THE DISPOSITIONAL THEORY OF CAUSATION

Fahad Zakaria Al-Witri

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016
The counterfactual approach to causation pioneered by David Lewis appears to be a degenerating research project. The dissertation develops an approach to the analysis of the concept of causation by other means. I argue that dispositions come first when analyzing causation. I develop a theory of causation as the ancestral of the dispositional stimulus-manifestation relation and show it to be adequate to handle counterexamples fatal to alternative theories.

The metaphysical theory of the causal relation is part of a package that includes pragmatic side theories of cognates of causation, such as causal explanation and moral responsibility. Certain problem cases in the causation literature, primarily cases of so-called causation by omission, prevention and double prevention, have long been thought to concern the metaphysical relation. I argue that they are best captured by theories of causal explanation and moral responsibility.

The dispositions-first package, while violating Humean strictures against primitive modality, is offered as an alternative to skepticism or eliminativism about causation.
Contents

List of Figures iii

1 Introduction 1
   1.1 Lewis .......................................................... 1
   1.2 Skepticism ..................................................... 8
   1.3 A Different Approach ........................................ 15

2 The Dispositional Theory of Causation 23
   2.1 Introduction .................................................. 23
   2.2 A framework for reduction .................................. 25
   2.3 The Manifestation Relation ................................ 29
   2.4 Preemption and the “In Response To” Relation ........... 38
   2.5 An Alternative Solution to the Problem of Trumping ....... 49
   2.6 A Causal Basis? ................................................ 52

3 Double Prevention 56
   3.1 Introduction .................................................. 56
   3.2 Is double prevention causation? ............................ 61
   3.3 Causation by disconnection? ................................. 66
   3.4 Causal Explanation .......................................... 72

4 Value and Responsibility 78
4.1 Introduction ................................................. 78
4.2 Does Responsibility Entail Causation? ..................... 80
4.3 Counterfactuals and Extrinsic Value ....................... 92
4.4 Responsibility without Causation by Omission .......... 104

5 Conclusion .............................................. 106

Bibliography ............................................ 111
List of Figures

1.1 Directed Acyclical Graph (DAG) of a causal model ............... 13
1.2 DAG of a variant model of fig. 1.1 with an intervention ........ 14

2.1 Manifestation-disposition relation and manifestation-stimulus (response) relation 29
2.2 Two dispositions, one of which lacks stimulus .................. 30
2.3 Two dispositions, one of which is masked ....................... 32
2.4 Forces, components and resultant ............................... 36
2.5 Competing dispositions with masking ............................ 38
2.6 Competing stimuli ............................................. 38
2.7 Regress of causal bases ....................................... 55

3.1 Ned Hall’s double prevention case .............................. 56
3.2 Bolt-action rile trigger mechanism ............................... 67
3.3 Trigger mechanism neuron diagram .............................. 67
3.4 At the high level of description: a stubborn neuron ............ 71
3.5 Inside the stubborn neuron C: the lower level ................. 71
Acknowledgements

I owe special thanks to Ralf Bader and John Collins. John Collins’s dispositions seminar at Columbia in the spring of 2011 and the subsequent metaphysics reading group that met during summers that followed were especially important in the development of this project.

I am grateful to Wolfgang Mann and Achille Varzi for their feedback and discussions; and to Yang Liu, Xing Zhao, Robby Finley, Mary Freeman, Christine Susienka, Usha Nathan, Jeremy Wolos, Alison Fernandes, Porter Williams, Simona Aimar, Haim Gaifman, Jeff Helzner, David Albert, and Arnold Koslow. Thanks also to the graduate school at Columbia for financial support, and to the Port Gamble S’Klallam Tribe Foundation for a philosophical residence of sorts at Heronswood Gardens, Washington in 2015 where the bulk of chapter 2 was completed. For their love and support, I am deeply grateful to my family: Clare, the Beuckes, my sister, and my mother.
For Clare
Chapter 1

INTRODUCTION

This chapter frames the dissertation as a defense of the metaphysical concept of causation against skeptical challenges. I argue that to resist skepticism, certain methodological constraints need to be relaxed. (1) We shouldn’t insist on a theory that reduces causation to Humean first principles. (2) We shouldn’t require the theory of metaphysical relation of causation to account for all intuitions about cases of causation; cases of “causation by omission,” and others that somehow involve absence in causation, can be understood in terms of cognate notions such as responsibility and explanation. The main chapters of the dissertation go on to defend a metaphysically sound view about causation within the framework presented here.

1.1 Lewis

In the 1973 paper “Causation,” Lewis wrote that, faced by a set of stubborn counterexamples, “the prospects look dark” for existing regularity analyses of causation (Lewis, 1973, 557). Lewis offered an alternative approach based on his own work on the semantics of the counterfactual conditional. There was reason for hope.

Nearly three decades later, in “Causation as Influence” (2000), Lewis, it seemed to some, had quietly given up on an analysis of the singular causal relation altogether. The lesson that might be learnt from the history of Lewis’s lifelong attempt to understand causation is that the project
of reducing the everyday notion of causation is deeply flawed.

It is worth reflecting on the differences and similarities between Lewis’s early, 1973 theory of causation and Mackie’s famous (1980) “INUS” analysis. Mackie’s theory is an example of the kind of flawed regularity theory that Lewis hoped to replace with a counterfactual theory of causation. In its schematic structure, Mackie’s theory held that a cause is something essential for deriving the effect. The idea is that there is causation between C and E where there is a true proposition (or the conjunction of a set of true propositions) F, such that C together with F entails E, but F alone without C does not entail E. For Mackie, the following pair of propositions are needed for causation:

\[ \square((C \land F) \rightarrow E) \]
\[ \neg \square(F \rightarrow E), \text{ i.e. } \Diamond(\neg C \land F \land \neg E) \]

For Lewis, the basis of causation is the notion of counterfactual dependence: if C had not occurred, then E would not have occurred. Dependence is a matter of the following pair of propositions:

\[ C \square \rightarrow E \]
\[ \neg C \square \rightarrow \neg E \]

It’s natural to see Lewis’s 1973 notion of dependence as offering a fix to Mackie’s attempt to approximate the notion of “difference making” causation. Instead of relying on a rigid notion of holding fixed a proposition F and varying the truth of C, Lewis’s use of the counterfactual allowed for the selection function and the notion of similarity to decide what to hold fixed.

Replacing the notion of INUS conditions with that of counterfactual dependence was a way that Lewis hoped to succeed where Mackie failed. It’s worth distinguishing this part of Lewis’s strategy, namely, getting right the basic notion of “difference making,” from another. For as well as disagreeing with Mackie about the basic building block of the causal relation, Lewis showed...
that it’s almost as important to consider carefully what causation is as a structure of dependence relations. Lewis’s 1973 theory famously argued that causation is the ancestral of the dependence relation. Causation isn’t exactly difference-making: it is rather chains of difference-making relations, pairs of events related by counterfactual dependence.

It will be helpful to refer by shorthand names to the two aspects to the theory just distinguished. Call the first task of understanding the basic relation the “A project”; and call the second task of understanding causation in terms of the basic notion the “B project.”

Under the purview of the B project are general questions such as whether causation is transitive, as Lewis in 1973 supposed; whether it is local, requiring spatiotemporal contiguity; and whether it is intrinsic (holding fixed the actual laws), such that either all duplicates of a pair (or chains) of events are causally related or none of them are.

Within the A project fall the more fundamental questions about the nature of the basic causal relation. Is it a matter of counterfactual dependence? Does it supervene on non-causal, non-dispositional facts about the world? What are the truth-conditions or grounds of a counterfactual conditional? Is there an objective similarity relation for correctly understanding causation? The aim of the A project is to reduce the basic, building-block causal notion to ontologically more basic facts.

In the 1973 paper, Lewis offers a B-type solution to one problem of preemption. An example of “early preemption” is the following. An assassin A1 shoots and kills some victim. But there is a back-up assassin, A2, who does not actually shoot the victim, but is ready to fire to ensure that the victim dies if A1 fails at his job. That is, if A1 had not shot the victim, A2 would have done so.

The backup A2 ensures that the death of the victim does not depend on its actual cause, namely, A1 shooting. The victim would have died even if A1 had not fired, since A2 would have killed him otherwise. But what this example shows, according to Lewis, is not that there is a problem for use of counterfactual dependence to analyze causation. Rather Lewis offers a B-type solution: the problem of “early” back-up preemption is to be solved by appealing to the fact that causation
is *transitive*, and relying on chains of dependence. This gives the right result since the victim’s death depends on his being shot by A1’s bullet. A1’s bullet depends on A1 firing his rifle and not on A2’s.

In the 1986 “Postscripts,” Lewis offers a B solution to another type of preemption, namely the case of “late preemption.” This is the classic case involving Billy and Suzy, each of whom throw a rock at a glass bottle. Suzy’s rock hits the bottle before Billy’s does. Suzy causes the bottle to shatter, but Billy does not.

Late preemption shows that not all cases of preemption turn on the existence of chains of dependence. Lewis argued that late preemption could be dealt with by exploiting the fact that the causal relation is *intrinsic* as well as transitive. Even though there is no chain of dependence in the actual world between Suzy’s throwing the rock and the bottle shattering, there is such a chain of dependence in a possible world where: (i) the series of events running from Suzy’s throw to the breaking occur exactly as they actually do, and (ii) the actual laws hold. This duplicate chain exhibits counterfactual dependence; therefore, the duplicate chain is a causal process; therefore, the actual chain (and indeed all duplicate chains of events) are also causal processes.2

In addition to the A and B projects, Lewis recognized a role for a third project. This further project we may call the “C project.” This third project aims to untangle the metaphysical notion of causation, given by the A and B theories, from related but derivative notions, such as causal explanation. Lewis (1986a) gives a theory of causal explanation, according to which, giving a causal explanation is communicating information about an event’s causal history. Lewis thereby distinguishes the metaphysical notion of the causal relation from the pragmatic notion of causal explanation.

Lewis wrote:

> this paper is not meant to rely on my views about the analysis of causation... Whatever causation may be, there are still causal histories, and what I shall say about causal explanation should still apply (Lewis, 1986a, 216).

---

2Incidentally, the intrinsnicness thesis also resolves issues about overdetermination where two causes acting simultaneously produce an effect. Example of overdetermination: two assassins fire bullets at exactly the same time, simultaneously killing their victim.
While it doesn’t rely on the details of Lewis’s views about dependence, the causal explanation theory is not unrelated. The C project should be seen as part of the package deal even if it does not directly address the question of what to say about the metaphysics of causation. It allows Lewis to defend the conception of the metaphysical relation of causation as a difference-making relation, as a relation between events rather than facts, and as a relation that is not context-sensitive or subject to pragmatic restrictions.

Lewis did not put too much weight on the notion of causal explanation in handling problematic counterexamples, but focused instead on B solutions. But I will make the case below (sec. 1.3) that the C project is an important and overlooked opportunity to make advances in responding to putative counterexamples.

Lewis’s B solutions to counterexamples, of the 1973 paper and the 1986 postscripts, keep the core part of his theory, the “A” part, as it is. That is, B solutions keep unchanged the idea that at the heart of the causal relation is the notion of counterfactual dependence. In Lewis (2000), a new theory of causation as influence is offered. Here, Lewis believed that the limits of B solutions have been reached and that a more fundamental revision to the core concept of counterfactual dependence is needed. Under the later causation-as-influence view, Lewis replaces the notion of counterfactual dependence with the notion of counterfactual covaration or, as he calls it, “influence.” In the 2000 theory, Lewis defends the view that causation is the ancestral of the relation of influence, keeping some of the B solutions while changing the heart of the A theory.

Recall that dependence was originally understood in terms of the pair of propositions $C \rightarrow E$ and $\neg C \rightarrow \neg E$. The not-C counterfactual considers the closest world where C does not occur. That is, the breaking of the bottle depends on Suzy throwing her rock if the bottle doesn’t break in the closest world where Suzy does not throw. Lewis now considers influence to obtain when there is a range of counterfactuals $C_1 \rightarrow E_1, C_2 \rightarrow E_2, \ldots$, where each event $C_i$ can be a minor “alteration” to $C$ with respect to its timing or the way it occurs. Thus, we now ask not just what would happen if Suzy doesn’t throw, but also what would happen if Suzy delayed her throw by a second, or if she throws the same rock with a greater force, or if she aims the rock at a slightly
different angle, etc.

None of the alternatives where Suzy throws the rock at a different time or in a different way would previously have been considered cases where it’s true to say that Suzy does not throw the rock, or that the event of Suzy’s throwing the rock does not occur. The reason is that such an event is not individuated in too modally fragile a manner. The event of Suzy’s throwing the rock tolerates minor alternations to timing and manner of throw. But the range of events $C_1, C_2, \ldots$ relevant for understanding influence are individuated more finely than would be ordinarily considered.

I have described this as a change in the A project. The suggestion was that, just as Lewis’s view of causal dependence was meant to replace Mackie’s notion of INUS condition, so perhaps influence is meant to fix the problems of dependence. But influence constitutes a bigger break than dependence did. Collins (2000, 114) voices this concern:

This new theory of causal influence amounts to a change of topic. Lewis is not offering a new answer to the old question: “what is it for this event to be a cause of that event?” He is rather, offering an answer to a quite new question: “what is it for this event to have a causal influence on that event?” Certainly one event can have a causal influence on another without being among its causes. An example is provided by the story of the “Poison and the Pudding” [(Lewis, 1986b, 198-199)]. If a poison kills its victim more slowly and painfully when taken on a full stomach, then the victim’s eating pudding before he drinks the poisoned potion has a causal influence on his death, since the time and manner of the death depend counterfactually on the eating of the pudding. Yet the eating of the pudding is not a cause of his death.

In “Causation as Influence,” Lewis takes himself to be answering the old question, the same one that Mackie was concerned with. But in his response to the case of the Poison and the Pudding, and to other cases of what he calls “spurious causation,” we find that Lewis has changed his views about causation more deeply. Lewis’s response is that the pudding delays and affects the manner of the death; therefore it is a cause of the death. This is counterintuitive. But according to Lewis, it’s counterintuitive not because it’s false that the pudding is a cause. While it’s not false to say that the pudding is a cause, according to Lewis it’s odd to say that it is given certain assumptions about our aims in communication. Not everything true is worth saying: Lewis proposes that “the pudding is a cause of death” is a case of a true proposition that is not worth asserting. According to
Lewis, we ignore the relatively minor influencers of an event and treat only those things that exert a greater influence as candidate salient causes. The pudding has a smaller degree of influence on the death compared with the poison, and so the causal role of the pudding is largely ignored.

Is it always true that things that have less influence are ignored while only greater influencers are assertable as “causes”? As it stands, Lewis’s explanation of the pragmatics of causal talk is inadequate. Consider a case of asymmetric overdetermination. Assassin A1 shoots the victim with an ordinary rifle. Ordinary rifles will fire anywhere you aim them. Assassin A2, however, shoots the victim with a “smart” rifle. A2’s bullet will somehow always find its way to the victim’s right lung. The two assassins fire at the same time, their bullets enter the victim at the same time, and each cause the death in their own right. Now, A1 has a greater range of influence on the death of the victim, depending where his bullet lands. By contrast there’s not much that A2 could’ve done differently to affect the manner or timing of the victim’s death. So, according to Lewis’s account, A2’s involvement may as well be ignored. A2, counterintuitively, is not a cause according to Lewis’s theory.

