused* this war. It is important to recognize, however, that this does not illustrate the explanation of a 'particular' event by 'resort to the general'; rather, the event of interest (the war) is explained by invoking knowledge of the conditions under which a known causal propensity ('revolutions cause wars') unfolds fully.[13]

There is also a deeper problem. Levy's contention that historians explain particular events by 'resort to the general' appears to suggest that there is an inherent order to causal inquiry: first, it is necessary to develop 'general' theoretical knowledge; second, this theoretical knowledge is 'applied' to the event of interest in order to explain it. Even if this view is refined by clarifying that causal theories are propensity statements, not generalizations, the idea that causal inquiry must proceed in this way remains flawed.

As we pointed out in Chapter 5, there are *two* ways of generating a causal explanation. One way is to apply a theory—that is, knowledge of a causal propensity and of the conditions under which it unfolds fully. We have just provided an example of how this might be done. A second way is directly to present empirical evidence about the process which generated the event being explained—evidence which rules out competing explanations for its occurrence. It is crucial to recognize not only that there are these two different

[13] Some of the ways in which historians and philosophers of history have described the role of theoretical ideas in historical inquiry bear close comparison to our account of causal theories as abstract statements of causal propensities. Consider, for example, Scriven's discussion of 'normic statements' (1959), Ringer's characterization of historians as employing '"characteristically and for the most part" generalizations' and as seeking to explain deviations from (what is taken to be) the 'ordinary course of events' (1989: 156–7), and Trachtenberg's contention that the kind of 'theory' that historians require is 'a general sense for how things are supposed to work' (2006: vii).

ways in which a causal explanation can be generated, but also that the second way is, in a crucial sense, basic. It is possible to draw on a causal theory to explain a particular event only if that theory is well established, but a causal theory becomes well established only when it is inferred from a persuasively supported explanatory statement about what caused a specific event of the relevant kind, and an explanatory statement is persuasively supported only when empirical evidence serves to rule out competing explanations for the occurrence of the event of interest.

It is therefore crucial to recognize not only that there is an overlap between causal and historical inquiry—that some instances of one are also instances of the other—but also that *all* causal inquiry depends on precisely the kind of knowledge which historians routinely provide—that is, knowledge of how specific events were brought about.

We are not arguing here that all causal inquiry *involves* historical inquiry: that is clearly not the case. A social scientist administering an experiment is not conducting historical inquiry, nor, typically, will they be *aiming* to develop knowledge of the causes of specific, past events. However, a causal theory should be accepted, on the basis of an experiment, only if it can be shown that the treatment contributed causally to the observed outcome—that is, only if the experiment shows that the treatment helped to bring about the specific event of interest: the experimental outcome. A social scientist applying a causal theory to generate a prediction or policy prescription will not be engaged in historical inquiry either; unlike the experimenter, they will not even be engaged in *developing* knowledge of what brought about specific events. However, they will *depend* upon such knowledge, for the basis on which the theory they are applying will have been accepted is that the propensity it describes has been shown to have unfolded fully in one or more specific situations, contributing to the bringing about of one or more specific events.

The key implication of all this for the discipline of IR is that historical inquiry is logically on a par with experimental and statistical methods as a potential source of causal knowledge. *All* causal inquiry depends on knowledge of how specific events were brought about, and it is precisely this kind of knowledge which historians are in a good position to provide. It might even be thought that in a discipline such as IR, in which what is studied is principally the unfolding of causal relations in open systems, this kind of knowledge will often be more readily achieved through historical inquiry than through the application of experimental and statistical methods! If so, then IR really would merit the description 'a historical science' (cf. Wohlforth 1999; Lawson 2018)! That, however, is not an issue we wish to press further at this point. The crucial point to recognize is the dependence

of all causal knowledge on accounts of how specific, past events were brought about. This point has been almost entirely missed in existing discussions of the relationship between causal and historical inquiry in IR.

A brief final word

What this brief review indicates is that the analysis of causal inquiry which we have provided in this book is valuable not only because of what it reveals about how causal inquiries can and should be conducted, but also because it helps to develop a more accurate picture of how various approaches that are employed in the study of world politics relate to one another. It would, of course, be foolish to hope that our analysis has been wholly free of errors. We are optimistic, however, that our arguments will contribute to an enhanced appreciation both of the logic of causal inquiry and also of what it can potentially contribute to the study of world politics; we hope that this will open up productive lines of thought, and of empirical investigation, which might otherwise have remained closed.

APPENDIX

Causal Statements and Statements of Necessary and/or Sufficient Conditions

Over the years, it has been suggested that causal statements are equivalent to, or can be translated into, statements of sufficient conditions, of necessary conditions, or of necessary *and* sufficient conditions (Sosa 1975b: 1–2; see also Sosa 1975a). In addition, Mackie has advanced an influential argument to the effect that a causal statement will often describe neither a necessary nor a sufficient condition, but rather an 'INUS condition': it will identify a putative cause as 'an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result' (1965: 245; cf. 1980: 62).

For reasons we will shortly examine, attempts to analyse causal statements by reference to statements of necessary and/or sufficient conditions have fallen out of fashion among philosophers. In the social sciences, however, links are commonly drawn between statements of these two kinds (see, for example, Collier, Brady, and Seawright 2010: 145–52; Gerring 2012: 335–42). In particular, it is often suggested that there is a close association between case study methods and efforts to identify necessary and/or sufficient conditions (see, for example, George and Bennett 2005: 25–7; Gerring 2006: 716; Mahoney and Goertz 2006: 232; Levy 2008; Mahoney 2008; Beach and Pedersen 2016: 65–71).[1] Consequently, recent decades have seen a range of efforts to develop methodological tools for the analysis of necessary and/or sufficient conditions (see, for example, Braumoeller and Goertz 2000; Ragin 2000; Seawright 2002; Ragin 2014).[2]

One reason why it might seem attractive to explore the correspondence between causal statements and statements of necessary and/or sufficient conditions is that knowledge of such conditions, if it could be acquired, would be valuable for manipulating the world around us. If policymakers wish to *produce* a particular kind of outcome, then it will be helpful to know if there are any manipulable factors which are sufficient conditions for the occurrence of such an outcome. Conversely, if they wish to *prevent* a particular outcome, then it will be helpful to know if there are any manipulable factors which are necessary conditions for its occurrence (because if such a factor can be removed, then the undesirable outcome will not occur).

Acquiring knowledge of necessary and/or sufficient conditions is, however, quite difficult. Contrary to what Charles Ragin appears to suppose (see, for example, 2014: 99), such conditions cannot be discovered simply by inspecting outcomes within a given sample of cases. The assertion that an *A*-type event is necessary or sufficient for the occurrence of a *B*-type event implies that an event of one of these types has never occurred and will never occur without the other; it is an assertion which does not permit of exceptions. Consequently, in order to establish that *A* is *necessary* for *B* it will not be enough just to show that, *in a particular sample*, *B*-type events have not occurred without *A*-type events; what is required is a demonstration

[1] Mahoney, Kimball, and Koivu (2009) also associate efforts to identify necessary and/or sufficient conditions with historical inquiry.

[2] There has also been a burgeoning interest in the analysis and investigation of counterfactual claims linked to statements of necessary and/or sufficient conditions (see, for example, Goertz and Levy 2007; Levy 2015; Mahoney and Barrenechea 2019).

that *B*-type events *have never and will never* occur without *A*-type events. Similarly, in order to establish that *A* is sufficient for *B*, it will have to be shown that *A*-type events are *always* accompanied by *B*-type events. Even a controlled experiment cannot generate such conclusions: it can establish that *A*-type events lead to *B*-type events *under the experimental conditions*, but it cannot demonstrate either that *A*-type events *always* lead to *B*-type events, or that *B*-type events *only* occur following *A*-type events.[3]

We accept, of course, that it may *sometimes* be possible confidently to advance claims about necessary and/or sufficient conditions. Yet, such claims will often be either tautological or trivial. For example, being infected by the yellow fever virus is a necessary condition of contracting yellow fever, but that is because yellow fever is the name given to the disease caused by the yellow fever virus: this is tautological. Being alive is also a necessary condition for contracting yellow fever, but this is trivial! Consequently, when a researcher claims that empirical inquiry has revealed a necessary or sufficient condition, it is reasonable to wonder whether this claim may be partly rhetorical: whether the aim is just to highlight a particularly striking pattern of co-occurrence. For example, Steve Chan cites Nils Petter Gleditsch as arguing that shared democracy is 'a near-perfect sufficient condition for peace' (1997: 60). In our view, this is best understood as a rhetorical claim which highlights the unusual robustness, in world politics, of the empirical association between shared democracy and peace (see also Suganami 1996: 73–4). It is certainly crucial to appreciate the difference between (i) the weaker claim that, under the conditions which prevailed during a particular historical period, shared democracy has always or nearly always led to peace and (ii) the far stronger claim that shared democracy is a properly sufficient condition for peace, such that *all* instances of shared democracy will lead to peace, regardless of the prevailing conditions.

In addition to the practical challenge of identifying necessary and sufficient conditions, there are two deeper problems with efforts to analyse causal statements in relation to statements of necessary and/or sufficient conditions.

The first is that many causal statements do not imply a statement either of a necessary or of a sufficient condition, let alone of a necessary *and* sufficient condition! In a well-known example of this, Mackie (1965: 245) constructs a hypothetical scenario in which experts are tasked to investigate the causes of a house fire and 'conclude that it was caused by an electrical short-circuit'. This conclusion, he points out, does *not* imply that short circuits are *necessary* conditions for fires: the experts 'know perfectly well that … any one of a number of other things might … have set the house on fire'. Equally, it does *not* imply that short circuits are *sufficient* conditions of fires, 'for if the short-circuit had occurred, but there had been no inflammable material nearby, the fire would not have broken out'. What, then, are the experts saying? On Mackie's account, their finding that a short circuit caused the fire identifies the short circuit as a crucial component of a set of factors which, together, brought about the fire, while allowing that there are other ways in which such a fire could also have occurred. In his terminology: the experts identify the short circuit as an *INUS condition* of the fire. He presents a similar analysis of (what he terms) 'general' causal statements, such as 'the restriction of credit causes … unemployment' and 'the eating of sweets causes dental decay' (1965: 252). He argues that '[a] great deal of our ordinary causal knowledge is of this form'—that is, knowledge only of INUS conditions and not of necessary and/or sufficient conditions (1965: 252; see also 245).[4]

[3] This limitation of controlled experiments is not always appreciated (see, for example, Ragin 2014: 26–30).

[4] Although we endorse Mackie's central contention that many causal statements do not imply statements either of a necessary condition or of a sufficient condition, we believe his exposition, presented here, to be misleading in a subtle but noteworthy way. He argues that the short circuit was an INUS condition of 'the fire' (1965: 245). This appears to suggest that there are other ways in which *this specific fire*

The second problem with efforts to analyse causal statements in relation to statements of necessary and sufficient conditions is that there is a significant and systematic, though often overlooked, difference between these two kinds of statements. Consider what it means for *A* to be a necessary condition of *B*: it means that *B*-type events occur *only* when *A*-type events occur. Now consider what this further implies: that *whenever* there is a *B*-type event, there will also be an *A*-type event. In short, if *A* is a necessary condition of *B*, then *B* is a sufficient condition of *A*. Similarly, if *A* is a sufficient condition of *B*, then *B* is a necessary condition of *A*. If *A* is a necessary and sufficient condition of *B*, then *B* is a necessary and sufficient condition of *A*. If *A* is an INUS condition of *B*, then *B* is an INUS condition of *A*. This may take a bit of thinking through, but to anyone well versed in logic it will be very familiar, so we will not linger on it. The important point is that causal statements do *not* have this quality. Whereas the claim '*A* is a necessary and sufficient condition of *B*' implies '*B* is a necessary and sufficient condition of *A*', the claim '*A* causes *B*' does *not* imply '*B* causes *A*'! Causal statements have a 'direction' (Mackie 1965: 261) or '*asymmetry*' (Von Wright 1975: 97) which statements of necessary and conditions lack and which they appear incapable of capturing (see also Van Fraassen 1980; Brady 2008).

This has two very important implications. The first is that statements of causal relations cannot be *reduced* to statements of necessary and/or sufficient conditions; it is not possible to take a causal statement and render it, without remainder, as a statement of necessary and/or sufficient conditions (see Mackie 1965: 261; Taylor 1975: 42–3). The second is that causal relationships cannot be *inferred* from statements of necessary and/or sufficient conditions. Even if it can be discovered that *A* is a necessary and/or sufficient condition of *B* (and we have indicated, above, why this might be difficult), this discovery will not, *on its own*, justify a causal claim. Even if this pattern of co-occurrence is non-spurious, it will still be necessary to establish the direction of causation. Cartwright (2007: 34–5) makes the same point in relation to the discovery that *A* is an INUS condition of *B*: this information does not, *on its own*, justify the claim '*A* causes *B*', for the correlation may be spurious; even if there is a causal relationship between *A* and *B*, the direction of causation will still have to be established.

What, then, can be said about the relationship between causal statements and statements of necessary and/or sufficient conditions? We have argued that concrete causal statements imply abstract statements and that the abstract statement '*A* causes *B*' implies that an *A*-type event will produce a *B*-type event *if the conditions are right*. Consequently, although statements of necessary and/or sufficient conditions do not imply causal statements, causal statements do imply statements of necessary and/or sufficient conditions! More precisely: the statement '*A* causes *B*' implies that *A* is *at least* an INUS condition of *B*.

The reason why we employ the qualifier 'at least' is to indicate that '*A* causes *B*' is compatible with *A* being *merely* an INUS condition of *B* but *also* with *A* being a necessary, sufficient, or necessary and sufficient condition of *B*. Let us elaborate.

If there were *no* conditions under which an *A*-type event would produce a *B*-type event, then the statement '*A* causes *B*' would not be acceptable: the assertion '*A* causes *B*' implies that there are at least some conditions under which an *A*-type event will produce a *B*-type

could have come about. As Donald Davidson points out, however, something else 'could not have caused *this* fire' (1975: 84, emphasis original); something else could have caused *a* fire at this moment, in this location, but it could not have caused *this* fire—this fire was caused by the short circuit! Hence, rather than identifying 'the short circuit' as an INUS condition of 'the fire', it seems to us more accurate to say that the statement 'the short circuit caused the fire' identifies *short circuits* as INUS conditions of *fires* (see Kim 1975: 58). In his later work, Mackie uses the INUS condition concept principally to analyse the structure of a complex regularity statement, rather than to analyse the structure of a statement of the form '*a* caused *b*' (see Mackie 1980).

event. However, four logically distinct patterns of co-occurrence are consistent with this requirement:

(i) It might be the case, as in Mackie's hypothetical example of the short circuit which caused the fire, that *A*-type events produce *B*-type events under *some* (but not all) conditions, but that there are also other ways in which *B*-type events can come about; if so, then it would be the case that '*A* causes *B*' and that *A* is an INUS condition of *B*.
(ii) It might be the case that *A*-type events produce *B*-type events under *some* (but not all) conditions and that there is no other way in which *B*-type events can come about; if so, then it would be the case that '*A* causes *B*' and that *A* is a necessary condition of *B*.
(iii) It might be the case that *any A*-type event will produce a *B*-type event, regardless of the conditions, but that there are also other ways in which *B*-type events can come about; if so, then it would be the case that '*A* causes *B*' and that *A* is a sufficient condition of *B*.
(iv) It might be the case that *any A*-type event will produce a *B*-type event, regardless of the conditions, and that there is no other way in which *B*-type events can come about; if so, then it would be the case that '*A* causes *B*' and that *A* is a necessary and sufficient condition of *B*.

In short, '*A* causes *B*' implies that *A* is either an INUS condition, a necessary condition, a sufficient condition, or a necessary and sufficient condition of *B*. This is why we say that the implication of '*A* causes *B*' is that *A* is *at least* an INUS condition of *B*.

Although this finding may seem significant, it is important to bear in mind that if presented with an abstract causal statement, it will typically *not* be possible to tell, just by inspecting that statement, which of these four possibilities holds. Consider some examples:

(1) 'Vitamin C deficiency causes scurvy'. As scientific knowledge developed, it was discovered that vitamin C (that is, ascorbic, or 'anti-scurvy', acid) prevents scurvy. Given that scurvy can, by definition, occur only when there is a deficiency of vitamin C, it seems correct to describe vitamin C deficiency as a cause, and also a necessary condition, of scurvy.
(2) 'Beheading causes death'.[5] Our knowledge of how the human body works suggests that it is not possible for humans to live once beheaded. It therefore seems correct to describe beheading (in humans) as a cause, and also a sufficient condition, of death.
(3) 'Revolution causes war'. Accept, for a moment, that there is such a causal relation. Inspection of the historical record will reveal that revolutions do not always produce wars and that there are other ways in which wars can come about. It therefore seems correct to describe revolution as a cause, and also an INUS condition, of war.

In each case, the relation of co-occurrence which accompanies the causal relation cannot be inferred directly from the abstract causal statement to which it is linked; in each case, discovering which pattern of co-occurrence is generated by the corresponding causal relation requires other knowledge. Consequently, although there is a systematic relationship between causal statements and statements of necessary and/or sufficient conditions, we do not believe that analysing these two kinds of statements in relation to one another offers much benefit for those conducting causal inquiries in IR: it tells us something about the logical implications of causal statements, but nothing about how to generate causal knowledge.

[5] We apologize for the gruesome example; it illustrates the difficulty of identifying truly sufficient conditions in which literally nothing can prevent the entailed event from occurring.

Rather than explicating causal statements in relation to statements of necessary and/or sufficient conditions, we believe that it is clearer, and hence preferable, to analyse causal statements in their own right, focusing on the feature of such statements which is most commonly misunderstood—that is, the fact that a statement of the form '*A* causes *B*' is a propensity statement and not a generalization: it does not imply 'whenever *A*, then *B*', but rather implies that 'an *A*-type event will produce a *B*-type event under the right conditions'. This is what we did in Chapter 4. We believe that our approach not only reduces confusion, but also helpfully directs attention towards one of the central questions that causal inquiry in IR can potentially answer, viz. under what real-world conditions do causal processes of interest unfold without interference?

Bibliography

Abbott, Andrew. 1988. 'Transcending General Linear Reality'. *Sociological Theory* **6**(2): 169–86.

Achinstein, Peter. 1977. 'What is an Explanation?' *American Philosophical Quarterly* **14**(1): 1–15.

Achinstein, Peter. 1983. *The Nature of Explanation*. New York: Oxford University Press.

Achinstein, Peter. 1993a. 'Can There Be a Model of Explanation?' In *Explanation*, edited by David-Hillel Ruben, 136–59. Oxford: Oxford University Press.

Achinstein, Peter. 1993b. 'The Pragmatic Character of Explanation'. In *Explanation*, edited by David-Hillel Ruben, 326–44. Oxford: Oxford University Press.