What kind of theory is Lewis’s influence view? Ultimately, I think we must agree with Collin’s view that it is not a theory of causation. The notion of influence is undiscriminating. Given what we know about physics, everything in the past light cone of an event is an influencer and hence a cause.

But causation is far more discriminating. On Lewis’s view, it is the pragmatics of causal talk that is supposed to do the major work of whittling down the set of influencers to the set of causes. It’s not the theory of influence that gives us the everyday notion of causation.

The causation-as-influence view represents a change in view regarding the basic character of the causal relation. The notion of causation that Lewis previously thought to be explaining through a theory based on dependence is not so fundamental. Fundamentally, what there is is influence between events. The Mackie/Lewis 1973 notion is derivative and, as Lewis seems to suggest, is determined by pragmatic considerations such as the relative degrees of influence amongst all influences of an event. Lewis, in other words, is suggesting that we can’t have a metaphysically
robust and objective notion of causation as previously understood.

1.2 Skepticism

The narrative of Lewis’s thought about causation, which moved towards giving up on reducing causation to fundamental facts, is mirrored in the larger state of the literature on causation. Summing up their exhaustive survey “Causation: A User’s Guide”, Paul and Hall (2013) write:

After surveying the literature in some depth, we conclude that, as yet, there is no reasonably successful reduction of the causal relation. And correspondingly, there is no reasonably successful conceptual analysis of a philosophical causal concept. No extant approach seems able to incorporate all of our desiderata for the causal relation, nor to capture the wide range of our causal judgments and applications of our causal concept. Barring a fundamental change in approach, the prospects of a relatively simple, elegant and intuitively attractive, unified theory of causation, whether ontological reduction or conceptual analysis, are dim (Paul and Hall, 2013, 249).³

The difficulties faced by the project of reduction may seem to vindicate those who have long argued that nothing corresponds to the philosophical concept of causation as a metaphysical relation between events in the world. A recent volume, Price and Corry (2007), collects a number of deflationary theories of causation inspired by Bertrand Russell’s famous claim that causation is “a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm” (Russell, 1912, 1).

What’s wrong with anti-metaphysical theories of causation? I will consider two interesting, substantive but skeptical positions represented in Price and Corry (2007). I will point to objections that I think get to the heart of the problems of deflationary accounts. The first representative skeptical position is the idea that causation is best thought to be a folk rather than a philosophical concept. The second is a subjectivism that holds that causation is dependent on agent’s beliefs.

³Paul and Hall are optimistic that “the pursuit of a reductive account of causation—both of its ontology and its concept—is a valuable enterprise” (Paul and Hall, 2013, 259).
“FOLK” CAUSATION

John Norton argues in “Causation as Folk Science” (Norton, 2003) that the concept of cause is like the concept of the caloric, a fluid substance that was once postulated to explain heat. It might be useful to think of heat as if there is caloric fluid. But the concept of caloric fluid is at best a crude approximation to the actual mechanisms that explain things like heat transfer. We know now that the thermodynamic behavior of physical systems that explain phenomena such as heat are best understood in terms of the aggregate interactions of many molecules and not in terms of the existence of some further material substance besides those molecules. The caloric theory of heat has been replaced by the kinetic theory.

The analogy with causation runs as follows. The appearance of causal interaction is generated by a microphysical or fundamental behavior that does not itself involve any causal interaction. There is no causation at the microphysical level. It’s just that when certain arrangements of physical systems obtain, in the aggregate, they can behave as if there is causation. It might then be more convenient to apply the terms “cause” and “effect” and to apply a “folk theory” of causation, since this is a simpler and more tractable way to describe things, than it is to apply the most precise descriptions of the pattern of events available.⁴

Consider the claim that there is no causation at the fundamental level: if we look at the level of physical particles (whatever they are), there is no causation and no cause and effect. To support the claim, Norton gives several examples where laws of determinism are violated, examples, for instance, where things happen apparently uncaused, where particles begin to move at an arbitrary moment in time without changes in the state of the system elsewhere.

Examples of violation of determinism are not the examples that can show that there is no causation at the fundamental level. So we don’t need to consider the details of the examples. We can grant Norton that there are classical violations of determinism. That there are physical

⁴That at least some motivation for the view comes from the failures of analysis is suggested by Norton, who writes: “That no correspondingly precise theory [of causation] is possible or, at least, presently available is implicit in the continuing proliferation of different accounts of causation in the literature” (Norton, 2003, 31).
sibilities where things begin to move from an inertial state without there being a candidate event that causes them to do so does not show that no fundamental event is ever caused by anything. Imagine a universe with just two particles. Particle 1 is stationary at time t when particle 2 collides into it. A short while later, at t', particle 1 begins to move. It seems at the very least possible that particle 1 caused particle 2 to move. Indeed it’s hard to deny that particle 2’s motion is the effect of its collision with particle 1.

This of course is not the end of the story. In response, the argument might be made that (1) the fundamental laws are exceptionless and perfectly uniform; (2) all and only those things mentioned by a fundamental law are fundamentally real; (3) there is no exceptionless regularity concerning causal interaction; therefore (4) there is no fundamental causation. Such arguments against the fundamentality of causation clearly turn on conceptions of what it is to be metaphysically or fundamentally real. These are contentious issues that are independent of the debate on causation. I would dispute, for instance, premise (2). Things can be fundamental without being mentioned by fundamental laws. But this transforms the issue of the reality of causation into another issue entirely, namely, what fundamentality amounts to.

Luckily, we don’t need to worry about that broader level of debate. For, if the idea is that there only appears to be causation when we are looking at large, medium-sized everyday objects, as opposed to fundamental particles interacting with each other, then this, it seems to me, is false. We have no trouble in describing two particles interacting with each other in causal terms. That is, we can apply the “folk theory” of causation even to the fundamental level.

This is not the case with regards to the concept of the caloric fluid. The concept of the caloric does not apply at the fundamental level because that concept requires the existence of a material substance that simply isn’t there. That is, the reason why the concept of the caloric is erroneous, or at best a folk theory, is that the concept is empty at the fundamental level: there is no object that the concept denotes. Once we have described the features of the physical system as consisting of the existence of, say, particle 1 and particle 2, there is no room then to say that there is, in addition, any caloric fluid.
By contrast, the relation of causation is not suddenly vacuous when applied to fundamental events as opposed to higher-level ones. Even if sometimes there are disputes as to whether some particular event $x$ caused another event $y$ at the fundamental level—disputes which can also arise at the level of tables and chairs—there is no problem in thinking about fundamental events as being causally related. It’s meaningful to ask, once we have described the events involving particle 1 and particle 2, whether the one event caused the other.

So what exactly is the idea of Norton’s claim that causation is merely a folk concept? At first sight, it seemed to turn on the view that causation only makes sense as an approximately true, but ultimately false theory describing large scale patterns in the behaviors of physical systems. But I claim causation can make sense in other domains, even at the level of microphysical particles. In other words, causal thinking is as general as we would expect a metaphysical concept to be.

Perhaps the aim of describing causation as merely a folk theory is to make a slightly different point. It might be the view that causal thinking is learnt or developed at the level of macroscopic things and insofar as we extend it to other domains, we are merely projecting habits of thinking in the area such habits were suited for to areas where such habits are not suited. But now, to argue that our causal thinking is indeed inapplicable at the fundamental level would require showing that the judgment, for instance, that particle 1 caused particle 2 to move, is misleading or inaccurate or false. We can agree with Norton and others that laws that predict patterns of causation, such as the law that nothing moves unless something moves it, can be misleading. But such laws are not invoked when we use the concept of causation to make the simple judgment “$x$ caused $y$.”

What exactly, one might wonder, is at stake? Suppose that Norton is right and that it is wrong to apply causation to some domains of events, such as microphysical events. So what? The implication is the restriction of the domains of many other crucial concepts. If I kill you with a gun that zaps particles at you, did the particles cause you to die? If not, then perhaps I’m not responsible for your death.

This isn’t a reductio. I merely point out that the folk theory of causation might have unintended costs to other areas of philosophy about which the folk theorist might have wished to
remain neutral. Causation is constitutive of many philosophical theories. For almost any important philosophical issue X, one can find a defense of a causal theory of X. There are causal theories of explanation, responsibility, perception, meaning, memory, etc.

SUBJECTIVE CAUSATION

An increasingly popular approach to analyzing causation is guided by the thought that a cause is a way of controlling or manipulating the effect. In “Causal Perspectivalism,” Huw Price (2007) argues that the only way to get an interventionist theory of causation off the ground is to adopt a subjectivist view of the causal relation. Subjectivism of this kind is a form of causal skepticism, even though it aims to give an analysis of causation. It is skeptical of the idea that causation is an objective, metaphysical relation.

The interventionist approach takes something like the following as a sufficient condition for causation. There is a causal relationship between C and E if E is correlated with and is not statistically independent of a possible intervention on C; that is, if the probability of E conditional on the intervention that affects C is not equal to the probability of E. Where I(C) is the event that is an intervention on C, the condition may be stated as: if P(E|I(C)) \neq P(E), then C causes E.

An intervention I(C) is an event; but how is the relationship of I(C) intervening on C to be defined? Following Pearl (2000), an important and common characteristic of an intervention for C is the idea that C’s occurrence is independent of everything except for the intervention I(C). That is, an intervention I(C) is such that C is statistically dependent on I(C) and not statistically dependent on anything else, other than the effects of C.

Ignore the immediate concern that intervention, when partly characterized in terms of the pattern of dependence on things except its effects, incurs a circularity in analyzing causation. Price argues that there is no pattern of dependence, where C is only dependent on I(C), if we appeal to objective probabilities to understand this kind of dependence. Price claims, “[i]nterpreted in an objectivist manner, then, interventionism makes the metaphysics of causation hostage to the possibility that there may be no causation, literally speaking, because there are no genuinely open
systems” (Price and Corry, 2007, 280-281). An event will usually be statistically dependent on many events. An adequate intervention will be hard to find.

Let’s begin by trying to see why we can’t use any actual probabilities to characterize interventions. Consider the following causal structure.

\[ Z' \rightarrow X \]
\[ \quad \quad \downarrow \]
\[ C \rightarrow E \]
\[ \quad \quad \uparrow \]
\[ Z \]
\[ \quad \quad \downarrow \]
\[ Y \rightarrow E' \]

\[ \text{Figure 1.1: Directed Acyclical Graph (DAG) of a causal model} \]

In this structure, C depends on X, Z and Z'. C also probabilistically depends on Y. The probability of C conditional on Y is greater than the probability of C conditional on the absence of Y.

Which of the available events can be considered an intervention on C? X is not an intervention on C because C does not depend on X alone. C also depends on Z, Y and Z'. By similar reasoning, Z, Y and Z' are not interventions on C either.

Perhaps the intervention is an aggregate event: all of the events C depends on taken together. However, now we have the problem that E', which is not an effect of C can turn out, on this interventionist theory of causation, to be an effect of C. If we take XYZZ' as the intervention on C, then intervening on C affects E'. Since \( P(E'|Y) \neq P(E') \), and Y is part of the intervention on C, it follows that intervening on C (which entails manipulating Y) results in changes to E'. This incorrectly yields the result that C causes the epiphenomenal event E'.

This is why Pearl and others see interventions as changes to the causal model itself. Interventions are meant to be possibilities in which the actual dependencies “break.” Pearl writes that an intervention “amounts to lifting [the target variable, e.g. C] from the influence of the old functional mechanism... and placing it under the influence of a new mechanism... while keeping all other mechanisms unperturbed” (Pearl, 2000, 70).
Rather than the above model, we are to consider some possibility (as in the model below) in which C depends only on a new event I. The actual relationships between C and X, Y, Z' and Z no longer hold. But the rest of the system is left as it is.

\[
Z' \rightarrow X
\]

\[
I \rightarrow C \rightarrow E
\]

\[
Y \rightarrow E'
\]

\[
Z
\]

**Figure 1.2:** DAG of a variant model of fig. 1.1 with an intervention.

The upshot seems to be that, instead of actual probabilities, interventionist theories appeal to counterfactual possibilities. Indeed, it’s hard to see why interventionism isn’t just a sophisticated version of a counterfactual theory of causation. For instance, suppose that the intervention model is a counterfactual possibility—the closest possible world in which some intervention I controls C. This would reduce the interventionist theories to a version of a counterfactual theory of causation that utilizes counterfactuals of the form: if C depends only on I, then E would depend on C. Hall (2007, 110) writes that “far from being indispensable, causal models merely provide a useful means for selectively representing aspects of an antecedently understood counterfactual structure.”

Price’s suggestion of appealing to subjective probabilities, instead of objective but counterfactual probabilities, is a way in which the interventionist approach to causation can be sharply distinguished from its counterfactual rivals. Price takes the probabilities that underlie the dependencies in the model to be the probabilities that an agent deliberating rationally ascribes to the events in question.

That is, within the first model above representing the actual dependencies, if we use agent probabilities, then E' is independent of C. The agent probabilities, unlike the actual objective probabilities, establish that \( P_A(E'|C) = P_A(E') \).

Interpreting the probabilities in this subjectivist manner is a suggestion that Price finds in the
writing of F. P. Ramsey. Ramsey writes:

This seems to me the root of the matter; that I cannot affect the past, is a way of saying something quite clearly true about my degrees of belief. Again from the situation when we are deliberating seems to me to arise the general difference of cause and effect (Ramsey, 1929, 146).

Price’s suggestion is that it is a constitutive feature of our agency and human action that, at the point of deliberation, we have fixed and settled beliefs about some events but not others.

Either interventions are in some sense counterfactual possibilities, in which case interventionism is a new counterfactual theory of causation; or interventions require, as Price suggests, agent probabilities. It’s interesting that for the interventionist framework to constitute a departure from the counterfactual theory, we are led to a subjectivism about causation.

A brand of interventionism that adopts agent probabilities does not give us a metaphysical and agent-independent notion of causation. The upshot of that is somewhat surprising. The cost of Price’s theory is a commitment to an evidentialist decision theory.

The argument [that agent probabilities take an action to be probabilistically independent of everything except their effects] turns on a defence of evidential decision theory against Newcomb-style objections. Indeed, I think the viability of the approach in general depends on this defence, for it is this that ensures that the probabilities in question need only be evidential, and hence not dependent on a prior modal notion (as invoked in causal decision theories) (Price, 2007, 281).

All things considered, it would be better to have a theory that is not tied to the mast of evidentialism. By contrast, a metaphysical account of causation would allow us to defend a causal decision theory. This illustrates the difficulties and costs incurred with non-metaphysical views about causation.

1.3 A Different Approach

As Paul and Hall suggest in the quote above, success in the analysis of causation will require a change in approach. Perhaps what has hobbled the project of analysis is that the desiderata for
a final theory of causation almost guarantee failure. Consider the range and broadness of the cases that have been submitted as counterexamples to various theories. These include cases of preemption, overdetermination, trumping, double prevention, causation by omission, cases that suggest intransitivity, not to mention cases of causation under indeterministic or probabilistic laws. It’s hard to see how a single metaphysical relation can be adequate to the full range of examples.