Angrist, Joshua D. 1990. 'Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records'. *The American Economic Review* **80**(3): 313–36.

Anievas, Alexander, Nivi Manchanda, and Robbie Shilliam, eds. 2015. *Race and Racism in International Relations: Confronting the Global Colour Line*. Abingdon: Routledge.

Ankersmit, F. R. 1986. 'The Dilemma of Contemporary Anglo-Saxon Philosophy of History'. *History and Theory* **25**(4): 1–27.

Anscombe, G. E. M. 1975. 'Causality and Determination'. In *Causation and Conditionals*, edited by Ernest Sosa, 63–81. Oxford: Oxford University Press.

Archer, Margaret, Roy Bhaskar, Andrew Collier, Tony Lawson, and Alan Norrie, eds. 1998. *Critical Realism: Essential Readings*. London: Routledge.

Ashworth, Lucien. 2014. *A History of International Thought: From the Origins of the Modern State to Academic International Relations*. Abingdon: Routledge.

Basson, A. H. 1958. *David Hume*. Harmondsworth: Penguin.

Beach, Derek and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor: University of Michigan Press.

Beach, Derek and Rasmus Brun Pedersen. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor: University of Michigan Press.

Beath, Andrew, Christina Fontini, and Ruben Enikolopov. 2012. 'Winning Hearts and Minds through Development: Evidence from a Field Experiment in Afghanistan'. World Bank Working Papers. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-6129 (accessed 16 Aug. 2023).

Beath, Andrew, Christina Fontini, and Ruben Enikolopov. 2013. *Randomized Impact Evaluation of Afghanistan's National Solidarity Programme: Final Report*. World Bank Group. https://openknowledge.worldbank.org/handle/10986/16637 (accessed 16 Aug. 2023).

Beebee, Helen. 2006. *Hume on Causation*. London: Routledge.

Beebee, Helen, Christopher Hitchcock, and Peter Menzies, eds. 2009a. *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Beebee, Helen, Christopher Hitchcock, and Peter Menzies. 2009b. 'Introduction'. In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 1–18. Oxford: Oxford University Press.

Bennett, Andrew. 2010. 'Process Tracing and Causal Inference'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 207–19. Lanham: Rowman & Littlefield.

Bennett, Andrew. 2013. 'The Mother of all Isms: Causal Mechanisms and Structured Pluralism in International Relations Theory'. *European Journal of International Relations* **19**(3): 459–81.

Bennett, Andrew and Jeffrey T. Checkel, eds. 2015a. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Bennett, Andrew, and Jeffrey T. Checkel. 2015b. 'Process Tracing: From Philosophical Roots to Best Practices'. In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffery T. Checkel, 3–37. Cambridge: Cambridge University Press.

Benton, Ted. 1977. *Philosophical Foundations of the Three Sociologies*. London: Routledge & Kegan Paul.

Berman, Eli, Joseph H. Felter, and Jacob N. Shapiro, with Vestal McIntyre. 2018. *Small Wars, Big Data: The Information Revolution in Modern Conflict*. Princeton: Princeton University Press.

Bernstein, Sara. 2017. 'Causal Idealism'. In *Idealism: New Essays in Metaphysics*, edited by Tyron Goldschmidt and Kenneth L. Pearce, 217–30. Oxford: Oxford University Press.

Bevir, Mark and Jason Blakely. 2018. *Interpretive Social Science: An Anti-Naturalist Approach*. Oxford: Oxford University Press.

Bevir, Mark and R. A. W. Rhodes. 2018. 'Interpretive Political Science: Mapping the Field'. In *Routledge Handbook of Interpretive Political Science*, edited by Mark Bevir and R. A. W. Rhodes, pbk. ed., 3–28. Abingdon: Routledge.

Bhaskar, Roy. 1998. *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*, 3rd ed. London: Routledge.

Bhaskar, Roy. 2008. *A Realist Theory of Science*, 3rd ed. London: Verso.

Bird, Alexander. 2005. 'Abductive Knowledge and Holmesian Inference'. In *Oxford Studies in Epistemology*, edited by Tamar Szabo Gendler and John Hawthorne, 1–31. Oxford: Clarendon Press.

Bird, Alexander. 2007. 'Inference to the Only Explanation'. *Philosophy and Phenomenological Research* **74**(2): 424–32.

Bird, Alexander. 2010. 'Eliminative Abduction: Examples from Medicine'. *Studies in History and Philosophy of Science* **41**(4): 345–52.

Blair, Graeme, C. Christine Fair, Neil Malhotra, and Jacob N. Shapiro. 2013. 'Poverty and Support for Militant Politics: Evidence from Pakistan'. *American Journal of Political Science* **57**(1): 30–48.

Boyd, Nora Mills and James Bogen. 2021. 'Theory and Observation in Science'. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/science-theory-observation/ (accessed 30 Nov. 2022).

Brady, Henry E. 2008. 'Causation and Explanation in Social Science'. In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 217–70. Oxford: Oxford University Press.

Brady, Henry. 2010. 'Doing Good and Doing Better: How Far does the Quantitative Template Get Us?' In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 67–82. Lanham: Rowman & Littlefield.

Brady, Henry E. and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 1st ed. Lanham: Rowman & Littlefield.

Brady, Henry E. and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Lanham: Rowman & Littlefield.

Braumoeller, Bear F. and Gary Goertz. 2000. 'The Methodology of Necessary Conditions'. *American Journal of Political Science* **44**(4): 844–58.

Broadbent, Alex. 2008. 'The Difference Between Cause and Condition'. Proceedings of the Aristotelian Society **108**(3): 355–64.

Bueno de Mesquita, Bruce. 1985. 'Toward a Scientific Understanding of International Conflict: A Personal View'. *International Studies Quarterly* **29**(2): 121–36.

Bull, Hedley. 1966. 'International Theory: The Case for a Classical Approach'. *World Politics* **18**(3): 361–77.

Bunge, Mario. 1997. 'Mechanism and Explanation'. *Philosophy of the Social Sciences* **27**(4): 410–65.

Bunge, Mario. 2004. 'How Does it Work? The Search for Explanatory Mechanisms'. *Philosophy of the Social Sciences* **34**(2): 182–210.

Campbell, David. 1998a. *National Deconstruction: Violence, Identity, and Justice in Bosnia*. London: University of Minnesota Press.

Campbell, David. 1998b. *Writing Security: United States Foreign Policy and the Politics of Identity*, 2nd ed. Minnesota: University of Minnesota Press.
Campbell, David. 1998c. 'MetaBosnia: Narratives of the Bosnian War'. *Review of International Studies* **24**(2): 261–81.
Carr, David. 2008. 'Narrative Explanation and its Malcontents'. *History and Theory* **47**(1): 19–30.
Carr, E. H. 2002. *What Is History?* 40th Anniversary ed., with a new introduction by Richard J. Evans. Basingstoke: Palgrave.
Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.
Cartwright, Nancy. 1989. *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.
Cartwright, Nancy. 1993. 'In Defence of "This Worldly" Causality: Comments on Van Fraassen's Laws and Symmetry'. *Philosophy and Phenomenological Research* **53**(2): 423–9.
Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Cartwright, Nancy. 2002. 'From Causation to Explanation and Back'. *LSE Centre for the Philosophy of Natural and Social Science*. https://core.ac.uk/download/pdf/73975.pdf (accessed 10 Sep. 2023).
Cartwright, Nancy. 2004. 'Causation: One Word, Many Things'. *Philosophy of Science* **71**(5): 805–19.
Cartwright, Nancy. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.
Cartwright, Nancy. 2010. 'What are Randomized Controlled Trials Good For?' *Philosophical Studies* **147**(1): 59–70.
Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
Chakravartty, Anjan. 2007. *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.
Chakravartty, Anjan. 2017. 'Scientific Realism'. In *Stanford Encyclopedia of Philosophy*: https://plato.stanford.edu/entries/scientific-realism (accessed 22 Sep. 2023).
Chan, Steve. 1997. 'In Search of Democratic Peace: Problems and Promise'. *Mershon International Studies Review* **41**(1): 59–91.
Checkel, Jeffrey T. 2006. 'Tracing Causal Mechanisms'. *International Studies Review* **8**(2): 362–70.
Chernoff, Fred. 2002. 'Scientific Realism as a Meta-Theory of International Politics'. *International Studies Quarterly* **46**(2): 189–207.
Chernoff, Fred. 2005. *The Power of International Theory: Reforging the Link to Foreign Policy-Making through Scientific Inquiry*. London: Routledge.
Chernoff, Fred. 2007a. *Theory and Metatheory in International Relations: Concepts and Contending Accounts*. Basingstoke: Palgrave.
Chernoff, Fred. 2007b. 'Critical Realism, Scientific Realism, and International Relations Theory'. *Millennium: Journal of International Studies* **35**(2): 399–407.
Chernoff, Fred. 2014. *Explanation and Progress in Security Studies: Bridging Theoretical Divides in International Relations*. Stanford, CA.: Stanford University Press.
Churchland, Paul M. 1985. 'The Ontological Status of Unobservables: In Praise of the Superempirical Virtues'. In *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, edited by Paul M. Churchland and Clifford A. Hooker, 35–47. Chicago: University of Chicago Press.
Churchland, Paul M. and Clifford A. Hooker. 1985. *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*. Chicago: University of Chicago Press.
Clarke, Stephen. 2010. 'Transcendental Realisms in the Philosophy of Science: On Bhaskar and Cartwright'. *Synthese* **173**(3): 299–315.
Collier, Andrew. 1998. 'Explanation and Emancipation'. In *Critical Realism: Essential Readings*, edited by Margaret Archer, Roy Bhaskar, Andrew Collier, Tony Lawson, and Alan Norrie, 444–72. London: Routledge.