Another shackle is the emphasis on having a full and clean reduction of causation to an unmysterious and unproblematic base. It is thought that if a theory does not show that causation is reducible to the very parsimonious resources of, for instance, Lewis’s Humeanism, which ultimately allows only categorical properties (particular matters of fact) and the notion of naturalness as primitive concepts, then the central project has failed to address the concerns of skeptics.

We can relax each of these methodological assumptions. I’ll argue that we need to be discriminating in deciding which cases are compulsory as cases of causation and which can be set aside or treated in a different way. I’ll follow the strategy of disambiguating notions of causation, explanation and responsibility. Our intuitions about various cases of causation may be mistaking notions of explanation or responsibility for that of causation. I’ll then argue for a non-reductive analysis. I’ll point out that a non-reductive analysis can do more than has been previously advertised. In particular, a non-reductive analysis can counter skeptical challenges.

CLARITY ABOUT COGNATES OF CAUSATION

It’s usual to insist that any adequate analysis allow for causation by omission, prevention and double prevention. But cases where absences are thought to be causally implicated give rise to what Paul and Hall call “the question of uniform treatment.”

Given that we are making room for causation by omission in our ontology, a main topic of investigation is whether a philosophical account of causation should provide a uniform treatment of omission-involving causation and ordinary causation between events. Are omission-involving causation and ordinary causation just apples and apples, mixed together without incident? Or is omission-involving causation like
apples, while ordinary causation is like oranges, where mixing can cause trouble? (2013, 176).

In “Two Concepts of Causation,” Hall (2004) argues that uniform treatment is not possible. Hall argues that if we are to solve problems of overdetermination and preemption, the theses that causation is intrinsic, transitive and local are indispensable. But causation by double prevention and causation by omission is not intrinsic, transitive or local.\(^5\)

One immediate result is that counterfactual analyses are doomed to failure (unless, as I think, they are understood to be targeted narrowly at just one kind of event-causation): for they need the first three theses [i.e. intrinsicness, transitivity and locality] if they are to cope with the well-known counterexamples involving overdetermination, but they cannot abide these theses if they are to cope with the counterexamples involving double-prevention (or, for that matter, if they admit omissions as causes and effects) (Hall, 2004, 226).

So, Hall argued, there must be two concepts of causation and separate analyses for each concept. One concept—“production”—gets ordinary causation right. Production is intrinsic to its relata, is transitive, local, and does not admit of absences as relata. The other concept of causation is “dependence” and it involves simple counterfactual dependence. Things can depend on absences, or more precisely, on propositions of event non-occurrence, a kind of causation (dependence) by omission. It’s also possible for absences to depend on events, which is causation (dependence) by prevention.

Hall embraces a disjunctivism about causation: it is either production or dependence. But disjunctivism is a friend to skepticism. A skeptic might well point to Hall’s result to suggest that the concept of causation is a mutable folk concept, if not hopelessly confused. Norton (2003, 43), for instance, suggests as much. Norton writes that “the present literature in philosophy of causation shows considerable variation in views on what counts as causal. Therefore we do not have one unambiguous notion that must be generated by the reduction relations, but many possibilities.” This supports his overall skeptical position.

\(^5\)For the details, see Chapter 3, page 57.
One of the central claims of this dissertation is that there is no causation by omission or prevention: absences cannot be causes, causal intermediaries, or effects. Now, often an author will settle the question whether to treat such cases as genuine causation by appealing to a verdict of commonsense intuition. If it seems right, intuitive or commonly asserted that an event C caused E, then that is taken to show that we have a case of causation. One reason for claiming that there is causation by omission or prevention is precisely the appeal to intuitive cases where some omission is said to cause an outcome, or some event is said to prevent another from happening. But a more critical appraisal of such an appeal to intuition is necessary if we are to avoid slipping into skepticism about causation.

One strategy for rejecting particular intuitions is suggested in the introduction to Collins et al. (2004). Collins, Hall and Paul write that one way to reject an example presented by commonsense as a case of causation is to say:

The example is misleading, because it naturally draws our attention to other, related concepts; we therefore mistake the conditions appropriate for applying these other concepts with the conditions appropriate for applying the concept of causation (Collins et al., 2004, 34).

I adopt this strategy in chapters 3 and 4 to argue that in cases of omission, prevention and double prevention, commonsense, roughly speaking, is confusing causal idioms of causation on the one hand, and idioms of explanation and responsibility on the other. The suggestion that we confuse explanatory and causal idioms of “cause” has been made in the literature by Davidson (1967), Beebee (2004), and Varzi (2007).

In chapter 3, I will explore the suggestion in some detail with respect to double prevention. There I will show how counterfactual reasoning implicit in our understanding of double prevention is an instance of providing information about the causation history of the explanandum. In chapter 4, I extend the thought to the notion of responsibility. I will suggest that the notions of responsibility and causation come apart and, more precisely, that responsibility does not entail causation. Nonetheless, we sometimes use causal terminology to assert responsibility. I show
how it’s possible that there can be responsibility by omission and responsibility for some non-occurrence (prevention).

But one has to be careful in applying the strategy. It’s not enough to say that the case is confusing various causal idioms. We have already encountered in Lewis’s attempted pragmatic treatment of counterexamples to his theory of causation as influence how such a strategy can fail. Lewis claimed that certain intuitions that are inconsistent with his theory can be recovered by a context-sensitive pragmatic theory. Lewis however did not establish what that theory is, or show that a pragmatic treatment can adequately distinguish causes from non-causes, or make plausible the idea that there is context-sensitivity in cases of preemption. To employ the strategy convincingly, chapters 3 and 4 therefore apply substantive theories of explanation and responsibility and show that the theories can capture the commonsense intuitions about omission, prevention and double prevention.

What emerges is an approach to causation that utilizes the area of the “C” theory of causation. As well as (what I’ve called) a B inquiry into the structural properties of the causal relation, an analysis of causation requires attending to the further question of causal cognates—the C theory. Since intuitions and commonsense are indispensable to the basic methodology of the metaphysics of causation, investigation of the cognates of causation is as necessary to understanding causation as providing a basic theory of the nature and structure of the metaphysical relation itself.

NON-REDUCTIVIST ANALYSIS

Let us return now to the second methodological issue. Paul and Hall (2013, 25ff.) note that reductive analysis is not the only possible or interesting project in the metaphysics of causation. There are other projects that one might be engaged in when giving a biconditional of the form “C causes E iff .... .” Interventionists such as Woodward (2003) are engaged in non-reductive analysis, explaining one causal notion (i.e. the singular causal relation) in terms of another (i.e. the notion of causal intervention). Such a circular analysis is thought to be enlightening if one buys into the reality of causation. But one might reasonably think that it can’t reasonably make a
skeptic about causation change their mind. For the skeptic, it seems, it’s either Humean analysis or bust. I want to challenge this idea. Skeptics can be swayed by a non-reductive analysis.

A non-reductive analysis, while circular, can nonetheless be non-trivial and illuminating. This is argued, for instance, by (Woodward, 2003, 21). Suppose you thought that counterfactuals cannot be reduced to purely non-causal concepts and that counterfactuals, in the final analysis, are based on causal facts. Now, one can still defend an interesting counterfactual theory of causation. It’s not obvious, for instance, that causation is a chain of counterfactual dependence. That would be a non-trivial claim of the theory. Similarly, Lewis thought that the dependence between events C and E is to be understood in terms of a conditional whose antecedent is simply the proposition that C is false. But this has been challenged. Yablo (2000), for instance, suggests that the relevant conditional is not a not-C counterfactual. Rather, we have to hold fixed some true proposition F. For Yablo, E depends on C only if \( \neg C \land F \equiv \neg C \neg E \). This disagreement between Lewis’s and Yablo’s counterfactual theories of causation is a disagreement about which counterfactuals to use. Whether counterfactuals are fully reducible or not, we can still learn something about causation by considering which conditionals are relevant on the right hand side of the biconditional: “C causes E iff...”.

Still, the causal skeptic wants to know what causation ultimately is in terms of fundamental facts. They take the non-existence of a reductive analysis as evidence that causation isn’t grounded in fundamental reality. A skeptic would see no reason to think that causation is reducible even if we show them the variety of counterfactual theories on offer. But given a strong connection between two distinct concepts, it becomes difficult to deny the reality of either. The burden of proof is shifted. If, that is, we supply an analysis of the form X iff Y, where X and Y are distinct concepts but neither of which is perfectly unmysterious, we offer something important by way of philosophical understanding. A necessary connection between unrelated concepts is evidence of the existence of a unifying explanation in more fundamental terms.

Reduction and simplification to a parsimonious base is one way to give a philosophical explanation when faced by the question, “what is X?” or “what is causation?” Provided that what
is taken as the reductive base is in fact metaphysically fundamental, such a reduction would appear to show not merely the existence of the explanation, but to give the explanation directly. By contrast the mere existence of a biconditional doesn’t say what the explanation is. But it does tell us that there is one to be had. This is why non-reductive explanations can be anti-skeptical.

But these issues concern larger questions of philosophical methodology and meta-metaphysics. One might respond that a biconditional could be explained by connections in meaning at the level of language, not the world. It might be that a given necessary biconditional is merely to a conventional or stipulated truth. All and only bachelors are unmarried men: but what does that prove about the reality of bachelors or unmarried men? Fortunately, for our aims, we don’t need to respond to these questions. For, in all parts of philosophy there are attempts to give necessary and sufficient conditions to clarify concepts. It’s a general problem for everyone to show that not all biconditionals are trivial analytic descriptions of the sort “all and only bachelors are unmarried men.” A non-reductive analysis of causation of the form “C is a cause of E iff X, Y Z” that isn’t stipulative or merely conventional points to the existence of a fundamental explanation.

In chapter 2, I argue for an analysis of causation in terms of dispositionality. As I see it, the two concepts—cause and disposition—are further apart than cause and intervention. But, of course, skeptics might think that dispositionality is a modal notion and not itself a thing one might wish to accept as primitive. The thorough-going skeptic, in other words, wants an analysis that reduces causation to notions that she is not skeptical about. But even for such a skeptic, the mere fact of a biconditional connection should count for something, as I’ve argued. It’s not a knockdown argument that causality is fundamental, since the promise of an explanation is not as good as the provision of one. But it does enough to shift the burden of proof onto those who say that there is no way to make sense of the concept in terms of fundamental reality.

In addition to the grand scale advantages, I argue in the dissertation that the dispositional theory of causation offers some concrete insights into particular puzzles about causation. In particular, I argue in chapter 2 that the theory gives us a way to understand trumping preemption. In chapter 3, I show how it allows one to resolve puzzles about causation by disconnection, which
is traditionally thought to be an instance of double prevention. There, I show that applying the dispositional theory shows that in fact disconnection is different and can be regarded as causation pure and simple.

Non-reductivism is contrasted with reduction. It is also to be contrasted with anti-reduction. Anti-reductionists deny that causation can be reduced to any non-causal concepts. In the literature, notable voices of anti-reduction include Carroll (2009), Carroll (1994) and Armstrong (1999). They argue that causation is irreducible and, further, that causation does not need to be reduced in order to be admitted as an acceptable metaphysical notion. What is more familiar than the concept of causation? Thus Carroll (1994, 118), writes: “With regard to our total conceptual apparatus, causation is at the center of the center.” In that spirit, anti-reduction does not offer any biconditional of the form “C causes E iff....”

But the position is merely the Moorean flip side of the skeptical response to failed attempts at reduction. It’s hard to see how it will convince a skeptic. Indeed, anti-reductionists and skeptics learn different lessons from the failure of analysis. Carroll (2009, 287) writes:

A primary motivation for anti-reductionism is the repeated failures of reductive analyses, and rightfully so. That no successful non-causal analysis of causation exists ought to lead us to consider the possibility that there cannot be one.

Since it cannot be reduced, skeptics say causation is nothing; anti-reductionists say it is primitive.

Non-reduction takes the middle road. It gives a theory of causation. But it does not stake the value of analysis as a means for the reduction of causation to notions free of any metaphysical baggage. Rather, a non-reductive account aims at systematizing different aspects of our conceptual economy, tying loose concepts closer together. The project is worthwhile and it guides the dispositional theory of causation advanced in this dissertation.
2.1 Introduction

Dispositions are commonly understood in terms of counterfactual conditionals. On the simple version, an object has a disposition D to manifest M when subject to stimulus S iff S $\rightarrow$ M.\(^1\) According to the simple counterfactual theory, that the vase is fragile is to be understood in terms of the claim that if it were struck, it would break.

Unfortunately, the counterfactual analysis is susceptible to virulent counterexamples. Finks, masks, and mimicks show that no non-trivial counterfactual is sufficient or necessary for a dispositional ascription. These are cases in which realizing conditions for the manifestation of a disposition interferes with the disposition (in the case of finks) or the manifestation of the disposition (masks and mimicks).

- Fink-off: x has a disposition D to M when S but if S were to occur, x would lose its disposition D before the disposition manifests. The original example given in Martin (1994) involves an “electro-fink.” A live electric wire has the disposition to conduct electricity when touched by a conducting material. However, Martin’s electro-fink, a device connected to the live wire, acts in such a way that it immediately shuts off the current to the

---

\(^{1}\) Except where quoting other authors, throughout this chapter, upper case letters S, M, C or E are used as shorthand for type events. Thus “S occurs” means: “an event of type S occurs.” Lower case letters will be used to denote particular or token events of the corresponding type. Token events are individuated by property instance at a time. Thus “s occurs” means: “the particular event s occurs.”
wire when the live wire is in contact with a conductor. No current passes through the live wire if it were touched by a conducting material.

- Fink-on: x does not have a disposition D to M when S but if S were to occur, x would gain the disposition D in time to manifest. Martin’s electro-fink can work in reverse: a dead wire can be made live just when it is touched.

- Masking: x has a disposition D to M when S but D would not manifest when S because a mask interferes with and overrides the disposition; e.g. applying the car’s brakes is a way of masking its disposition to roll down a hill.

- Mimicking: x does not have a disposition D to M when S but D would manifest when S because of a mimick, an external factor responsible for the manifestation M occurring just when S occurs. Johnston (1992) gives the example of the watchful angel who is ready to intervene whenever a gold chalice is dropped. The chalice is not fragile. However, if dropped, the angel—the mimicking factor—would act to make the chalice shatter.

The counterexamples suggest that dispositionality ought to be regarded as a more fundamental concept, and philosophers have proposed using dispositions to account for a variety of metaphysical notions such as modality (Goodman, 1954; Vetter, 2015), laws of nature (Bird, 2007), and causation (Bird, 2010).

In this chapter, I will be concerned specifically with the prospects of a dispositional theory of causation. The dispositional theory of causation has not been systematically investigated. There are some common themes amongst those who see causation as reducing to dispositions, but no precise theory that can handle the standard problem cases has been developed in the causation literature. For instance, it has been suggested that the nature of dispositions to come in reciprocal pairs is important to understanding causation, as is the idea that dispositions can compose and combine. Many have identified the causal relation with the relation between the stimulus and manifestation of dispositions. These ideas all play a role in a final theory of causation. The purpose of this chapter is to provide a systematic view that makes clear in one package the roles of
reciprocity and compositionality. I argue that these factors allow us to handle cases of preemption and the deeply problematic cases of trumping preemption.