Collier, David. 2011. 'Understanding Process Tracing'. *PS: Political Science and Politics* **44**(4): 823–30.

Collier, David, Henry E. Brady, and Jason Seawright. 2010a. 'Critiques, Responses, and Trade-Offs'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 125–59. Lanham: Rowman & Littlefield.

Collier, David, Henry E. Brady, and Jason Seawright. 2010b. 'Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 161–99. Lanham: Rowman & Littlefield.

Collier, David, James Mahoney, and Jason Seawright. 2004. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 85–102. Lanham: Rowman & Littlefield.

Collier, David, Jason Seawright, and Gerardo L. Munck. 2010. 'The Quest for Standards: King, Keohane, and Verba's *Designing Social Inquiry*'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 33–63. Lanham: Rowman & Littlefield.

Collingwood, R. G. 1937–8. 'On the So-Called Idea of Causation'. Proceedings of the Aristotelian Society **38**: 85–112.

Cox, Robert W. 1981. 'Social Forces, States, and World Orders: Beyond International Relations Theory'. *Millennium: Journal of International Studies* **10**(2): 126–55.

Crost, Benjamin, Joseph H. Felter, and Patrick B. Johnston. 2014. 'Aid Under Fire: Development Projects and Civil Conflict'. *American Economic Review* **104**(6): 1833–56

Crost, Benjamin, Joseph H. Felter, and Patrick B. Johnston. 2016. 'Conditional Cash Transfers, Civil Conflict and Insurgent Influence: Experimental Evidence from the Philippines'. *Journal of Development Economics* **118**: 171–82.

Cudworth, Erika and Stephen Hobden. 2012. 'The Foundations of Complexity, the Complexity of Foundations'. *Philosophy of the Social Sciences* **42**(2): 163–87.

Davidson, Donald. 1963. 'Actions, Reasons, and Causes'. *Journal of Philosophy* **60**(23): 685–700.

Davidson, Donald. 1975. 'Causal Relations'. In *Causation and Conditionals*, edited by Ernest Sosa, 82–94. Oxford: Oxford University Press.

Dean, Kathryn, Jonathan Joseph, John Michael Roberts, and Colin Wight. 2006. *Realism, Philosophy and Social Science*. Houndmills: Palgrave Macmillan.

De Carvalho, Benjamin, Julia Costa Lopez, and Halvard Leira, eds. 2021. *Routledge Handbook of Historical International Relations*. London: Routledge.

Dessler, David. 1991. 'Beyond Correlations: Toward a Causal Theory of War'. *International Studies Quarterly* **35**(3): 337–55.

Dessler, David. 2003. 'Explanation and Scientific Progress'. In *Progress in International Relations Theory: Appraising the Field*, edited by Colin Elman and Miriam Fendius Elman, 381–404. Cambridge, MA MIT Press.

Diamond, Jared and James A. Robinson. 2011. *Natural Experiments of History*. Cambridge, MA: Belknap Press of Harvard University Press.

Douven, Igor. 2021. 'Abduction'. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/abduction/ (accessed 27 July 2023).

Doyle, Michael W. 2005. 'Three Pillars of the Liberal Peace'. *American Political Science Review* **99**(3): 463–6.

Dray, William. 1957. *Laws and Explanation in History*. Oxford: Oxford University Press.

Dray, W. H. 1971. 'On the Nature and Role of Narrative in Historiography'. *History and Theory* **10**(2): 153–71.

Dray, W. H. 1978. 'Concepts of causation in A. J. P. Taylor's account of the origins of the Second World War'. *History and Theory* **17**(2): 149–74.

Ducasse, C. J. 1975. 'On the Nature and the Observability of the Causal Relation'. In *Causation and Conditionals*, edited by Ernest Sosa, 114–25. Oxford: Oxford University Press.

Dunning, Thad. 2010. 'Design-Based Inference: Beyond the Pitfalls of Regression Analysis?' In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 273–311. Lanham: Rowman & Littlefield.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.

Durkheim, Emile. 1952. *Suicide: A Study in Sociology*, edited with an Introduction by George Simpson. London: Routledge and Kegan Paul.

Durkheim, Emile. 1976. *The Elementary Forms of the Religious Life*, 2nd ed., with a new Introduction by Robert Nisbet, translated by Joseph Ward Swain. London: Allen & Unwin.

Egami, Naoki and Erin Hartman. 2022. 'Elements of External Validity: Framework, Design, and Analysis'. *American Political Science Review*, First View: 1–19. https://www.cambridge.org/core/journals/american-political-science-review/article/elements-of-external-validity-framework-design-and-analysis/2D0914404C84B3F169732FF1D5E39420 (accessed 5 July 2023).

Ehring, Douglas. 2009. 'Causal Relata'. In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 387-413. Oxford: Oxford University Press.

Elias, Norbert. 1991. *The Symbol Theory*, edited with an Introduction by Richard Kilminster. London: Sage.

Elman, Colin, Miriam Fendius Elman, Jack S. Levy, Stephen H. Haber, David M. Kennedy, Stephen D. Krasner, Alexander L. George, Edward Ingram, Paul W. Schroeder, and John Lewis Gaddis. 1997. 'Symposium on Diplomatic History and International Relations'. *International Security* **22**(1): 5–85.

Elman, Colin and Miriam Fendius Elman, eds. 2001a. *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*. Cambridge, MA: MIT Press.

Elman, Colin and Miriam Fendius Elman. 2001b. 'Introduction: Negotiating International History and Politics'. In *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*, edited by Colin Elman and Miriam Fendius Elman, 1–36. Cambridge, MA: MIT Press.

Elman, Colin, Miriam Fendius Elman, Yale H. Ferguson, Richard W. Mansbach, Edward Keene, Christian Reus-Smit, John M. Hobson, and George Lawson. 2008. 'Forum on the Role of History in International Relations'. *Millennium: Journal of International Studies* **37**(2): 357–435.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Elster, Jon. 2007. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Fair, C. Christine, Rebecca Littman, Neil Malhotra, and Jacob N. Shapiro, 2018. 'Relative Poverty, Perceived Violence, and Support for Militant Politics: Evidence from Pakistan', *Political Science Research and Methods* **6**(1): 57–81.

Falletti, Tulia G. and Julia F. Lynch. 2009. 'Context and Causal Mechanisms in Political Analysis'. *Comparative Political Studies* **42**(9): 1143–66.

Findley, Michael G., Kyosuke Kikuta, Michael Denly. 2021. 'External Validity'. *Annual Review of Political Science* **24**: 365–93.

Finnemore, Martha and Kathryn Sikkink. 1998. 'International Norm Dynamics and Political Change'. *International Organization* **52**(4): 887–917.

Friedman, Milton. 1994. 'The Methodology of Positive Economics'. In *The Philosophy of Economics: An Anthology*, edited by Daniel M. Hausman, 145–78. Cambridge: Cambridge University Press.

Gaddis, John Lewis. 1997. 'History, Theory, and Common Ground'. *International Security* **22**(1): 75–85.

Gaddis, John Lewis. 2001. 'In Defense of Particular Generalization: Rewriting Cold War History, Rethinking International Relations Theory'. In *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*, edited by Colin Elman and Miriam Fendius Elman, 301–26. Cambridge, MA: MIT Press.

Gaddis, John Lewis. 2002. *The Landscape of History: How Historians Map the Past*. Oxford: Oxford University Press.

Galiani, Sebastian, and Ernesto Schargrodsky. 2010. 'Property Rights for the Poor: Effects of Land Titling'. *Journal of Public Economics* **94**(9–10): 700–29.
Gardiner, Patrick, ed. 1959. *Theories of History*. Glencoe: Free Press.
George, Alexander L. 1979. 'Case Studies and Theory Development: The Method of Structured, Focused Comparison'. In *Diplomatic History: New Approaches*, edited by Paul Gordon Lauren, 43–68. New York: Free Press.
George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
George, Jim and David Campbell. 1990. 'Patterns of Dissent and the Celebration of Difference: Critical Social Theory and International Relations'. *International Studies Quarterly* **34**(3): 269–93.
Gerring, John. 2005. 'Causation: A Unified Framework for the Social Sciences'. *Journal of Theoretical Politics* **17**(2): 163–98.
Gerring, John. 2006. 'Single-Outcome Studies: A Methodological Primer'. *International Sociology* **21**(5): 707–34.
Gerring, John. 2008. 'The Mechanismic World View: Thinking Inside the Box'. *British Journal of Political Science* **38**(1): 161–79.
Gerring, John. 2010. 'Causal Mechanisms: Yes, But …'. *Comparative Political Studies* **43**(11): 1499–526.
Gerring, John. 2011. 'The Case Study: What it Is and What it Does'. In *The Oxford Handbook of Political Science*, edited by Robert E. Goodin, 1133–65. Oxford: Oxford University Press.
Gerring, John. 2012. *Social Science Methodology: A Unified Framework*, 2nd ed. Cambridge: Cambridge University Press.
Gerring, John. 2017. *Case Study Research: Principles and Practices*, 2nd ed. Cambridge: Cambridge University Press.
Getachew, Adom. 2019. *Worldmaking After Empire: The Rise and Fall of Self-Determination*. Princeton: Princeton University Press.
Gies, Anna, Lothar Brock, and Harald Müller, eds. 2006. *Democratic Wars: Looking at the Dark Side of Democratic Peace*. Basingstoke: Palgrave Macmillan.
Goddard, Stacie E. 2015. 'The Rhetoric of Appeasement: Hitler's Legitimation and British Foreign Policy, 1938–39'. *Security Studies* **24**(1): 95–130.
Goddard, Stacie E. 2018. *When Right Makes Might: Rising Powers and World Order*. Ithaca: Cornell University Press.
Goertz, Gary and Jack S. Levy (eds.). 2007. *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*. London: Routledge.
Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press.
Goldstein, Judith and Robert O. Keohane. 1993. 'Ideas and Foreign Policy: An Analytical Framework'. In *Ideas and Foreign Policy: Beliefs, Institutions, and Political Change*, edited by Judith Goldstein and Robert O. Keohane, 3–30. Ithaca: Cornell University Press.
Goldthorpe, John H. 2001. 'Causation, Statistics, and Sociology'. *European Sociological Review* **17**(1): 1–20.
Grynaviski, Eric. 2013. 'Contrasts, Counterfactuals, and Causes'. *European Journal of International Relations* **19**(4): 823–46.
Guilhot, Nicolas. 2008. 'The Realist Gambit: Postwar American Political Science and the Birth of IR Theory'. *International Political Sociology* **2**(4): 281–304.
Guilhot, Nicolas, ed. 2011. The Invention of International Relations Theory: Realism, the Rockefeller Foundation, and the 1954 Conference on Theory. New York: Columbia University Press.
Gunitsky, Seva. 2013. 'Complexity and Theories of Change in International Politics'. *International Theory* **5**(1): 35–63.
Guzzini, Stefano, ed. 2012. *The Return of Geopolitics in Europe: Social Mechanisms and Foreign Policy Identity Crises*. Cambridge: Cambridge University Press.

Haas, Michael. 1974. *International Conflict*. Indianapolis: Bobbs-Merrill.
Hacking, Ian. 1985. 'Do We See Through a Microscope'. In *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, edited by Paul M. Churchland and Clifford A. Hooker, 132–52. Chicago: University of Chicago Press.
Hanson, N. R. 1955. 'Causal Chains'. *Mind* **64**(255): 289–311.
Harman, Gilbert H. 1965. 'The Inference to the Best Explanation'. *Philosophical Review* **74**(1): 88–95.
Harré, Rom. 1964. 'Concepts and Criteria'. *Mind* **73**(291): 353–63.
Harré, Rom. 1972. *The Philosophies of Science: An Introductory Survey*. Oxford: Oxford University Press.
Harré, Rom. 1981. *Great Scientific Experiments: Twenty Experiments that Changed our View of the World*. New York: Dover
Harré, Rom and E. H. Madden. 1975. *Causal Powers: A Theory of Natural Necessity*. Oxford: Basil Blackwell.
Harrison, Neil E., ed. 2006a. *Complexity in World Politics: Concepts and Methods of a New Paradigm*. Albany: State University of New York Press.
Harrison, Neil E. 2006b. 'Thinking about the World We Make'. In *Complexity in World Politics: Concepts and Methods of a New Paradigm*, edited by Neil E. Harrison, 1–23. Albany: State University of New York Press.
Hart, H. L. A. and Tony Honoré. 1985. *Causation in the Law*, 2nd ed. Oxford: Clarendon Press.
Hayes, Jarrod. 2012. 'The Democratic Peace and the New Evolution of an Old Idea'. *European Journal of International Relations* **18**(4): 767–91.
Hedström, Peter and Richard Swedberg, eds. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
Hedström, Peter and Petri Ylikoski. 2010. 'Causal Mechanisms in the Social Sciences'. *Annual Review of Sociology* **36**: 49–67.
Hempel, Carl G. 1942. 'The Function of General Laws in History'. *Journal of Philosophy* **39**(2): 35–48.
Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.
Hempel, Carl G. 1966. *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.
Hempel, Carl G. 2001. 'Explanation in Science and History'. In *The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality*, edited by James H. Fetzer, 276–96. Oxford: Oxford University Press.
Hempel, Carl G. and Paul Oppenheim. 1948. 'Studies in the Logic of Explanation'. *Philosophy of Science* **15**(2): 135–75.
Hesslow, Germund. 1988. 'The Problem of Causal Selection'. In *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality*, edited by Denis J. Hilton, 11–32. New York: New York University Press.
Hitchcock, Christopher Read. 1995. 'The Mishap at Reichenbach Fall: Singular vs. General Causation'. *Philosophical Studies* **78**(3): 257–91.
Hitchcock, Christopher. 2018. 'Probabilistic Causation'. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/causation-probabilistic/ (last accessed 27 Jan. 2022).
Hobson, John M. and George Lawson. 2008. 'What is History in International Relations?' *Millennium: Journal of International Studies* **37**(2): 415–35.
Holland, Paul W. 1986. 'Statistics and Causal Inference'. *Journal of the American Statistical Association* **81**(396): 945–60.
Hollis, Martin and Steve Smith. 1991a. *Explaining and Understanding International Relations*. Oxford: Clarendon Press.
Hollis, Martin and Steven Smith. 1991b. 'Beware of Gurus: Structure and Action in International Relations', *Review of International Studies* 17(4): 393–410.

Hume, David. 1962. *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, prepared from the posthumous edition of 1777 and edited with Introduction, Comparative Tables, and Analytical Index by L. A. Selby-Bigge. Oxford: Clarendon Press.
Hume, David. 1969. *A Treatise of Human Nature*, edited with an Introduction by Ernest C. Mossner. First published in 1739 and 1740. London: Penguin Books.
Humphreys, Adam R. C. 2011. 'The Heuristic Application of Explanatory Theories in International Relations'. *European Journal of International Relations* **17**(2): 257–77.
Humphreys, Adam R. C. 2012. 'Another Waltz? Methodological Rhetoric and Practice in *Theory of International Politics*'. *International Relations* **26**(4): 389–408.
Humphreys, Adam R. C. 2017a. 'Introduction: Problems of Causation in International Relations'. *Journal of International Relations and Development* **20**(4): 659–66.
Humphreys, Adam R. C. 2017b. 'Complexity, Causation, and the Concert: The Pragmatics of Causal Explanation in International Relations'. *Journal of International Relations and Development* **20**(4): 717–36.
Humphreys, Adam R. C. 2019. 'Realism, Empiricism, and Causal Inquiry in International Relations: What is at Stake?' *European Journal of International Relations* **25**(2): 562–87.
Hunt, Lynn. 2014. *Writing History in the Global Era*. New York: W. W. Norton.
Hyde, Susan D. 2007. 'The Observer Effect in International Politics: Evidence from a Natural Experiment'. *World Politics* **60**(1): 37–63.
Illari, Phyllis and Federica Russo. 2014. *Causality: Philosophical Theory Meets Scientific Practice*. Oxford: Oxford University Press.
Imai, Kosuke 2017. *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.
Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. 'Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies'. *American Political Science Review* **105**(4): 765–89.
Isacoff, Jonathan B. 2002. 'On the Historical Imagination of International Relations: The Case for a "Deweyan Reconstruction"'. *Millennium: Journal of International Studies* **31**(3): 603–26.
Jackson, Patrick Thaddeus. 2011. *The Conduct of Inquiry in International Relations: Philosophy of Science and its Implications for the Study of World Politics*. London: Routledge.
Jackson, Patrick Thaddeus. 2016. *The Conduct of Inquiry in International Relations: Philosophy of Science and its Implications for the Study of World Politics*, 2nd ed. London: Routledge.
Jackson, Patrick Thaddeus. 2017. 'Causal Claims and Causal Explanation in International Studies'. *Journal of International Relations and Development* **20**(4): 689–716.
Jackson, Patrick Thaddeus. Forthcoming. *Facts and Explanations in International Studies … and Beyond*. London: Routledge.
James, William. 1907. *Pragmatism: A New Name for some Old Ways of Thinking*. Cambridge, MA: Harvard University Press.
Jervis, Robert. 1997. *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.
Johnson, James. 2002. 'How Conceptual Problems Migrate: Rational Choice, Interpretation, and the Hazards of Pluralism'. *Annual Review of Political Science* **5**: 223–48.
Johnson, James. 2006. 'Consequences of Positivism: A Pragmatist Assessment'. *Comparative Political Studies* **39**(2): 224–52.
Johnson, Janet Buttolph and H. T. Reynolds. 2012. *Political Science Research Methods*, 7th ed. Los Angeles: CQ Press.
Joseph, Jonathan. 2007. 'Philosophy in International Relations: A Scientific Realist Approach'. *Millennium: Journal of International Studies* **35**(2): 345–59.
Joseph, Jonathan. 2010. 'Is Waltz a Realist?' *International Relations* **24**(4): 478–93.
Joseph, Jonathan and Colin Wight, eds. 2010. *Scientific Realism and International Relations*. London: Palgrave Macmillan.
Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War*. Cambridge: Cambridge University Press.
Keat, Russell and John Urry. 1975. *Social Theory as Science*. London: Routledge.