I will argue that the dispositionalist aims to give a reductive account of causation in the traditional sense through an analysis of causation in terms that do not presuppose causation (sec. 2.2). I will argue that causation does not reduce to facts about what dispositions are instantiated. We must further include in the reductive base facts about how dispositions compose (sec. 3-4). These relations of composition, and the facts about when a collection of dispositions gives rise to another, are, I suggest, primitive. The analysis I advance is therefore given in terms of dispositions and compositionality.

Having formulated a dispositional theory of causation, I will consider an important objection to the hopes of reduction, namely that dispositions must have causal bases that are causally relevant to their manifestation (sec. 2.6). The dispositionalist about causation must deny the causal basis thesis. I argue that the most plausible way of doing so requires appealing to a distinction between causes and conditions. I suggest that the base of a disposition is a condition for its manifestation and not a cause.

2.2 A framework for reduction

In general, the idea behind a dispositional theory of causation is to identify the cause-effect relation with the stimulus-manifestation relation of a disposition.

... the most obvious dispositional analysis of causation says that C causes E iff E is the manifestation of some disposition whose stimulus is C (Bird, 2007, 161).

There are however surprising differences in the implications of this view that are tied to foundational questions about the nature of dispositions. For instance, Charlie Martin thinks that once we see causation in terms of the triggering and manifestations of dispositions, we are pushed to the radical view that cause and effect are identical and not two distinct events.
The two-event cause-and-effect view is easily avoided and replaced by the view of mutual manifestation of reciprocal disposition partners... the reciprocal dispositional partnering and their mutually manifesting are identical... It is not a matter of two events, but of one and the same event—a reciprocal dispositional partnering as a mutual manifesting (Martin, 2007, 46).

We will return to the notion of reciprocal partnering. Concerning the identity of manifestation and stimulus, as I read Martin, it may be wrong to think that the stimulus condition first obtains and then the dispositions interact to manifest. When the hammer strikes the fragile vase, part of the striking event involves the fragility of the vase. It would not be a striking, strictly speaking, if one were to swipe the hammer at empty air or a viscous liquid. In order to achieve the striking, the hammer needs to interact with something appropriately disposed to resist it. The dispositions are already interacting in the obtaining of the stimulus. Manifestation and stimulus are therefore not to be thought of as separate states. Or so Martin claims.

I mention Martin to point out two things. First of all, regardless of the plausibility of Martin’s view that the stimulus and manifestation of dispositions are identical (or at least constitutively related), we can see that the antecedent theory about dispositionality can have significant implications for the resulting views about causation. A radical view about dispositionality like Martin’s can have radical implications for how we understand causation. In particular, how we think of the stimulus-manifestation relation will determine how to understand the cause-effect relation.

Secondly, Martin’s view is an example of an informative and non-trivial dispositional theory about causation. We move from a substantive view about dispositionality that is not formulated in causal terms to an interesting view about causation. This may be contrasted with a somewhat less informative dispositional theory. According to Alexander Bird, the causal relation is the stimulus-manifestation relation. The stimulus-manifestation relation, however, is in turn taken as primitive. Bird writes, “the relation ‘Mx is the manifestation of disposition Dx in response to stimulus Sx’ is ontologically basic and is not reducible” (Bird, 2010).

By taking the “in response to” relation as basic, Bird is building causation into the dispositional base. For, this “in response to” relation looks to be simply the relation of causation. One
reason to take manifestation relations as basic is that it avoids certain kinds of counterexamples to be addressed in the next sections. However, doing so undercuts the force and motivation of a dispositional theory of causation. This puts our aim to find a conceptual analysis of causation in terms of dispositions in conflict with a dispositionalist such as Bird who sees the project to be one of providing an “ontological reduction” or “metaphysical analysis.” Bird writes, “While the analysis in terms of dispositions provides no conceptual reduction, it does provide insight into the metaphysics of causation. We know (if the account is correct) what causation is—it is the stimulation and manifestation of a disposition” (Bird, 2010, 7-8).

It is worth mentioning the approach to a dispositional analysis of causation of Ellis (2000, 2001, 2014). Ellis’s view is best understood within the process theory of causation. The process theory aims to distinguish genuine causal process from non-causal processes and other “spatiotemporal junk” (as Philip Kitcher has put it). Causal processes, it has been suggested, are those capable of transmitting a mark or signal, or ones that involve the transfer of some physical property such as energy or momentum. The suggestion Ellis makes, as an alternative to these accounts, is that bona fide causal processes are just those that are manifestations of dispositions.

Consider refraction. For anything to be a case of refraction, it must occur in a certain way. It is not enough that it look like a case of refraction. It is not even enough that it obey the manifest laws of refraction... For anything to be a case of refraction, it has to be a display [i.e. manifestation] of refractivity (Ellis, 2001, 50).

Dispositionality serves to identify the class of processes that may be regarded as genuinely causal. Spatiotemporal junk by contrast is not the manifestation or display of any natural disposition; spatiotemporal junk does not “occur in the right way.”

Ellis’s account faces similar questions about building-in causation as does Bird’s. Distinguishing certain processes as those that “occur in the right way” presupposes the notion of occurring in the right way, i.e. of being a particular manifestation of a particular disposition. Facts about whether some event/process is a display or manifestation of its disposition are taken as basic.

Bird and Ellis’s theories of causation fall short of achieving an insightful reduction of causation to dispositions insofar as they presuppose as primitive a relation of “being a manifestation of” or
“occurring in response to.” What is problematic, to be clear, is not that they presuppose something as primitive without explaining what it is. As we shall see, our account will presuppose relations of composition between dispositions. Rather, the problem is specifically that the relations Bird and Ellis take as primitive seem to be distinctively causal relations.

The account of causation I offer avoids building causation into the basis of the analysis. I follow Bird in taking causation to be given in terms of the “in response to” relation, or more precisely, the relation of x manifesting its disposition D in response to a concrete event s. Since causation is transitive, causation is to be identified as the ancestral of the response relation. That is, causation allows for chains of stimulus manifestations of different dispositions, where the manifestation of one disposition is the stimulus of another. Thus the framework for our analysis of causation is the following biconditional:

\[ c \text{ causes } e \iff (i) \text{ there is a disposition } D \text{ to } E \text{ when } C, \text{ and where } e \text{ is the manifestation of } D; \text{ or there is a series of events } c, c_1, c_2, \ldots, c_k, e, \text{ each of which is a manifestation of a corresponding disposition in a series of dispositions } D_0 \text{ (to } C, D_1 \text{ (to } C_2 \text{ when } C_1), \ldots, D_k \text{ (to } E \text{ when } C_k). \]

The response relation holds between two events, the stimulus and the manifestation. This for example is the relation between the hammer striking the vase and the vase breaking. The manifestation relation holds between a disposition and the manifestation. This is the relation between the fragility of the vase and the vase breaking. In short, the breaking of the vase is the response to the hammer blow, and the manifestation of the vase’s fragility.

I want to suggest that the response relation is reducible to the manifestation relation. What it is for the vase to break in response to being struck is for its fragility to manifest when the stimulus occurs, i.e. when it is struck. As a first pass, I suggest that the response relation is reducible to the manifestation relation:

\[ m \text{ occurs in response to } s \iff (i) \text{ there is a disposition } D \text{ to } M \text{ when } S, (ii) \text{ s is a token of } S \text{ and (iii) } m \text{ is a manifestation of } M. \]

The relation between the stimulus s and the disposition D, represented by the dashed line in fig. 2.1, is a type/token relation and holds just when the event s is a token of the type of event.
that is a stimulus for the disposition D. The particular hammer striking may be said to satisfy the stimulus condition for fragility because it is the type of event that is associated with fragility.

It might be argued that the specification of a stimulus for some disposition is determined causally. That is, one might think that what makes something a stimulus is that it can cause the manifestation to arise. On our view, rather, the specification of the stimulus is part of the primitive notion of dispositionality and part of the reductive base. That is, what is being assumed as primitive in our theory are triples \( \langle D, S, M \rangle \) consisting of the unique dispositional identifier D (which may be the categorical base of the disposition), possible stimulus condition S, which can be disjunctive (i.e. dropping or striking), and a property M that constitutes the manifestation of the disposition.

In the next section I will turn to accounting for the manifestation relation that holds between a particular event m and a disposition to M. Under what conditions does the occurrence of an event m that is the type of event associated with a disposition D ensure that m is a manifestation of D and not some other disposition? Then I will return in sec. 2.4 to the response relation “m occurs in response to s.”

2.3 The Manifestation Relation

In this section, we shall progress through three attempts to analyze the manifestation relation in terms of the instantiation of properties or the occurrences of events alone before reaching a fourth and final analysis.

The first and simplest proposal is that a disposition to M manifests just when a token of M
occurs. That is:

1. m is a manifestation of D iff D is a disposition to M and m is a token of type M.

(1) faces a simple counterexample. It’s possible for there to be two distinct dispositions that have the same type of manifestation M. Moreover, it’s possible for the one but not the other of the two dispositions to actually manifest.

We may imagine something which is disposed to explode because it’s highly flammable and also because it’s radioactive. Let’s suppose that the resulting explosive event is the same type of event whether it occurs because of oxidation, or because of a nuclear chain reaction. Either this is because we are individuating events somewhat coarsely as an explosion rather than as an explosion-by-oxidation or nuclear-explosion. Or it may be that we are imagining a possible world, somewhat unlike ours, where the two processes would yield precisely the same event.

![Figure 2.2: Two dispositions, one of which lacks stimulus](image)

The natural suggestion to distinguish an event as a manifestation of D₁ from the manifestation of D₂ is to look to the stimulus conditions. The disposition that manifests is the one that has its conditions for manifestation actually satisfied.

2. m is a manifestation of disposition D to M when S iff m is of type M and there occurs an event s of type S.

One might worry that not all dispositions need stimuli in order to manifest; this analysis will appear too strong. An example of a stimulus-free disposition might be the property responsible for inertial motion. Objects are disposed to continue at their current acceleration provided there are no disturbances. The absence of disturbances, moreover, should not be considered as a stimulus.

30
But rather than treat such dispositions as dispositions without stimulus condition, we can treat them instead as dispositions with trivially-satisfied stimulus conditions.

There is also the possibility that any disposition can manifest, as a brute matter of fact, without its stimulus. Take a world where all that occurs is that a vase shatters. Such an occurrence may be consistent with the laws of nature of that world. It might be a world that starts off in a highly unlikely initial state that is nonetheless possible. The occurrence of M without the occurrence of S, for any disposition, may be nomologically possible. Such occurrences of M are brute and uncaused.

The possibility of brute events does not clearly refute (2). For it may be denied that the brute event is the manifestation of a disposition. In such cases, the occurrence of M, while of the correct type to be a display of the disposition, is not an instance of the disposition manifesting. The lonely and spontaneously shattering fragile vase is not to be construed as manifesting its fragility. In order for fragility to manifest, the conditions for the manifestation of the disposition—the occurrence of a stimulus—would need to be satisfied. In other words, in such cases, we can bite the bullet and agree with the present analysis.

Even putting the two foregoing cases aside, there is a counterexample to the sufficiency of (2). This example involves the masking of dispositions. Masking is commonplace. A fragile vase may be stuffed with packaging material to protect it during shipping. The packaging ensures that the vase would not break when impacted. It prevents impacts (stimulus) from bringing about a breaking (manifestation). Poison is disposed to kill when ingested. But taken with its antidote, ingesting poison is short of deadly. The antidote masks the deadliness of the poison.

Now, it is possible to have two dispositions for the same manifestation, as before, but where only one of them is being masked. I am disposed to die from ingesting poison A as well as from ingesting poison B. Suppose that I swallow both poisons and thereby satisfy the conditions for the manifestation of each poison’s lethal power. Suppose, moreover, I am given an antidote that works only to counteract poison A. The antidote is a mask for the disposition of A only. My resulting death would then be the manifestation of poison B alone. It would have been caused
by ingesting B and not A. However, the stimulus for A’s lethal powers (my ingesting A) and its manifestation (my death) both obtain. According to the analysis of manifestation at hand, poison A’s disposition to kill manifests, which is not the case since it is masked by the antidote.²

\[
\begin{align*}
\text{s}_1 & \rightarrow \text{D}_1 \\
& \quad \xrightarrow{} \text{m} \\
\text{s}_2 & \rightarrow \text{D}_2 \\
\text{Mask} &
\end{align*}
\]

**Figure 2.3:** Two dispositions, one of which is masked

Masked dispositions cannot manifest. This suggests that a disposition manifests just when its distinctive manifestation and stimulus obtain while the disposition is not masked. For, if masking is the only way by which dispositional manifestations can be suppressed, then a disposition manifests just when its stimulus obtains and it is not masked.³

3. m is a manifestation of disposition D to M when S iff (i) an event s of type S occurs, and (ii) D is not masked.

According to (3), the satisfaction of the stimulus and the absence of masking is sufficient for a disposition to manifest, and so also for the occurrence of the manifestation. Yet this may seem too strong. There may be dispositions which, when stimulated, manifest probabilistically. Such a disposition D manifests with a chance \( p < 1 \) when S. There is a non-zero chance that D fails to manifest even in ideal (non-masking) conditions when the stimulus obtains. Where a probabilistic disposition D fails to manifest when stimulated, it does not follow that D was masked.

---

² The case presented resembles cases of “veridical mimicking” in the literature. For instance, (Johnston, 1992, 233) gives the case where “x has the disposition to break when struck, but a deranged guardian angel has decided to break x when struck in a way that is independent of its fragility.” But in Johnston’s case, it is not made explicit whether the fragility of the vase is being masked or whether the breaking of the vase is somehow a manifestation of its fragility and the additional disposition triggered by the angel. If the fragility of the vase is not masked, then the shattering ought to be thought as overdetermined by the manifesting of fragility and the manifesting of whatever powers of the angel are responsible for shattering the vase.

³ We should note here that cases of finking, another well-known problem cases for dispositions, where the stimulus causes the loss of a disposition before the manifestation can occur are precluded by the requirement that the disposition is present.
Probabilistic cases ought to be put aside. Our focus will be restricted to deterministic dispositions and deterministic cases of causation. The difficulty here is that it is unclear whether to describe the disposition as one where S produces M with a probability p, or whether it is a disposition whose manifestation is that the change that M occurs is p. That is, it may be that the manifestation of D is the event of M occurring with probability p. In that case, D is best described as a disposition to manifest the state $\text{Prob}(M)=p$ rather than M. Thus, even if M fails to occur, it would be wrong to say that the disposition does not manifest. It \textit{does} manifest the chancy state that M occurs with probability p.

\textbf{MASKING: PREVENTION V COMPOSITION}

We need to address the question of how masking is to be understood and what it is for a disposition to be masked. One suggestion is that masking is the non-satisfaction of the stimulus condition. The stimulus condition for the poison, properly stated, is that it is ingested \textit{without its antidote} (Lewis, 1997, 153). This means that in the example I have given, we can say that poison A does not manifest because it is not stimulated since A is ingested with its antidote.