Keene, Edward. 2005. *International Political Thought: A Historical Introduction*. Cambridge: Polity Press.
Keohane, Robert O. 1988. 'International Institutions: Two Approaches'. *International Studies Quarterly* **32**(4): 379–96.
Kim, Jaegwon. 1975. 'Causes and Events: Mackie on Causation'. In *Causation and Conditionals*, edited by Ernest Sosa, 48–62. Oxford: Oxford University Press.
Kim, Jaegwon. 1993. 'Explanatory Realism, Causal Realism, and Explanatory Exclusion'. In *Explanation*, edited by David-Hillel Ruben, 228–45. Oxford: Oxford University Press.
King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
Kiser, Edgar and Michael Hechter. 1991. 'The Role of General Theory in Comparative-Historical Sociology'. *American Journal of Sociology* **97**(1): 1–30.
Kneale, William. 1949. *Probability and Induction*. Oxford: Clarendon Press.
Kratochwil, Friedrich. 1993. 'The Embarrassment of Changes: Neo-Realism as the Science of Realpolitik Without Politics'. *Review of International Studies* **19**(1): 63–80.
Kurki, Milja. 2006. 'Causes of a Divided Discipline: Rethinking the Concept of Cause in International Relations Theory'. *Review of International Studies* **32**(2): 189–216.
Kurki, Milja. 2007. 'Critical Realism and Causal Analysis in International Relations'. *Millennium: Journal of International Studies* **35**(2): 361–78.
Kurki, Milja. 2008. *Causation in International Relations: Reclaiming Causal Analysis*. Cambridge: Cambridge University Press.
Kurki, Milja and Hidemi Suganami. 2012. 'Towards the Politics of Causal Explanation: A Reply to the Critics of Causal Inquiries'. *International Theory* **4**(3): 400–29.
Ladyman, James. 2000. 'What's Really Wrong with Constructive Empiricism? Van Fraassen and the Metaphysics of Modality'. *British Journal for the Philosophy of Science* **51**(4): 837–56.
Ladyman, James, Igor Douven, Leon Horsten, and Bas van Fraassen. 1997. 'A Defence of Van Fraassen's Critique of Abductive Inference: Reply to Psillos'. *Philosophical Quarterly* **47**(188): 305–21.
Lake, David. 2011. 'Why "isms" are Evil: Theory, Epistemology, and Academic Sects as Impediments to Understanding and Progress'. *International Studies Quarterly* **55**(2): 465–80.
Lapid, Yosef. 1989. 'The Third Debate: On the Prospects of International Theory in a Post-Positivist Era'. *International Studies Quarterly* **33**(3): 235–54.
Lawson, George. 2012. 'The Eternal Divide? History and International Relations'. *European Journal of International Relations* **18**(2): 203–26.
Lawson, George. 2018. 'International Relations as a Historical Social Science'. In *The SAGE Handbook of the History, Philosophy and Sociology of International Relations*, edited by Andreas Gofas, Inanna Hamati-Ataya, and Nicholas Onuf, 75–89. London: SAGE.
Layne, Christopher. 1994. 'Kant or Cant: The Myth of the Democratic Peace'. *International Security* **19**(2): 5–49.
Lebow, Richard Ned. 2014. *Constructing Cause in International Relations*. Cambridge: Cambridge University Press.
Lebow, Richard Ned. 2020. *Reason and Cause: Social Science and the Social World*. Cambridge: Cambridge University Press.
Lentini, John J. 2013. 'Fire Patterns and their Interpretation'. In *Encyclopedia of Forensic Sciences*, edited by Jay A. Siegel and Pekka J. Saukko, 2nd ed., vol. 3, 396–405. Waltham: Academic Press. https://doi.org/10.1016/B978-0-12-382165-2.00213-0 (accessed 5 Jan. 2023)
Leon, David. 2010. 'Reductionism, Emergence and Explanation in International Relations Theory'. In *Scientific Realism and International Relations*, edited by Jonathan Joseph and Colin Wight, 31–50. London: Palgrave Macmillan.
Levy, Jack S. 1997. 'Too Important to Leave to the Other: History and Political Science in the Study of International Relations'. *International Security* **22**(1): 22–33.
Levy, Jack S. 2001. 'Explaining Events and Developing Theories: History, Political Science, and the Analysis of International Relations'. In *Bridges and Boundaries: Historians, Political Scientists, and*

the Study of International Relations, edited by Colin Elman and Miriam Fendius Elman, 39–83. Cambridge, MA: MIT Press.
Levy, Jack S. 2008. 'Case Studies: Types, Designs, and Logics of Inference'. *Conflict Management and Peace Science* 25(1): 1–18.
Levy, Jack S. 2015. 'Counterfactuals, Causal Inference, and Historical Analysis'. *Security Studies* 24(3): 378–402.
Levy, Jack S., Thomas J. Christensen, and Marc Trachtenberg. 1991. 'Correspondence: Mobilization and Inadvertence in the July Crisis'. *International Security* **16**(1): 189–203.
Lewis, David. 1975. 'Causation'. In *Causation and Conditionals*, edited by Ernest Sosa, 180–94. Oxford: Oxford University Press.
Lewis, David. 1993. 'Causal Explanation'. In *Explanation*, edited by David-Hillel Ruben, 182–206. Oxford: Oxford University Press.
Lipton, Peter. 1993. 'Contrastive Explanation'. In *Explanation*, edited by David-Hillel Ruben, 207–27. Oxford: Oxford University Press.
Lipton, Peter. 2004. *Inference to the Best Explanation*, 2nd ed. London: Routledge.
Lipton, Peter. 2007a. 'Précis of *Inference to the Best Explanation*, 2nd Edition'. *Philosophy and Phenomenological Research* **74**(2): 421–3.
Lipton, Peter. 2007b. 'Replies'. *Philosophy and Phenomenological Research* **74**(2): 449–62.
Little, Daniel. 1991. *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder: Westview Press.
Little, Daniel. 1995. 'Causal Explanation in the Social Sciences'. *Southern Journal of Philosophy* **34**(Supplement): 31–56.
Locke, John. 1996. *An Essay concerning Human Understanding*, abridged and edited by Kenneth P. Winkler. First published in 1689. Indianapolis: Hackett.
Lucas, Jeffrey W. 2003. 'Theory-Testing, Generalization, and the Problem of External Validity'. *Sociological Theory* **21**(3): 236–53.
Lyall, Jason. 2009. 'Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya'. *Journal of Conflict Resolution* **53**(3): 331–62.
Lyall, Jason, Graeme Blair, and Kosuke Imai. 2013. 'Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan'. *American Political Science Review* **107**(4): 679–705.
Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. 'Thinking about Mechanisms'. *Philosophy of Science* **67**(1): 1–25.
Mackie, J. L. 1965. 'Causes and Conditions'. *American Philosophical Quarterly* 4(2): 245–64.
Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.
Mahoney, James. 2001. 'Beyond Correlational Analysis: Recent Innovations in Theory and Method'. *Sociological Forum* **16**(3): 575–93.
Mahoney, James. 2008. 'Toward a Unified Theory of Causality'. *Comparative Political Studies* 41(4/5): 412–36.
Mahoney, James. 2012. 'The Logic of Process Tracing Tests in the Social Sciences'. *Sociological Methods and Research* **41**(4): 570–97.
Mahoney, James. 2015. 'Process Tracing and Historical Explanation'. *Security Studies* **24**(2): 200–18.
Mahoney, James and Rodrigo Barrenechea. 2019. 'The Logic of Counterfactual Analysis in Case-Study Explanation'. *British Journal of Sociology* 70(1): 306–38.
Mahoney, James and Gary Goertz. 2006. 'A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research'. *Political Analysis* 14(3): 227–49.
Mahoney, James, Erin Kimball, and Kendra L. Koivu. 2009. 'The Logic of Historical Explanation in the Social Sciences'. *Comparative Political Studies* **42**(1): 114–46.
Manicas, Peter T. 2006. *A Realist Philosophy of Social Science: Explanation and Understanding*. Cambridge: Cambridge University Press.
Maoz, Zeev and Bruce Russett. 1993. 'Normative and Structural Causes of Democratic Peace, 1946–1986'. *American Political Science Review* **87**(3): 624–38.

Martin, Raymond. 1982. 'Causes, Conditions, and Causal Importance'. *History and Theory* **21**(1): 53–74.
McDermott, Rose. 2011. 'Internal and External Validity'. In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Greene, James H. Kuklinski, and Arthur Lupia, 27–40. Cambridge: Cambridge University Press.
McKeown, Timothy J. 2004. 'Case Studies and the Limits of the Quantitative Worldview'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 1st ed., 139–67. Lanham: Rowman & Littlefield.
Mearsheimer, John J. and Stephen M. Walt. 2013. 'Leaving Theory Behind: Why Simplistic Hypothesis Testing is Bad for International Relations'. *European Journal of International Relations* **19**(3): 427–57.
Mellor, D. H. 1995. *The Facts of Causation*. London: Routledge.
Mink, Louis O. 1966. 'The Autonomy of Historical Understanding'. *History and Theory* **5**(1): 24–47.
Mink, Louis O. 1967–8. 'Philosophical Analysis and Historical Understanding'. *Review of Metaphysics*. 21: 667–98.
Mink, Louis O. 1970. 'History and Fiction as Modes of Comprehension'. *New Literary History* **1**(3): 541–58.
Mink, Louis O. 1987. 'Narrative Form as a Cognitive Instrument'. In *Historical Understanding*, edited by Brian Fay, Eugene O. Golob, and Richard T. Vann, 182–203. Ithaca: Cornell University Press.
Monton, Bradley, ed. 2007. *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*. Oxford: Oxford University Press.
Monton, Bradley, and Chad Mohler. 2021. 'Constructive Empiricism'. In *Stanford Encyclopedia of Philosophy*: https://plato.stanford.edu/entries/constructive-empiricism (accessed 22 Sep. 2023).
Monton, Bradley, and Van Fraassen, Bas C. 2003. 'Constructive Empiricism and Modal Nominalism'. *British Journal for the Philosophy of Science* **54**(3): 405–22.
Morgan, Jamie and Heikki Patomäki. 2017. 'Contrast Explanation in Economics: Its Context, Meaning, and Potential'. *Cambridge Journal of Economics* **41**(5): 1391–418.
Müller, Harald. 2004. 'The Antinomy of Democratic Peace'. *International Politics* **41**(4): 494–520.
Munslow, Alan. 2006. *Deconstructing History*, 2nd ed. London: Routledge.
Musgrave, Alan. 1985. 'Realism versus Constructive Empiricism'. In *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, edited by Paul M. Churchland and Clifford A. Hooker, 197–221. Chicago: University of Chicago Press.
Newton, Isaac. 1995. *The Principia*, translated by Andrew Motte. First published in 1713. New York: Prometheus Books.
Noonan, Harold W. 1999. *Hume on Knowledge*. London: Routledge.
Norman, Ludvig. 2021. 'Rethinking Causal Explanation in Interpretive International Studies'. *European Journal of International Relations* **27**(3): 936–59.
Onuf, Nicolas. 2009. 'Structure? What Structure?' *International Relations* **23**(2): 183–99.
Owen, John M. 1994. 'How Liberalism Produces Democratic Peace'. *International Security* **19**(2): 87–125.
Owens, Patricia and Katharina Rietzler, eds. 2021. *Women's International Thought: A New History*. Cambridge: Cambridge University Press.
Pape, Robert A. 1996. *Bombing to Win: Air Power and Coercion in War*. Ithaca: Cornell University Press.
Patomäki, Heikki. 2002. *After International Relations: Critical Realism and the (Re)construction of World Politics*. London: Routledge.
Patomäki, Heikki. 2008. *The Political Economy of Global Security: War, Future Crises and Changes in Global Governance*. London: Routledge.
Patomäki, Heikki. 2010. 'Exploring Possible, Likely and Desirable Global Futures: Beyond the Closed vs Open Systems Dichotomy'. In *Scientific Realism and International Relations*, edited by Jonathan Joseph and Colin Wight, 147–66. London: Palgrave Macmillan.