A problem with this proposal is that it makes the task of characterizing the stimulus condition for most dispositions intractable. One reason is that masks themselves are susceptible to masking, preventing them from doing the work of interfering with the manifestation of disposition. I may ingest poison A with its antidote but also with a mask for its antidote. The antidote’s mask prevents the antidote from counteracting the poison. So at the very least it seems that the stimulus condition for poison should be that either the poison is ingested without its antidote or that it is ingested with its antidote and the mask for that antidote. It is easy to see that stimulus conditions for dispositions, in order to account for all possible masks, becomes highly disjunctive and unwieldy.

Another proposal is to understand masking in terms of causation. Johnston (1992, 233) suggests that the masking of a disposition to M be understood as a case where something causes an event that is incompatible with M occurring. This however is too strong. A disposition may be masked even where it remains possible for the manifestation to occur. As we have seen, we can
mask and prevent poison A from causing death while leaving poison B unmasked. In that case, the manifestation event characteristic of poison A occurs and is not precluded by the fact that there is a mask acting on poison A.

Revising a little, we can say that a mask is something that prevents the stimulus from causing a disposition to manifest. While preventing one stimulus from causing the manifestation, a mask may thus allow a different stimulus to cause the manifestation and thereby allow the manifestation to occur. That is, on the causal understanding of masks, a mask for a disposition to M when S is something that prevents S from bringing about M. It is not something that prevents M from occurring altogether.

(M1) \( x \) masks \( D \) iff \( D \) does not manifest and \( x \) prevents stimulus \( s \) from causing \( m \)

(M2) \( D \) is masked iff if \( s \), then something would prevent \( s \) from causing \( m \)

This proposal is incompatible with our reductive aims. In the first place, prevention is causation and masks are preventers. In the second place, on the revised view, masks prevent the very relation that we are attempting to analyze, the relation between stimulus and manifestation. The challenge is to give a way to understand masking in a non-causal way.

Johnston’s approach of understanding masking in terms of prevention is unavailable to us. But the idea that masks somehow cancel the manifestation of a disposition might be understood in terms of non-causal canceling. The kind of non-causal canceling and the structure of the solution that we require can be found by considering Newtonian forces.

Consider a simple case of billiard ball mechanics. A ball is disposed to roll to the right when pushed in that direction. But whether the ball does in fact move in that direction depends on other forces present. The ball won’t move to the right if, for instance, it is also pushed to the left. The force acting to the left, in that case, combines with it a contrary force pushing to the right. The relationship between the forces themselves is one of opposition and canceling that isn’t exactly one of prevention. It’s true that the rightward force prevents a leftward motion. But that is explained by the more fundamental fact that the rightward force opposes the leftward force.
The notion of “real opposition” is discussed by Kant in the 1763 essay “Attempt to introduce the concept of negative magnitudes into philosophy.” Kant begins that essay by contrasting real with logical opposition.

Two things are opposed to each other if one thing cancels [aufhebt] that which is posited by the other. The opposition is two-fold: it is either logical through contradiction, or it is real, that is to say, without contradiction.... The motive force of a body in one direction and an equal tendency of the same body in the opposite direction do not contradict each other... It is, nonetheless, a true opposition (2:171/211).

Kant’s view suggests a non-causal notion of opposition that is primitive and more basic than prevention. It also suggests that such opposition is grounded in facts about composition that are more basic than canceling. Forces compose in such a way as governed by mathematics of vector addition. It is the rules of force composition, codified by the vector addition, that explains why two equal but opposing forces cancel out and give a net zero force resultant.

An analogous structure of property composition can be applied to dispositions and can be leveraged to understanding masking. Just as two forces can oppose each other and compose to give rise to a resultant, so two dispositions can oppose each other and give rise to a resultant disposition. The one disposition can be said to be masked by its opposing disposition.

We do not need to think of forces as dispositions in order to appeal to talk about composition in the case of dispositions. Rather, it is that the relation that is exhibited by forces when we apply the rules of vector addition allows us to understand how the dispositions which give rise to the forces in question can compose.

Consider an object with two ropes attached to it. The object has a disposition to accelerate in the direction of force applied to it as a result of any rope being pulled. It has a disposition to accelerate to the left when pulled leftward. It also has a disposition to accelerate to the right when pulled rightward.

Now suppose that the object is pulled simultaneously in opposing directions with equal magnitude of force. We may describe the case as one in which the leftward pull prevents the manifes-

---

4 Thanks to Ralf Bader for pointing me to the essay.
tation of disposition of the object to accelerate to the right when pulled in that direction, and vice versa. But we may also describe the case in terms of the fact that joint stimulation of the left- and right-acting dispositions compose such that they jointly manifest a zero acceleration.

There are reasons to prefer the compositional description as more fundamental. In the first place, composition allows us to capture the intuitive fact that the dispositions are jointly manifesting as components of a larger dispositional whole. By contrast, the masking story would suggest that neither disposition is doing any work, so to speak, since the manifestation of each disposition has been prevented. In fact, all that is prevented are the events of the object moving left and the object moving right but not that the dispositions in question manifest.

Secondly, the compositional story allows us to make sense of partial masking in a way that unifies it with the cases of full canceling. Suppose for instance that the object is pulled by one rope directly north, applying a force of 4 newtons, and another rope pulled directly east, applying a force of 3 newtons. Then the object accelerates at 5 ms\(^{-2}\) in the direction 30 degrees east of north. According to the causal story, the disposition to accelerate east is masked. But we should distinguish the case from one in which there is no net motion whatsoever. For we might say that while the disposition to move east is masked, it is not completely masked. The compositional picture therefore allows for a more fine-grained understanding of masking that the Johnstonian causal theory in terms of event prevention.

\[
\begin{array}{c|c}
R \text{ is masked} & R \text{ is not masked} \\
\hline
\text{Components} & L \rightarrow R & R \\
\text{Net force} & & \\
\end{array}
\]

**Figure 2.4:** Forces, components and resultant

The billiard ball case provides a model for understanding masking in the case of the fragile vase stuffed with packing material, even though here matters are not as simple as the composition of individual forces. As Johnston explains the details:
Packing companies know that the breaking of fragile glass cups involves three stages: first a few bonds break, then the cup deforms and then many bonds break, thereby shattering the cup. They find a support which when placed inside the glass cup prevents deformation so that the glass would not break when struck.

The vase, like the billiard ball, has competing dispositions. On the one hand, it is disposed to deform in the way that leads to shattering. On the other hand, it is disposed to retain its shape when subjected to many little forces in the way exerted by the packaging. In other words, the vase is flexible as well as fragile. When the vase is struck and supported by the packaging, both dispositions are activated. It is the resultant of the composite of the two dispositions that manifests.

According to our example, a mask is something that stimulates a component disposition that counteracts another. The billiard ball’s disposition to move to the right is masked by the leftward push. The vase’s disposition to shatter is masked by the packing material, since the packing material activates the vase’s disposition to regain its shape.

Masking understood in terms of composition is revealed to be a species of a more general phenomena. Masking occurs when the components and the resultants disagree in their characteristic manifestation. The component that is masked would behave differently if it were acting alone rather than as a component part with other dispositions in play. Thus the more fundamental property of a manifesting disposition is not that it is a disposition that lacks masks, but that it is an overall disposition. The suggestion, more precisely, is that manifestation is only possible and reserved for maximal dispositions, where a maximal disposition is one that is not a proper component of any other.

4. m is a manifestation of the disposition D to M when S iff (i) an event of type S occurs, (ii) m occurs and is an event of type M, and (iii) there is no disposition D’ to M when S’, such that D is a proper component of D’ and an event of type S’ occurs.
2.4 Preemption and the “In Response To” Relation

We provided an analysis of what it is for an event $m$ to be a manifestation of a disposition $D$. It is time now to give an account of what it is for $m$ to occur in response to some stimulus event $s$. The suggestion sketched in section 2.2 was that $m$ occurs in response to $s$ just when $m$ is the manifestation of a disposition $D$ to $M$ when $S$ and $s$ is an event of type $S$. Unfortunately, the possibility of preemption between different tokens of $S$ causes trouble for the account. It’s possible for an event of type $S$ to be preempted by an event $s'$ that is also of type $S$. In that case, $m$ does not manifest in response to $s$, but rather in response to $s'$. Being an event that satisfies the characteristic manifestation conditions for a disposition that manifests is not sufficient for that event to be the actual stimulus.

\[ S_1 \xrightarrow{D_1} M_1 \]

\[ S_2 \xrightarrow{D_2} M_2 \]

\[ \text{Masking} \]

\[ S_1 \xrightarrow{D} M \]

\[ S_2 \]

\[ \text{Figure 2.5: Competing dispositions with masking} \]

\[ \text{Figure 2.6: Competing stimuli} \]

It should be noted that preemption between different stimuli for one and the same disposition does not constitute a counterexample to the analysis of what it is for a disposition to manifest given by (4). In the case where $s$, a stimulus appropriate for $D$, is preempted by another stimulus $s'$, also appropriate for $D$, the disposition manifests. According to (4) it is necessary for a disposition $D$ to manifest that some stimulus for $D$ obtains. In cases of preemption, there are several competing events each of which is of $D$’s stimulus type.
Our account faces a more serious preemption problem. This is the problem of trumping preemption given by Schaffer (2000b). Schaffer (2000b, 165) asks us to imagine the law of magic that holds that “the first spell cast on a given day match the enchantment that midnight.” Consider a world in which such a law obtains and which is inhabited by the wizards Merlin and Morgana, and a Prince. Now imagine the following events. Merlin casts a spell at 12pm that would transform the Prince into a frog. Merlin’s spell is the first cast that day. Later that day at 6pm, Morgana casts exactly the same Prince-to-frog spell at 6pm. At midnight, the Prince is turned into a frog. Given the laws of magic, it seems we are forced to say that Merlin but not Morgana causes the event at midnight. Merlin’s spell trumpingly preempts Morgana’s.

Before we return to trumping, let’s consider a version of preemption that I think can be handled by our account. This is the case of late preemption. In the standard example of late preemption, Billy and Suzy each throw a rock at window but Suzy’s rock hits and destroys the window before Billy’s does. Late preemption is a problem for Lewis’s 1973 counterfactual theory of causation, as is well known, since the breaking of the window depends neither on Billy throwing his rock nor on Suzy throwing hers and, moreover, there is no chain of counterfactual dependence running between the breaking of the window and either of the two throws.⁵ According to our dispositional account, causation is the ancestral of the response relation. It can be seen that the problem with late preemption from a dispositionalist perspective is that the chain of stimulus and manifestation running from Billy’s throw to the shattering of the window is broken. The window ultimately shatters as a result of manifesting its fragility, which requires that the window be struck in order to manifest. Billy’s rock, however, does not strike the window. The disposition of Billy’s rock to strike the window when thrown therefore does not manifest. The absence of this dispositional manifestation is what blocks the chain of stimulus and manifestation necessary for causation in this case.

⁵For there to be such a chain, there must be at least one intermediate event D between the throw and the breaking of the window such that the window’s breaking depends on D. Consider for instance Suzy’s throw: at any point along the trajectory of her rock, had the rock been absent from that point or deviated sufficiently from the trajectory, Billy’s rock would hit and broken the window. So, the window would have broken had D failed to occur, for any D.
Lewis (1986b), in the Postscripts to “Causation,” adopted the principle that causation is intrinsic to processes of events in order to handle cases of late preemption. Roughly, according to his thesis, if there is a chain of counterfactual dependence amongst any duplicate of actual events, then the actual events are causally related. By intrinsicness, all duplicate pairs of event must agree whether they are causally related or not. The problem, however, is that trumping preemption shows precisely that causation is not intrinsic in this way. Trumping, in other words, is a counterexample to the intrinsicness of causation to events. The pair of events consisting of Morgana’s spell and the Prince turning to a frog are actually not causally related since Morgana’s spell is trumped. But there is an exact duplicate of the pair of events that does involve causation, viz. where Morgana’s spell happens to be the first spell of the day. This means that no theory that assumes intrinsicness of the relation to the events can handle trumping. Lewis’s solution to late preemption provides no adequate solution to trumping. The dispositional theory is in a better position than Lewis’s 1973 counterfactual theory with respect to late preemption. The theory hold that causation is transitive, but does not hold that it is intrinsic to events.

Let’s return to trumping preemption and see why it causes a problem for our account. For simplicity, let’s suppose that there is only one spell known to Merlin and Morgana that transforms the Prince into a frog. Let’s grant, further, that if Merlin casts his spell earlier than Morgana, then he and not Morgana would be considered the cause of the Prince’s transformation into a frog at midnight. The reverse is the case if Morgana casts the Prince-to-frog spell before Merlin casts the same spell. The earlier spell preempts the later spell.

Whether the Prince turns into a frog because of Merlin or Morgana, it’s plausible that the Prince’s transformation obtains in virtue of the same disposition. Let’s identify the disposition of the Prince to turn into a frog as D*. If Merlin trumps Morgana, then Merlin’s spell is a stimulus of D*. Similarly, if Morgana trumps Merlin, her spell would then be a stimulus of the same disposition D*. The stimulus condition of D*, in other words, can be satisfied by either Merlin or Morgana.

S1: the event type of Merlin casting a frog spell
S2: the event type of Morgana casting a frog spell

D*: the disposition of the Prince to turn into a frog when either S1 or S2

Now suppose that the events that are actually tokened are the following:

\[ s = (S1, 12\text{noon}) \]
\[ s' = (S2, 6\text{pm}) \]

That is, Merlin casts the frog spell at noon while Morgana casts the same spell later in the day at 6pm. Then we should say, given the laws of magic, that the enfrogment that occurs at midnight occurs in response to \( s \) and not \( s' \). According to our theory, however, the enfrogment does occur in response to \( s' \). This is because \( s' \) satisfies the stimulus condition of \( D^* \) and the disposition \( D^* \) does in fact manifest when the Prince becomes a frog.

Now, there is a temptation to complicate the characterization of the stimulus condition of \( D^* \). Rather than manifesting when either \( S1 \) or \( S2 \), the suggestion goes that we should characterize the disposition \( D^* \) as manifesting when either \( S1 \) or \( \text{not-}S1 \) and \( S2 \). In that case, \( s' \) would not satisfy the stimulus condition of \( D^* \).

But now suppose that Morgana acts before Merlin does. Suppose, in other words, that the following events occur:

\[ s = (S1, 6\text{pm}) \]
\[ s' = (S2, 12\text{noon}) \]

In this scenario, \( s' \) trumps \( s \). For that to be the case, the relevant disposition of the Prince ought to be one that would be manifested when either \( S2 \) or \( \text{not-}S2 \) and \( S1 \). Call this disposition \( D^{**} \). The event \( s' \) satisfies \( D^{**} \); the manifestation of \( D^{**} \) occurs in response to \( s' \).

Now, merely changing the chronological order of \( S1 \) and \( S2 \) should not affect the intrinsic character of the Prince. Whether \( S1 \) happens before \( S2 \) or \( S2 \) happens before \( S1 \), the Prince’s dispositions should be the same in both situations. Since the Prince has \( D^* \) when Merlin acts first,
he will still have \( D^* \) when Morgana happens to act first. But in that case, \( s \) will be satisfy the stimulus of \( D^* \). So it seems that the event of the Prince turning into a frog would be overdetermined. Both \( s \) and \( s' \) will count as causes of the enfrogment. This is contrary to the laws of magic that state that only the first spell of the day is effective.

Complicating the dispositions of the Prince is an unpromising strategy. The solution that I want to propose appeals to facts about the disposition of the Prince \( D^* \) and the magical powers of each wizard:

\[
\begin{align*}
D1: & \text{ Merlin’s disposition to a cast a Prince-to-frog spell} \\
D2: & \text{ Morgana’s disposition to cast a Prince-to-frog spell}
\end{align*}
\]

I want to suggest that the three dispositions combine together to give rise to an aggregate disposition \((D1,D2,D^*)\). This disposition has two key features. It has the same manifestation and stimulus condition as \( D^* \). That is, when \((D1,D2,D^*)\) is stimulated by the occurrence of a Prince-to-frog spell, the Prince is turned into a frog at midnight. Like many dispositions, once the disposition manifests, it ceases to be receptive to further stimuli. Once the window is broken, for example, the window ceases to be fragile. So, once \((D1,D2,D^*)\) has been stimulated, the disposition ceases to exist.

The manifestation of the aggregate disposition occurs at midnight. But this disposition ceases to exist just after 12pm, when Merlin completes intoning the magical words. That’s why when Morgana casts her spell at 6pm, there’s no disposition for her spell to activate. But isn’t there a temporal confusion here if the disposition ceases to exist before it manifests?

It is strange. But this is a strangeness built into Schaffer’s trumping example. His example involves action at a temporal distance. We are asked to suppose that spell castings can bring about their effects hours later without intermediaries. Why not then suppose that a disposition at \( t_1 \) can manifest an event at \( t_2 \) without any intermediaries?

As we saw when handling masking, we said that the salient disposition must be the maximal disposition. It was a condition for \( m \) to be a manifestation of \( D \) that \( D \) is not itself part of any larger,
active disposition. I shall now suggest a similar maneuver. I want to suggest that the disposition \( D' \) that is the Prince’s disposition to turn into a frog is in a sense incomplete and partial. There are further dispositions that must be taken into account. These are the reciprocal dispositions that are responsible for the stimulus.

RECIROCITY

C. B. Martin gives the following examples of reciprocal dispositional partners:

- “the soluble salt and the solvent water for salt in solution in the water” (Martin, 2007, 3)
- “The dispositional partners of key and lock” (Martin, 2007, 46)
- “the manifestation of the dispositional state of the match needs, amongst others, the cooperation of the reciprocal dispositional state of the enfolding oxygen” (Martin, 1996, 63).

The phenomenon is quite general. Whenever there is a disposition to \( M \) when \( S \), there is also a power that ensures that \( S \) can indeed cause \( M \). Salt can only be said to have the disposition to dissolve in water if water has the power to cause salt to dissolve in water. Moreover, the reverse is also true: if water is to have its power to dissolve salt, salt must be soluble.

Described in terms of powers, we seem to be invoking causation. The manifestation of a power is a causal relation. The power of water, for instance, is to cause salt to dissolve. The power of the key is to cause the lock to open.\(^6\)

Many have treated powers as dispositions by a different name. On this view, when \( x \) and \( y \) have reciprocating dispositions, they each have a disposition to \( M \). Even if we eschew the natural understanding of powers in terms of causation, we run into trouble with the idea that reciprocating dispositions are each disposition for the same manifestation. More precisely, the problem raised is one of overdetermination. When salt dissolves, this event would appear to be simultaneously

\(^6\) For a causal account of the notion of a power, see Bader (2015a, 7-11). According to Bader, a power of \( x \) to cause \( E \) in \( y \) is a disposition of \( x \) whose manifestation can serve as the stimulus for a disposition of \( y \) to \( E \).
the manifestation of a disposition of the salt and the manifestation of the reciprocal power of the water.

I suggest that we can make sense of the notion of reciprocity, alternatively, in terms of composition. If D and D' are reciprocal dispositions to manifest M, then M is the manifestation only of the resultant of D and D'. The event of the salt dissolving in water is not the manifestation of either the disposition of the salt or the water alone. Rather, it is the manifestation of the disposition of the salt and water together. When we ascribe to salt the disposition to dissolve when put in water or to water the power to dissolve salt, we are put things misleadingly. Rather than possessing a disposition to dissolve when put in water, salt possesses a component of a disposition for salt to dissolve when put in water. Rather than possess the power to dissolve salt, water possess a component of that same disposition for salt to dissolve when put in water.

Reciprocity is not a fundamental feature of reality. We sometimes individuate a disposition by the stimulus and manifestation conditions of the resultant disposition of which it is only a component and not the whole. We say that it has the disposition, when really it has only a component of that disposition.

This idea that dispositions can be individuated merely in terms of their contribution to a resultant may seem to generate problems. Firstly, if both water and salt have components of the same disposition, then what distinguishes their respective dispositions? Surely the disposition of the salt is not the same thing as the disposition of the water. In that case, being a component of some disposition is not sufficient to uniquely individuate a disposition. All components have the property of being part of their resultant. Secondly, if, as it turns out, the disposition of salt is not a disposition to dissolve when put in water, but only a component of that disposition, then perhaps the resultant disposition itself is in turn merely a component of some other disposition. If every disposition is individuated by its resultant, then no disposition can be individuated at all, except perhaps the total disposition of the whole world.

We can resolve these problems in turn. Firstly, while we cannot distinguish the disposition of the salt from the disposition of the water merely in terms of their role as components of the larger
disposition, we can distinguish these in other ways. It may be, for instance, that the disposition of water is distinguished from the disposition of the salt in that, while both can be components of the disposition to dissolve salt, the two dispositions differ with respect to how they can compose with other disposition. For instance, it may be that the disposition of water can also be a component of dissolving sugar, or other compounds. The solubility of sugar clearly cannot be constituted by any dispositions of the salt.

Alternatively, if we accept that dispositions have categorical bases, then these categorical properties can serve to individuate dispositions. The individual dispositions of water and salt are to then be distinguished not in terms stimulus and manifestation, but in terms of their role as components and their categorical bases. Since water and salt have different categorical properties, their dispositions can be distinguished.

Let us turn now to the second problem of individuation. Only the total, world-disposition is to be individuated by its role as a component. Here, too, there are resources that can be recruited to identify which dispositions can be individuated by their stimulus and manifestation conditions. I suggest that a disposition can be individuated purely in terms of the manifestation and stimulus only if the stimulus and manifestation occur intrinsically to the possessor of the disposition. We can say that D has the “intrinsic character” of M-ing when S, or is intrinsically characterizable as the disposition to M when S, just when events S and M are constituted by intrinsic properties of any object that possesses D.

If D is intrinsically characterized as the disposition to M when S, then for any x that has D, S and M intrinsically to x

An event is a change or transition of properties. An event can be said to occur intrinsically to x just when the properties that constitute the event are all intrinsic properties of x or any of its parts, or any intrinsic relations between x’s parts.

Consider again the salt and the water. We can say that the salt has a disposition to dissolve in water, that the water has a disposition to dissolve salt, and that the pair (salt, water) jointly has a
disposition to form a saline solution when salt and water are mixed together. All three dispositions seem to have the same manifestation and stimulus condition. The manifestation is a state in which the salt is dissolved in water. The stimulus condition is that the salt is stirred and mixed with the water.

There are three dispositions with the same manifestation and stimulus. However, only when applied to the pair (salt, water), does the disposition for salt to dissolve when mixed with water correspond to a disposition that is *intrinsically characterized*. Consider the manifestation event, that the salt is dissolved. This clearly implies a relation between salt and water. Being dissolved is therefore not an intrinsic property of the salt. It is also extrinsic to the water, since having some salt dissolved in it implies a relation to salt. (If we extract all the salt out of some saline water, we do not change the properties of the water itself. So, being saline is not an intrinsic property of the water.) Consider the stimulus event, which is that salt and water are mixed together. This also implies a relation between salt and water. The stimulus is therefore extrinsic to the salt and it is extrinsic to the water. However, both the manifestation and the stimulus are intrinsic properties of the pair (salt, water). So, the disposition to for salt to dissolve when mixed with water is intrinsically characterized when it refers to the *resultant* disposition of the pair (salt, water).

Consider, further, the key and the lock. We can say that the key has the power to open the lock, that the lock has the disposition to be opened by the key, or that the key and the lock together have a disposition for the lock to open when the key and lock are brought together. The power of the key to open the lock has a manifestation which is extrinsic to the key, since it refers to the state of the lock. The disposition of the lock, in addition, has an extrinsic stimulus condition, since that refers to the key. Neither the power of the key nor the disposition of the lock is an intrinsically characterized disposition. But the disposition of the pair (key, lock), is an intrinsically characterized disposition, since this involves only the objects of that pair.

Dispositions that can be characterized intrinsically in this way play a special role. They enable us to identify reciprocal pairs of dispositions in that a reciprocal pair \( D \) and \( D' \) is one that has a resultant disposition that can be characterized in terms of intrinsic stimulus and manifestation.
We may define the notion of a resultant of a reciprocal pair of dispositions required by analysis (6) above.

\[(D1,D2)\] is a reciprocal pair of dispositions iff (i) \(D1\) and \(D2\) are each components of \((D1,D2)\) and (ii) \((D1,D2)\) can be individuated in terms of stimulus and manifestation events that are intrinsic to its possessor.

We are now in a position to state the analysis of the response relation.

- \(m\) occurs in response to \(s\) iff (i) \(m\) and \(s\) are events of the same type as the manifestation and stimulus, respectively, of a pair of reciprocal dispositions \((D1,D2)\) and (ii) there is no dispositional \(D'\) to \(M\) when \(S'\) such that \((D1,D2)\) is a proper component of \(D'\) and an event of type \(S'\) occurs.

Causation is the ancestral of this relation.

Return once again to the trumping example. We can now see that the relevant dispositional pairs are \((D1,D^*)\) and \((D2,D^*)\). These are the pairs of dispositions that involve, on the one hand, Merlin and Morgana’s powers respectively and, on the other, the Prince’s receptive disposition to be transformed. The pairs themselves in turn compose to give the maximal disposition \([(D1,D^*), (D2,D^*)]\), or more simply \((D1+D2,D^*)\). It is this disposition that is relevant for understanding the response relation. At 12pm, the disposition exists and Merlin’s spell casting, the event \(s\), satisfies the stimulus condition. At 6pm, \((D1+D2,D^*)\) no longer exists. So Morgana’s action \(s'\) is not the stimulus of any reciprocal pair responsible for the Prince turning into a frog at midnight.

2.4.1 BRUTE COMPOSITION

We now need to consider why \((D1+D2,D^*)\) exists only until 12pm but ceases to be once \(s\) has occurred. The crux of our solution to trumping preemption is that under certain, brute conditions, dispositions do not compose. At 6pm, we have all the relevant pieces: \(D1\), \(D2\), and \(D^*\). But they do not compose. This suggests that the rules of composition might be unfamiliar and might allow for
violations of principles of intrinsicness. Given the contingency of such laws, there is no reason to expect any metaphysical laws for the composition of disposition more generally. That is, not only are the rules of composition contingent, but the very fact that there is composition is contingent. This, it might be argued, is implausible, mysterious, and goes against our the intuitive view that whenever a pair of dispositions obtain then they will always compose in the manner dictated by the rules of composition.

Let’s distinguish two types of contingency for compositional rules. On the one hand, there is a question of what disposition a given set of components may give rise to. On the other hand, there is a question of whether some set of dispositions will compose at all. Perhaps one might accept contingency in the determination of the character of the resultant of some set of component without accepting contingency in the matter of whether they will compose at all. An analogous position concerning the composition of material bodies would be that one accepts contingency in the question, for instance, of whether some set of parts will compose to form a human being or a giraffe, but deny that there is any contingency in the question of whether those parts will compose to form a composite material object at all.

The objector would hold that there is too much contingency entailed by our view that, when trumped by Merlin’s earlier spell, Morgana and the Prince do not together give rise to an aggregate disposition. The objector would hold that it is a matter of necessity and, moreover, intuitively the case that whenever a pair of dispositions D and D’ exist, then they will give rise to a resultant if they have one.

In response, consider that the Moon is visibile from certain points on the Earth at certain times. Sometimes, we can see it, sometimes we can’t. Being visible isn’t a property of the Moon alone, though it is ostensibly predicated of the Moon. It is a composite disposition whose components are the dispositions of the Moon to reflect light and the dispositions of the viewer to detect the reflections. Even though the Moon is not visible during the day, this does not correspond to any change in the dispositions of either the Moon or any observer on Earth, but on things like the ambient light due to the Sun drowning the light reflected by the moon.
One may insist that, in a sense, the Moon remains visible during the day—in the sense that it’s true during the day that the Moon can be seen at night. But this sense of “visible” is hardly the one we commonly use. For, according to that sense, the Moon is (trivially) always visible. Rather, I think, the natural meaning attached to the notion of visible is a disposition that comes and goes. But it is one that comes and goes for reasons extrinsic to the observer and the object of observation.

A similar lesson applies to the world of magic. Merlin casts his spell before Morgana. We may ask, does Morgana still have the power to turn the Prince into a frog? One answer is, “of course.” It is still the case of Morgana, given the Prince’s passivity, that if she were to cast a spell with no trumps, the spell would have its full effect. The other answer, and the more fundamental one according to our view, is “no.” Once Merlin casts his spell, Morgana loses the power to affect the Prince. Merlin doesn’t cause any intrinsic change to Morgana. But his act constitutes the non-satisfaction of the conditions for Morgana to be able to turn the Prince into a frog. This, as I have suggested, is to be understood as a failure of composition.

2.5 An Alternative Solution to the Problem of Trumping

Our solution to trumping in the previous section held that the disposition \((D_1+D_2,D^*)\) ceases to exist because the component dispositions \(D_1, D_2\) and \(D^*\) no longer compose. The conditions under which the existence of \(D_1, D_2\), and \(D^*\) compose to form \((D_1+D_2,D^*)\) cease to obtain after the disposition has been activated once.

There is an alternative proposal. To see how the alternative proposal works, assume for simplicity that there are exactly two wizards (Merlin and Morgana) and that both have the ability to cast at most one spell (Prince-to-frog) during the day in question, that Merlin’s spell may only be cast at noon, and Morgana’s only at 6pm. The relevant spell-casting events may be characterized as follows:

\[ S_1: \] Merlin casts a spell at 12pm to turn the Prince into a frog
S2: Morgana casts a spell at 6pm to turn the Prince into a frog

Given the laws of trumping, the disposition \((D_1+D_2,D^*)\) to turn the Prince into a frog has the stimulus condition:

either \(S_1\) or \((\neg S_1 \land S_2)\)

Suppose that as before \(s\) and \(s'\), respectively tokens of \(S_1\) and \(S_2\) occur. Then, according to the definition on p. 47, the manifestation at midnight is a response to \(s\) and not \(s'\). Firstly, even though \(s'\) satisfies the stimulus of the dispositional pair \((D_2, D^*)\) consisting of Morgana and the Prince’s dispositions, this is a proper component of the disposition \((D_1+D_2,D^*)\) that has as a manifestation the prince’s transformation into a frog. The larger disposition is the relevant disposition and \((D_2,D^*)\) is ignored. Only \(s\) satisfies the stimulus of the larger disposition, i.e. the disjunction \(S_1\) or \((\neg S_1 \land S_2)\). So, only \(s\) is to a stimulus for the manifestation of the enfrogment at midnight.