Patomäki, Heikki. 2017. 'Praxis, Politics and the Future: A Dialectical Critical Realist Account of World-Historical Causation'. *Journal of International Relations and Development* **20**(4): 805–25.
Patomäki, Heikki and Colin Wight. 2000. 'After Positivism? Promises of Critical Realism'. *International Studies Quarterly* **44**(2): 213–37.
Paul, L. A. 2009. 'Counterfactual Theories'. In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 158–84. Oxford: Oxford University Press.
Posner, Daniel N. 2004. 'The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi'. *American Political Science Review* **98**(4): 529–45.
Psillos, Stathis. 1996. 'On Van Fraassen's Critique of Abductive Reasoning'. *Philosophical Quarterly* **46**(182): 31–47.
Psillos, Stathis. 2002. *Causation and Explanation*. Chesham: Acumen.
Psillos, Stathis. 2009. 'Regularity Theories'. In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 131–57. Oxford: Oxford University Press.
Putnam, Hilary. 1975. *Mathematics, Matter and Method*: *Philosophical Papers, Vol. 1*. Cambridge: Cambridge University Press.
Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
Ragin, Charles C. 2004. 'Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 1st ed., 123–38. Lanham: Rowman & Littlefield.
Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
Ragin, Charles C. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, with a new introduction. Oakland: University of California Press.
Ray, James Lee. 1993. 'Wars Between Democracies: Rare, or Nonexistent?' *International Interactions* **18**(3): 251–76.
Ray, James Lee. 1995. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition*. Columbia: University of South Carolina Press.
Ray, James Lee. 1998. 'Does Democracy Cause Peace?' *Annual Review of Political Science* **1**: 27–46.
Read, Rupert and Kenneth A. Richman, eds. 2007. *The New Hume Dabate*. London: Routledge. Rev ed.
Reid, Thomas. 1827. *Essays on the Powers of the Human Mind; to which are added An Essay on Quantity and An Analysis of Aristotle's Logic*. London: Thomas Tegg.
Reiss, Julian. 2009. 'Causation in the Social Sciences: Evidence, Inference, and Purpose'. *Philosophy of the Social Sciences* **39**(1): 20–40.
Reiter, Dan. 2017. 'Is Democracy a Cause of Peace?' *Oxford Research Encyclopedias*. https://doi.org/10.1093/acrefore/9780190228637.013.287 (accessed 24 Aug. 2023).
Ringer, Fritz K. 1989. 'Causal Analysis in Historical Reasoning', *History and Theory* **28**(2): 154–72.
Risse-Kappen, Thomas. 1995. 'Democratic Peace – Warlike Democracies? A Social Constructivist Interpretation of the Liberal Argument'. *European Journal of International Relations* **1**(4): 491–517.
Roberts, Geoffrey. 2006. 'History, Theory and the Narrative Turn in IR'. *Review of International Studies* **32**(4): 703–14.
Rohlfing, Ingo. 2014. 'Comparative Hypothesis Testing via Process Tracing'. *Sociological Methods and Research* **43**(4): 606–42.
Rosato, Sebastian. 2003. 'The Flawed Logic of Democratic Peace Theory'. *American Political Science Review* **97**(4): 585–602.
Rosen, Gideon. 1994. 'What is Constructive Empiricism?' *Philosophical Studies* **74**(2): 143–78.
Ruben, David-Hillel. 1993a. 'Introduction'. In *Explanation*, edited by David-Hillel Ruben, 1–16. Oxford: Oxford University Press.
Ruben, David-Hillel, ed. 1993b. *Explanation*. Oxford: Oxford University Press.
Rubin, Donald B. 1974. 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies'. *Journal of Educational Psychology* **66**(5): 688–701.

Ruggie, John Gerard. 1998. 'What Makes the World Hang Together? Neo-Utilitarianism and the Social Constructivist Challenge'. *International Organization* **52**(4): 855–85.
Rummel, Rudolph. 1972. *The Dimensions of Nations*. London: Sage.
Russo, Federica and Jon Williamson. 2007. 'Interpreting Causality in the Health Sciences'. *International Studies in the Philosophy of Science* **21**(2): 157–70.
Sabaratnam, Meera. 2020. 'Is IR Theory White? Racialised Subject-Positioning in Three Canonical Texts'. *Millennium: Journal of International Studies* **49**(1): 3–31.
Salmon, Wesley C. 1993. 'Scientific Explanation and the Causal Structure of the World'. In *Explanation*, edited by David-Hillel Ruben, 88–112. Oxford: Oxford University Press.
Salmon, Wesley C. 1998. 'Scientific Explanation: Causation and Unification'. In Wesley C. Salmon (ed.), *Causality and Explanation*, 68–78. Oxford: Oxford University Press.
Sayer, Andrew. 1992. *Method in Social Science: A Realist Approach*, 2nd ed. London: Routledge.
Schaffer, Jonathan. 2005. 'Contrastive Causation'. *Philosophical Review* **114**(3): 297–328.
Scheler, Max. 2017. *The Nature of Sympathy*, with a new introduction by Graham McAleer. London: Routledge.
Schmidt, Brian. 1998. *The Political Discourse of Anarchy: A Disciplinary History of International Relations*. Albany: State University of New York Press.
Schroeder, Paul W. 1997. 'History and International Relations Theory: Not Use or Abuse, but Fit or Misfit'. *International Security* **22**(1): 64–74.
Schroeder, Paul W. 2001. 'International History: Why Historians Do It Differently than Political Scientists'. In *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*, edited by Colin Elman and Miriam Fendius Elman, 403–16. Cambridge, MA: MIT Press.
Schweder, Rebecca. 1999. 'Causal Explanation and Explanatory Selection'. *Synthese* **120**(1): 115–24.
Scriven, Michael. 1959. 'Truisms as the Grounds for Historical Explanations'. In *Theories of History*, edited by Patrick Gardiner, 443–75. Glencoe: The Free Press.
Seawright, Jason. 2002. 'Testing for Necessary and/or Sufficient Causation: Which Cases are Relevant?' *Political Analysis* 10(2): 178–93.
Seawright, Jason and David Collier. 2010. 'Glossary'. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 313–59. Lanham: Rowman & Littlefield.
Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont: Wadsworth Cengage.
Shapiro, Ian and Alexander Wendt. 1992. 'The Difference that Realism Makes: Social Science and the Politics of Consent'. *Politics & Society* **20**(2): 197–223.
Sil, Rudra and Peter J. Katzenstein. 2010. 'Analytical Eclecticism in the Study of World Politics: Reconfiguring Problems and Mechanisms across Research Traditions'. *Perspectives on Politics* **8**(2): 411–31.
Singer, J. David. 1961. 'The Level-of-Analysis Problem in International Relations'. *World Politics* **14**(1): 77–92.
Singer, J. David. 1979. 'The Historical Experiment as a Research Strategy in the Study of World Politics'. In *The Correlates of War*, vol. 1: *Research Origins and Rationale*, edited by J. David Singer, 175–96. New York: Free Press.
Skovgaard Poulsen, Lauge N. 2014. 'Bounded Rationality and the Diffusion of Modern Investment Treaties'. *International Studies Quarterly* **58**(1): 1–14.
Smith, Steve. 1995. 'The Self-Images of a Discipline: A Genealogy of International Relations Theory'. In *International Relations Theory Today*, edited by Ken Booth and Steve Smith, 1–37. University Park: Pennsylvania State University Press.
Smith, Steve. 1996. 'Positivism and Beyond'. In *International Theory: Positivism and Beyond*, edited by Steve Smith, Ken Booth, and Marysia Zalewski, 11–44. Cambridge: Cambridge University Press.
Smith, Steve. 2000. 'The Discipline of International Relations: Still an American Social Science?' *British Journal of Politics and International Relations* **2**(3): 374–402.

Smith, Thomas W. 1999. *History and International Relations*. London: Routledge.
Snow, John 1965. *On the Mode of Communication of Cholera*, 2nd ed. Originally published in 1855. London: John Churchill.
Sober, Elliott. 1984. 'Two Concepts of Cause'. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2: 405–24.
Sosa, Ernest ed. 1975a. *Causation and Conditionals*. Oxford University Press.
Sosa, Ernest. 1975b. 'Introduction'. In *Causation and Conditionals*, edited by Ernest Sosa, 1–14. London: Oxford University Press.
Stanford, Kyle. 2023. 'Underdetermination of Scientific Theory'. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/scientific-underdetermination/ (accessed 27 July 2023).
Stone, Lawrence. 1979. 'The Revival of Narrative: Reflections on a New Old History'. *Past & Present* **85**: 3–24.
Strawson, Galen. 2007. 'David Hume: Objects and Power'. In *The New Hume Debate*, edited by Rupert Read and Kenneth A. Richman, rev ed., 31–51. London: Routledge.
Stroud, Barry. 1977. *Hume*. London: Routledge.
Suganami, Hidemi. 1996. *On the Causes of War*. Oxford: Clarendon Press.
Suganami, Hidemi. 1999. 'Agents, Structures, Narratives'. *European Journal of International Relations* **5**(3): 365–86.
Suganami, Hidemi. 2002a. 'On Wendt's Philosophy: A Critique'. *Review of International Studies* **28**(1): 23–37.
Suganami, Hidemi. 2002b. 'Explaining War: Some Critical Observations'. *International Relations* **16**(3): 307–26.
Suganami, Hidemi. 2008. 'Narrative Explanation and International Relations: Back to Basics'. *Millennium: Journal of International Studies* **37**(2): 327–56.
Suganami, Hidemi. 2009. 'Understanding *Man, the State, and War*'. *International Relations* **23**(3): 372–88.
Suganami, Hidemi. 2011. 'Causal Explanation and Moral Judgement: Undividing a Division'. *Millennium: Journal of International Studies* **39**(3): 717–34.
Suganami, Hidemi. 2013. 'Causation-in-the-world: A Contribution to Meta-Theory of IR'. *Millennium: Journal of International Studies* **41**(3): 623–43.
Tannenwald, Nina. 1999. 'The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use'. *International Organization* **53**(3): 433–68.
Taylor, Daniel M. 1970. *Explanation and Meaning: An Introduction to Philosophy*. Cambridge: Cambridge University Press.
Taylor, Richard. 1975. 'Metaphysics of Causation'. In *Causation and Conditionals*, edited by Ernest Sosa, 39–43. Oxford: Oxford University Press.
Thistlethwaite, Donald L. and Donald T. Campbell. 1960. 'Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment'. *Journal of Educational Psychology* **51**(6): 309–17.
Thomas, Ward. 2001. *The Ethics of Destruction: Norms and Force in International Relations*. Ithaca: Cornell University Press.
Toulmin, Stephen E. 2003. *The Uses of Argument*, updated ed. Cambridge: Cambridge University Press.
Trachtenberg, Marc. 1990–1. 'The Meaning of Mobilization in 1914'. *International Security* **15**(3): 120–50.
Trachtenberg, Marc. 1991a. 'The Coming of the First World War'. In *History and Strategy*, 47–99. Princeton: Princeton University Press.
Trachtenberg, Marc. 1991b. 'Making Sense of the Nuclear Age'. In *History and Strategy*, 261–86. Princeton: Princeton University Press.
Trachtenberg, Marc. 2006. *The Craft of International History: A Guide to Method*. Princeton: Princeton University Press.
Van Evera, Stephen. 1984. 'The Cult of the Offensive and the Origins of the First World War'. *International Security* **9**(1): 58–107.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.
Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.
Van Fraassen, Bas C. 1985. 'Empiricism in the Philosophy of Science'. In *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, edited by Paul M. Churchland and Clifford A. Hooker, 245–308. Chicago: University of Chicago Press.
Van Fraassen, Bas C. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.
Van Fraassen, Bas C. 1993. 'Armstrong, Cartwright, and Earman on Laws and Symmetry'. *Philosophy and Phenomenological Research* **53**(2): 431–44.
Van Fraassen, Bas C. 1994. 'Gideon Rose on Constructive Empiricism'. *Philosophical Studies* **74**(2): 179–92.
Van Fraassen, Bas C. 2001. 'Constructive Empiricism Now'. *Philosophical Studies* **106**(1–2): 151–70.
Van Fraassen, Bas C. 2002. *The Empirical Stance*. New Haven: Yale University Press.
Van Fraassen, Bas C. 2007. 'From a View of Science to a New Empiricism'. In *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. van Fraassen*, edited by Bradley Monton, 337–83. Oxford: Oxford University Press.
Vitalis, Robert. 2015. *White World Order, Black Power Politics: The Birth of American International Relations*. Ithaca: Cornell University Press.
Von Wright, George Henrik. 1971. *Explanation and Understanding*. Ithaca: Cornell University Press.
Von Wright, G. H. 1975. 'On the Logic and Epistemology of the Causal Relation'. In *Causation and Conditionals*, edited by Ernest Sosa, 95–113. Oxford: Oxford University Press.
Waever, Ole. 2009. 'Waltz's Theory of Theory'. *International Relations* **23**(2): 201–22.
Waldner, David. 2007. 'Transforming Inferences into Explanations: Lessons from the Study of Mass Extinctions'. In *Theory and Evidence in Comparative Politics and International Relations*, edited by Richard Ned Lebow and Mark Irving Lichbach, 145–75. Basingstoke: Palgrave Macmillan.
Waltz, Kenneth N. 1959. *Man, the State, and War: A Theoretical Analysis*. New York: Columbia University Press.
Waltz, Kenneth N. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.
Webb, Keith. 1995. *An Introduction to Problems in the Philosophy of Social Science*. New York: Pinter.
Weber, Max. 1948. 'Politics as a Vocation'. In *From Max Weber: Essays in Sociology*, edited, with an introduction by H. H. Gerth and C. Wright Mills, 77–128. London: Routledge.
Wendt, Alexander E. 1987. 'The Agent-Structure Problem in International Relations Theory'. *International Organization* **41**(3): 335–70.
Wendt, Alexander. 1991. 'Bridging the Theory/Meta-Theory Gap in International Relations'. *Review of International Studies* **17**(4): 383–92.
Wendt, Alexander. 1998. 'On Constitution and Causation in International Relations'. *Review of International Studies* **24**(5): 101–18.
Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.
Wetzel, Linda. 2006. 'Types and Tokens'. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/types-tokens/ (accessed 28 Jan. 2022)
White, Hayden. 1973. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Baltimore: Johns Hopkins University Press.
White, Hayden. 1984. 'The Question of Narrative in Contemporary Historical Theory'. *History and Theory* **23**(1): 1–33.
Wight, Colin. 2002. 'Philosophy of Social Science and International Relations'. In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse, and Beth A. Simmons, 23–51. London: SAGE.
Wight, Colin. 2006a. *Agents, Structures and International Relations: Politics as Ontology*. Cambridge: Cambridge University Press.

Wight, Colin. 2006b. 'Realism, Science and Emancipation'. In *Realism, Philosophy and Social Science*, edited by Kathryn Dean, Jonathan Joseph, John Michael Roberts, and Colin Wight, 32–64. Houndmills: Palgrave Macmillan.

Wight, Colin. 2007. 'A Manifesto for Scientific Realism in IR: Assuming the Can-Opener Won't Work!' *Millennium: Journal of International Studies* **35**(2): 379–98.

Wight, Colin. 2013. 'The Dualistic Grounding of Monism: Science, Pluralism and Typological Truncation'. *Millennium: Journal of International Studies* **41**(2): 326–45.

Wight, Colin and Jonathan Joseph. 2010. 'Scientific Realism and International Relations'. In *Scientific Realism and International Relations*, edited by Jonathan Joseph and Colin Wight, 1–30. London: Palgrave Macmillan.

Williamson, Jon. 2009. 'Probabilistic Theories'. In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 185–212. Oxford: Oxford University Press.

Winch, Peter. 2008. *The Idea of a Social Science and its Relation to Philosophy*, 2nd ed. Abingdon: Routledge.

Wohlforth, William C. 1999. 'A Certain Idea of Science: How International Relations Theory Avoids the New Cold War History'. *Journal of Cold War Studies* **1**(2): 39–60.

Woodward, James. 1984. 'A Theory of Singular Causal Explanation'. *Erkenntnis* **21**(3): 231–62.

Woodward, James. 1993. 'A Theory of Singular Causal Explanation'. In *Explanation*, edited by David-Hillel Ruben, 246–74. Oxford: Oxford University Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wright, Austin L., Luke N. Condra, Jacob N. Shapiro, and Andrew C. Shaver. 2017. 'Civilian Abuse and Wartime Informing'. Working Paper. *The Pearson Institute Discussion Paper No. 42* https://thepearsoninstitute.org/sites/default/files/2017-07/07272017%20Wright_WartimeInforming.pdf (accessed 16 Aug. 2023).

Zaks, Sherry. 2017. 'Relationships Among Rivals (RAR): A Framework for Analyzing Contending Hypotheses in Process Tracing'. *Political Analysis* **25**(3): 344–62.

Zehfuss, Maja. 2003. 'Forget September 11th'. *Third World Quarterly* **24**(3): 513–28.

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.