A similar proposal was raised earlier (p. 41). There, it was suggested that the Prince’s disposition \(D^*\) has a complicated stimulus condition that gets the case right. We argued that the strategy is undermined by the intrinsicism of dispositions. By contrast, the proposal at hand complicates only the stimulus condition of the larger, aggregate disposition \((D_1+D_2,D^*)\). It does not say that the stimulus of the disposition of the Prince to transform has stimulus condition \(S_1\) or \((\neg S_1 \land S_2)\).

The disposition \((D_1+D_2,D^*)\) characterized in terms of the stimulus condition \(S_1\) or \((\neg S_1 \land S_2)\) is not intrinsic to Merlin, Morgana and the Prince. Simply, if we suppose that Morgana can cast her spell at 11am as well as 6pm, then \(S_1\) would not be sufficient. There would be another spell casting event:

\[ S_3: \text{Morgana casts her spell at 11am} \]

The stimulus condition in that case must be:

\[ S_3 \text{ or } (\neg S_3 \land S_1) \text{ or } (\neg S_1 \land \neg S_3 \land S_2) \]
We can see that if we relax the simplifying assumptions and the wizards are limited in neither
the number nor type of spells they may cast, nor in the time at which they may cast them, then
a stimulus condition that is intrinsic to the relevant objects will be very complicated. It will be
difficult therefore to give a fully general characterization of the disposition \((D_1+D_2,D^*)\).

Here it might help to apply a lesson from Newtonian physics and the interaction of forces.
The language of forces provides an elegant way of describing the functional relationship between
particular token stimuli and particular token manifestations of pairs of interacting objects. Thus,
e.g., consider an object of mass 1 kg was pulled by a rope in an an eastward direction with a force
of 1 newton. Let the disposition of the rope be \(D_1\) and the disposition of the massive object be
\(D_2\). Newton’s second law of motion tells us that the object would have an eastward acceleration
of \(1 \text{ ms}^{-2}\). That is, it tell us that the pair \((D_1,D_2)\) manifests the object moving with an eastward
acceleration of \(1 \text{ ms}^{-2}\) when the rope is pulled eastward with force of 1 newton. This particular
stimulus-manifestation is only a partial characterization of the reciprocal pair \((D_1,D_2)\).

The laws of trumping dictate that the first spell of the day has its effect at midnight. This
describes a functional relationship between possible stimuli and manifestation of the dispositions
of the Merlin, Morgana and the Prince that allows us to determine, in the event that \(S_1\) and \(S_2\)
both occur, that the instance of \(S_1\) alone is the stimulus of the manifestation of the disposition
\((D_1+D_2,D^*)\).

According to the proposal at hand, how dispositions interact when they form reciprocal pairs
and which events count as stimuli and which are manifestations is determined by appealing to
the laws and mechanism of the world in which those dispositions obtain. While the individual
dispositions of objects are fixed, the interaction between any pair of dispositions \(D_1\) and \(D_2\) cannot
be determined simply by considering the manifestation and stimuli of \(D_1\) and \(D_2\).

This solution to trumping preemption and the one we gave in the previous both appeal to facts
about aggregate dispositions. But there is a tradeoff between the two solutions and one that can
inform the choice for a final dispositional theory of causation.

The solution of section 2.4 explains the mechanism by which trumps work in terms of dis-
positional loss. Trumping preemption, on our solution, is the same kind of phenomena as late
preemption. Compare: once Suzy’s rock hits and shatters the window, this no longer a dispo-
sition for the window to shatter when struck by a rock. So, Billy’s preempted rock has no disposition
for which to be a stimulus. Similarly, once the disposition \((D1+D2,D^*)\) is activated, there is no
longer a disposition for which Morgana’s spell can be a stimulus. Our solution leaves unexplained
the mystery of action at a temporal distance that allows a disposition to be lost before its manifes-
tation is realized. It also leaves unexplained the fact that \((D1+D2,D^*)\) “decomposes” once a spell
has been cast. Facts about composition are taken as brute.

The alternative proposal, by contrast, appeals neither to brute composition nor to brute action
at a temporal distance. But it does not explain the mechanism by which trumping works. Trump-
ing, on the alternative proposal, is just a way to describe the functional relationships between
stimuli and manifestations of reciprocal dispositions.

### 2.6 A Causal Basis?

Prior, Pargetter, and Jackson (1982) have argued that dispositions have causal bases. This is a
property which, together with the stimulus event, helps to produce the manifestation. As well
as helping to produce the manifestation, the basis “grounds” the disposition. On this view, for
instance, what makes an object fragile is that it has a property, such as a weak crystalline mi-
crostructure, which is one of the causes of any shattering that is also partly caused by impacts or
strikings, i.e. the distinctive stimulus condition associated with fragility.

Many have endorsed this causal basis thesis, including Lewis (1997). If this is right, then there
is a potentially serious problem for the dispositional theory of causation. I will argue that we can
save the dispositional theory of causation by giving up the view that counterfactual dependence
or nomological necessity implies causation. Thus, it will be shown that the dispositional theory
of causation differs in an important respect from the dominant contemporary understanding of
causation.
Let’s begin with the argument for the claim that dispositions have causal bases. The argument, due to Prior et al., aims to show that both deterministic and indeterministic dispositions have causal bases. We shall be concerned only with their argument that deterministic dispositions must have a causal basis. They define a causal basis as something which, together with the stimulus condition, “is causally operative sufficient for the manifestation” (Prior et al., 1982, 251). Their argument is to show that the stimulus condition is never sufficient for the manifestation. In that case, something other than the stimulus is necessary for the manifestation. From this it may be inferred that something other than the stimulus is a cause of the manifestation, i.e. there is a causal basis.

It may seem trivial that the stimulus condition is not causally sufficient for the manifestation. After all, dispositions can simply be masked. If, for example, a fragile vase were protected and the stimulus condition obtained, it might not break after all. But this would not show that there is a causal basis distinct from the stimulus. The stimulus can only be sufficient for manifestation together with the absence of masks or finks or any other types of dispositional inference. But this isn’t to say that the causal basis of a disposition is the absence of masks. Prior et al. (1982) take the view that if the vase does not break when the stimulus condition obtains, then it is not fragile. In effect, they either reject masking as a possibility, or would include the absence of masks in the stimulus condition, as per Lewis (1997).

This disagreement over the nature of masks does not matter. We can slightly revise the argument as follows. Take a causal basis to be whatever other than the stimulus and the absence of masks causes the manifestation. Let’s call the stimulus together with the absence of interferers the “broad” stimulus condition. The aim is then to show that if, say, a vase is struck and it breaks, then something other than the broad stimulus must be causally operative for the manifestation.

So the refined claim is that even the satisfaction of the broadly conceived stimulus conditions is not sufficient for a manifestation. This claim seems hard to deny. Striking an object, even under conditions ideal for the manifestation of fragility, does not guarantee that the object will shatter. Vases will shatter, but concrete walls will not. Striking is not sufficient for shattering. There must
be a difference between the two situations that explains why the one object shatters but the other does not. Let’s name the difference-maker “B.” B may be a categorical property or a dispositional property. Or perhaps all the relevant difference between the wall and the vase is that the vase, but not the wall, is fragile. S is insufficient for M and B is necessary.

The claim that B is the causal basis, i.e. B causes the manifestation, follows from the fact that B is necessary only if we assume that what is necessary for an event is must also be a cause of it. This assumption is true on the INUS analysis of causation given by Mackie (1980) as well as counterfactual theories of causation. On the counterfactual view, B is a cause of the manifestation since if B had not been present, the manifestation would not occur. Thus the argument for the causal basis thesis depends on accepting either of the following claims.

1. If □L(Γ ∧ C → E) and ¬□L(Γ → E), then C is a cause of E (where □L is nomologically necessity, Γ, C, E are conditions/events, and Γ and C occur before E)

2. If C□→E and ¬C □→ ¬E, then C causes E (where C and E are propositions that c and e, two mereologically distinct events, occur)

I make these explicit because I think that accepting the dispositional theory of causation gives us a reason to reject both of these deep-rooted assumptions about causation. The dispositional theory of causation conflicts with the causal basis thesis as will be presently explained. To avoid the conclusion of Prior et. al.’s argument that all dispositions have a causal basis, we ought to reject the claim that nomologically necessary events are causes.

The dispositional theory of causation is incompatible with the causal basis thesis. If both views were true, we would get a regress. Suppose that C causes E. Then according to the dispositional theory of causation, there is some disposition D whose manifestation is E and whose stimulus is C. And according to the causal basis thesis, D has a causal basis B that helps produce E. Now, the relation between B and E must have a dispositional basis D′ whose stimulus is B and whose manifestation is E, which in turn must have a causal basis B′, and so on ad infinitum. Thus, for any pair of causally related events, there is an infinite number of events that produce the second
The dispositional theory of causation implies the falsity of the causal basis thesis. On the dispositionalist view, if dispositions have bases, then those properties are not causes of the manifestation. Bases are nomologically necessary; the manifestation of a disposition may depend on the existence of a base as Prior et. al. have argued. But nomologically necessity and counterfactual dependence are not the same as causation and do not imply causation. An event may be a necessary condition without thereby being a cause.

According to the dispositional theory of causation, it is necessary for C to cause E that there exists a disposition D with C and E as stimulus and manifestation. Perhaps D is more fundamentally some other property, the “basis” B. Then B is necessary for C to cause E. B is a condition but not a cause. We may say then the D (or B) is a condition for C to cause E.

The distinction between causes and conditions is one of the major consequences of a dispositional theory of causation. Conditions are to be understood in terms of nomological necessity whereas causation is to be understood in terms of dispositions. A condition for some event to occur is a something which is necessary for its occurrence, but not as a stimulus is to a manifestation. In other words, conditions are necessary things that are not causes.
Chapter 3

DOUBLE PREVENTION

3.1 Introduction

Consider Ned Hall’s classic example of double prevention (Hall, 2004, 241). Suzy is a bomber pilot. Her goal is to drop bombs on a target site. A fighter called Enemy is dispatched to intercept Suzy. However, Enemy is shot down before he can get to Suzy. Billy, who is on Suzy’s side, takes out Enemy. Had Billy failed to shoot Enemy down, then Enemy would have caught up to Suzy and shot her down before she releases her bombs. Billy, in short, prevents the enemy from preventing the bombing. The neuron diagram (fig. 3.1) makes clear the structure of the case.¹

![Neuron Diagram](image)

**Figure 3.1:** Ned Hall’s double prevention case

More generally, where some event C prevents an event D from preventing E, we have double prevention. We may say that C “double prevents” E; or, more accurately, to avoid the misleading

¹A reminder about neuron diagrams: Neurons may either fire (filled black) or be dormant (blank). Neurons are linked by stimulatory (arrow) or inhibitory (ball-end) mechanisms. Neurons fire when receiving a stimulatory signal, and neurons that receives an inhibitory signal do not fire.
notion that prevention is happening twice, we may say that C cancels or prevents a threat to E.

These cases are puzzling. On the one hand, double prevention is strikingly unlike ordinary, more familiar cases of causation in many different respects and ought therefore to be treated separately. On the other hand, double prevention mechanisms seem to do some of the work we attribute to the concept of cause, insofar as events related by double prevention imply other kinds of causally-loaded relations. For instance, threat-cancellers (double preventers) can be relevant in explanation and claims about moral responsibility. This suggests that we should treat double prevention the same as other types of causation.

What, then, is double prevention—is it causation or something else? This chapter will argue that double prevention is not causation; we can make sense of the relevance of double prevention directly through the notion of causal explanation. Things can be causally explanatory without being causes.

Whether we take double prevention to be causation or not, there is a challenge down either road. Suppose we reject the view that holds double prevention to be causation, as I argue we should, on account of the striking differences. We must then provide an account of why and how a non-causal process such as that of double prevention can be relevant to causal explanation or transmit moral responsibility. I will dub this the Relevance problem: what makes double prevention relevant to the event brought about by double prevention?

Suppose, alternatively, as many have suggested, that we treat double prevention as a case of causation proper. We must then try to find an analysis of causation that will account for double prevention as well as the ordinary cases of causation. Given the striking differences, this is likely to be a difficult task. Let’s call this the problem of Uniform Treatment.²

Let us turn now to the striking difference between double prevention and cases of ordinary causes. The differences are explored systematically in Hall (2004). According to Hall, the structure of double prevention involves a violation of some of central features we ordinarily think are essential to causation. Causal relations are often held to be local, intrinsic to their causal relata,

² Cf. Paul and Hall (2013, 176ff.) on the “question of uniform treatment.”
and transitive. But if we mix double prevention mechanisms, we find violations of these basic features of causation.

- **Violation of locality.** We can see from the Billy/Suzy case that we have action at a spatial distance. Billy’s shooting down of Enemy can occur far away from Suzy and the bombing, and there is no connecting chain of events between Enemy and the bombing. Moreover, we have action at a temporal distance. The time of shooting down of Enemy is not immediately prior to the relevant bombing event. (See Hall (2004, 242) for discussion).

- **Violation of intrinsicness.** Consider the following variant of the case. Suppose that at the time of being shot down, Enemy didn’t know Suzy’s exact whereabouts. If he hadn’t been shot down by Billy, he would have received the coordinates from his home base. This does not change the structure of the case. By shooting down Enemy, Billy cancels a threat to Suzy as before. But the fact that the home base would have sent the crucial information is an extrinsic enabling factor for Enemy to be a genuine threat. An exact duplicate of the relevant events, i.e. Billy shoots down Enemy, and Suzy completes her bombing mission, is possible in a context where the home base has no knowledge of Suzy’s location and could not send Enemy Suzy’s coordinates. In that scenario, the relevant events are intrinsically the same as they actually are, but Billy does not cancel a genuine threat. Enemy never posed a danger to Suzy. So we have two duplicate chains of events that disagree over whether Billy was relevant to the bombing. But relations that are intrinsic to their relata do not vary between between duplicates of those relata in worlds with the same laws of nature (See Hall (2004, 245-6)).

- **Violation of transitivity.** Double prevention can be shown to involve a violation of transitivity in cases of “mixed chains” of ordinary causation and threat cancelling. If threat cancelling is a kind of causation and causation is transitive, then we should be able to chain together ordinary causation with double prevention. Counterexamples to transitivity try to show that some A causes an event C, which in turn cancels a threat to E (so, C “causes” E by
double prevention), but A does not cause E. Example: suppose the home base broadcasts unencrypted instructions for Enemy to shoot down Suzy (=the event “A” in the scheme above). The message reaches Billy who moves to shoot down Enemy. The instructions therefore cause Billy to shoot down Enemy, which in turn cancels a threat to the bombing just as before. But the broadcast of the instructions does not cause the bombing. This is in light of the fact that this event is also responsible for generating the threat in the first place. (See Hall (2004, 252)).

To these dissimilarities we may add that, in the terms of Lewis (2000), double prevention often lacks the how- and when-dependence usual in ordinary cases of causation. In short, a threat canceller behaves like an on/off switch, while a direct cause is more like a volume-dial with respect to the influence it has on the effect. It would not affect the intrinsic character of the bombing in any significant way if Billy had shot down Enemy earlier or later than he actually did, nor if Billy had shot down Enemy with bullets or a heat-seeking air-to-air missile. Now compare Suzy. The timing of the bombing depends on the time at which Suzy drops the bombs. Moreover, the type and number of bombs that she uses also affects the character of the resulting bombing.

Despite these peculiarities of double prevention mechanisms, philosophers have held that double prevention is an instance or type of causation, and that perhaps we should not think of causation as being a uniform relation after all. For instance, according to Jonathan Schaffer, the differences between ordinary causation and causation by double prevention indicate that “there is more than one way to wire a causal mechanism” (Schaffer, 2000a, 289). Hall (2004) argues that there are “two concepts of causation.” Garden-variety causation is a matter of production. Double preven-

---

3 Counterexamples to transitivity, in my view, are inconclusive. They trade on an ambiguity between explanation and causation that ultimately undermines their force. As I argue in the concluding chapter 5, it may well make sense to say, for instance, that A does not explain or is not responsible for the bombing, insofar as A is in a sense a self-cancelling threat. It may be that we mistakenly express this as the judgement that A is not a cause of E. However, A may well be a cause of E, but a non-salient one. In the case above, it seems odd to say that by broadcasting the location of Suzy, the enemy home base caused the bombing. But this oddness really only reflects the fact the broadcast is non-explanatory, not non-causal. See Lewis (2000, 96-99) for further arguments against counterexamples to transitivity.

While violations to transitivity are inconclusive, the violations of locality and intrinsicness are clear and sufficient examples of the way in which double prevention is strikingly different from ordinary causation.

4 See p. 5 in the Chapter 1 above for details on Lewis’s notion of influence.
tion, amongst other types of cases, exemplifies dependence. In Hall’s terms, Billy causes the bombing by dependence and Suzy causes the bombing by production.

Hall splits the concept of cause to accommodate double prevention. His account faces the problem of Uniform Treatment. We find “natural joints” between an intrinsic, local, covarying causal relation and a non-intrinsic, non-local, non-covarying kind. Production and dependence seem to be metaphysically distinct relations. An explanation is needed for why such distinct relations are unified under a single genus. To state that causation is the relation that consists of the heterogenous class of either \( R_1 \) or \( R_2 \) relations is ad hoc given that \( R_1 \) and \( R_2 \) bear no real resemblance to each other.

I think we should give up the common view that it is possible to cause something by double prevention. There is no causation by double prevention, and there is only one relation of causation, a relation close to Hall’s notion of production. But why think, in the first place, that double preventers are causally related to the effect? Why do Hall and Schaffer accept a heterogenous notion of causation?

One reason for taking double prevention to be causation is, as we have said, that cases of double prevention seem to be connected to other concepts in just the ways we expect causation to be. According to Schaffer, events related by double prevention involve

... all the central connotations of causation such as counterfactuals, statistics, explanation, inference, agency, and responsibility, involved in even the most paradigmatic cases of causation including all instances of human action, and involved in the most theoretically salient cases of causation relevant to the analyses of reference, decision, and perception. I cannot see a stronger argument that anything is causal (Schaffer, 2000a, 289).

I think we can resist this line of argument, and on its own terms. In section 3.2, I will consider some of the purported connections between double prevention and notions of explanation and show that things are not quite so clear. If the presumption is that double prevention is just like causation by production, in terms of how it supports causal explanation, then this presumption is

---

5 For Hall, production is the ancestral of intrinsic minimal nomological sufficiency. Dependence is understood in terms of the counterfactuals \( \neg C \iff \neg E \), where \( C \) and \( E \) are true propositions that distinct events \( C \) and \( E \) occur).
wrong. I will argue that explanation by double prevention is usually quite different from other types of explanation. There is no uniformity between double prevention and “ordinary” causation.

Moreover, as Beebee (2004) and Varzi (2007) have argued, we cannot infer causation from explanation. That is, it does not follow from the fact that C explains E that C is thereby a cause of E. This can be verified, for instance, on the theory of Lewis (1986a), that causal explanation is the provision of information about an event’s causal history. As well as undermining the argument of Schaffer and others, I think the theory of explanation as information provides a positive account of double prevention and a way to solve the Relevance Problem.

The second reason one might take double prevention to be causation is that many ordinary and paradigmatic instances of causation seem to turn out, on closer inspection, to be cases of double prevention. An example described below is where pulling the trigger of a gun causes a bullet to fire. Schaffer (2000a) has argued that this and many other cases of “disconnection” are cases of double prevention. But I shall argue in section 3.3, we can resist Schaffer’s description of these cases as involving double prevention or disconnection. Schaffer’s examples, I argue, involve causation by production, not causation by disconnection or double prevention. Here, we find support from the dispositional theory of causation of chapter 2.

3.2 Is double prevention causation?

Here is one argument that in Hall’s example, Billy, the threat-canceller, is not a cause of the target being bombed: Billy is not a cause because causation cannot involve absences. Billy is responsible for the absence of Enemy as a threat to Suzy’s bombing mission. It’s the absence of the Enemy that enables the bombing of the target. But absences are not events. They are nothing and cannot be causes.

Unfortunately this is not a promising line of argument since the possibility that double prevention is causation can be a reason to reject pre-theoretic principles concerning causation such as intrinsicness, locality, or that absences can’t be causes. For those who think that there is cau-
sation by double prevention, it would follow that there is absence involving causation. We need an independent reason to reject the idea that double prevention is causation.

Here is another argument that Billy is not a cause of the bombing. Suppose, as seems plausible, that there is an important connection between causation and moral responsibility. Given this connection, one may point to the intuition that Billy is less responsible, or responsible in a different way, for the bombing than is Suzy. Let’s suppose that both Billy and Suzy are similarly intent on ensuring that the bombing occurs, that they are both acting with full agency, that they are both of same military rank, that they stand equally with respect to all other norms, etc. That is, we are eliminating any differences with respect to the norms relevant for attributing moral responsibility.

It may seem at this point that the only explanation for the difference in moral responsibility is a difference in the causal relevance of the two agents for the bombing. However, there does seem to be a difference between Suzy and Billy with respect to the bombing besides the fact that one is a double preventer and the other a producer. As we saw above (p. 59), there is far less covariation between the bombing and Billy’s action. Perhaps this difference in influence can account for any intuitive difference in judgments about moral responsibility. The difference in responsibility can thereby be explained in terms of a difference in covariation. This, it should be noted, a difference in degree not in kind of responsibility: Billy bears less moral responsibility because he exerts less influence than does Suzy. As we saw, Suzy can control when, whether and how the bombing occurs to a large degree by, for instance, altering the timing at which she drops the bombs or the type and number of bombs she uses. Billy can at best determine whether the event occurs. Alterations to his shooting down Enemy would not have affected the character of the bombing event.

We shall put aside the issue of moral responsibility and causation for now and return to it in the next chapter (4). Suffice it to say, these two arguments for treating double prevention as non-causation are inadequate to make the case. I will argue now that the distinction between Billy and Suzy, and more generally between events that play the role of threat-cancellers and those that play the role of producers, is based on a difference in causal explanation. Citing the role of a
double preventer answers a very different “why” question than does citing a producer. There is a systematic difference in the explanatory role of causes of an event and the events that merely allow it.

Imagine that we are watching an action movie.⁶ We reach the point in the plot where Suzy is sent to destroy the enemy target. Just then you go into the kitchen to get a glass of water. When you come back, you see that the headquarters of the enemy is destroyed. You ask me, “How did the headquarters get blown up?”

Me: Well, you see, Billy shot down an enemy jet.
You: I don’t understand. What does that have to do with the tower that was destroyed three miles away?
Me: If Billy hadn’t shot down the enemy jet, the enemy would have prevented the bombing.
You: I still don’t understand. How would the enemy have prevented the bombing?
Me: Well, the enemy would have shot down Suzy. It was Suzy who dropped the bombs.

The scenario demonstrates that to explain by citing a double preventer, one must invoke further elements about the causal history, specifically that there is some threat to a direct producer of the event to be explained. That Billy shoots down the enemy cannot stand alone as an explanation. One must also be told that Suzy produces the bombing and that the enemy threatened that process.

By contrast, in cases of production, or ordinary cases of causation, citing the cause can be a satisfactory stand alone explanation of the event. Consider a different way in which I might have answered you.

You: How did the headquarters get destroyed?
Me: It was thanks to Suzy. She got to the target site and bombed the tower.

Unless you ask further how Suzy managed to accomplish this feat, i.e. ask for an explanation for a different explanandum, citing the cause of the bombing does not demand information about the

---

⁶ Thanks to Achille Varzi for suggesting the form of the example.
existence of threats and the threat-cancelers, at least not in order to understand the explanatory relevance of Suzy for the bombing.

In general, I claim that this explanatory asymmetry holds between causes and threat-cancelling mechanisms of double prevention: the relevance of a “cause” by double prevention requires further explanation in terms of threats to a causal process for the effect, whereas the relevance of a cause for the effect does not require any further explanation. This, moreover, is a difference in the kind of explanation that is adequate and not merely a difference in the degree of explanation.

There is an objection that may be made here. I am arguing that double prevention is fundamentally unlike causation. But explanation is a context-sensitive matter. There might well be a scenario in which citing the role of the threat-canceling cause does stand alone in the way that I say only causes do. That is, there might be a situation in which “Billy shot down the enemy” is a good enough explanation and does not demand to be further supported by auxiliary facts about threat-canceling. In that case, it would seem there is nothing fundamental about the difference between causation and double-prevention: the fact that an explanation stands alone is a factor of contextual salience and supposition and does not track any metaphysical difference between causing and mere allowing, or causes and threat-cancellers.

Consider such a scenario. We are watching a sequel to the original action movie above. The plot is slightly different, but the characters are the same. For the last twenty minutes of the movie, the good guys (Suzy, Billy and co.) have attempted to get close to the target site. But plane after plane is shot down by a disturbingly efficient bad guy, Enemy. At this point you leave the room, exhausted by the repetitive drama of the action. When you come back you see that the target has finally been destroyed. It might be a satisfactory explanation of the fact that the HQ was destroyed, in this context, to learn that Billy shot down Enemy. And so, “the bombing occurred because Billy shot down Enemy” could be a good explanation here.

But this does not show what the objection wants to show, that the difference in whether or not the explanation stands alone is merely contextually-sensitive. For, in the new context with the efficient Enemy, the claim that some good guy dropped the bombs on the target is implicit,
and may be presupposed rather than explicitly stated. In providing explanations, not all parts of the explanans need to be explicitly stated or communicated. We can give explanations partially by saying and partially by relying on commonly-known facts to fill in the rest of the necessary details that are left unsaid.

The significance of the context shift is that it makes implicitly available the additional premises needed to make the explanation work—the premise that the bombing is directly produced by a bomber pilot flying to and destroying the target. Whether the requisite information is provided implicitly by the context or explicitly as a premise in the explanation doesn’t matter. In either case, “the bombing occurred because Billy shot down Enemy” is only satisfactory when other pieces of information are presupposed or made available. In the scenario at hand, after all, you knew prior to leaving the room that Enemy was preventing a swathe of attempts to run the bombing mission with disturbing regularity. When I tell you that Billy shot down Enemy, you are able to complete the explanation by yourself, filling in the additional premises. Context doesn’t affect the structure of a full explanation; it merely affects which aspects of the full explanation to communicate and which can be assumed as known to the hearer.

You: How did the target finally get destroyed?
Me: Billy shot down Enemy!
You: Oh I see. That was the enemy jet that was shooting all the bombers. Clearly, if Billy hadn’t shot down Enemy, he would have prevented any bombing attempt. And, so now that the Enemy has been shot down, one of the good guys was finally able to reach the target and drop the bombs.

Once again, the matter is different when we explain by citing a cause. There is no assumption in the background about specific threats or threat-cancelers. You can fully understand that Suzy’s dropping the bombs is explanatory without knowing anything about the enemy or Billy. It’s true that the fact that Suzy did successfully drop the bombs and destroy the target rules out the success of threats. That is, if you understand that Suzy explains the bombing, then you must know that any threats have been canceled. But this is not specific information about any particular threats, such as Enemy, or any particular threat-cancelers, such as Billy. Put differently, if Suzy caused the
bombing, then this implies that nothing interfered or prevented Suzy from causing the bombing. It does not however imply that Enemy is a threat and potential preventer, let alone that Billy shot down Enemy.

The purpose of this section was to show that double prevention and causation are like apples and oranges. The default assumption should not be that double prevention is just like causation. In explanations, double preventions functions very differently to production, and threat-cancellers play a different role than causes in helping explain the explanandum event. This systematic explanatory difference itself suggests a metaphysical distinction between causation and double prevention. Any account that seeks to treat the cases uniformly will have trouble accounting for that.

3.3 Causation by disconnection?

Fig. 3.2 shows the internal trigger mechanism of a bolt-action rifle. It will be helpful for understanding the process by which pulling the trigger can cause a bullet to fire. At first, the gun is cocked. This involves retracting a coil into a compressed state. Once the coil is sufficiently compressed, the sear locks the coil, keeping it compressed. When in this “ready” state, pulling the trigger displaces the sear, which releases the compressed coil. The coil rapidly decompresses with the typical spring-like motion and propels the striker. The striker’s impact causes the bullet in the barrel of the gun to explode and shoot out.

Schaffer (2000a) claims that pulling the trigger causes the bullet to fire by “disconnection,” a double prevention mechanism. Fig. 3.3 shows the gun mechanism in the form of the neuron diagram for double prevention. According to Schaffer, pulling the trigger merely prevents the sear from preventing the coil from decompressing and causing the bullet to fire.

On our account, double prevention is not causation. If guns fire by double prevention, that would make pulling the trigger merely an enabler of the bullet firing and not a cause. This conflicts with everyday common sense, which tells us that firing a gun is causation. Since guns work by
double prevention, there must be causation by double prevention.

But is this really double prevention? Should we think of Schaffer’s examples as involving causation by “disconnection”?

Consider the following case. You are holding a perfectly ordinary ball four feet from the ground. You let go of the ball, which falls to the ground. Now consider the relation between these two events.

C: You release your grip on the ball

E: The ball hits the ground four feet below

It’s would be odd to describe the relation between C and E in terms of double prevention. That is, it would be odd to say that by releasing the ball, you are preventing your hand from preventing the ball’s fall. More simply described, when you release the ball, you are ceasing to hold the ball.
Releasing is not preventing a preventer because cessation is not prevention. Rather, cessation is the production of a positive event. In the case of the ball, ceasing to hold the ball is to cause the ball’s potential energy, had in virtue of its position in a gravitational field, to be converted into kinetic energy.

Consider other examples of cessation.

- You are on a walk and you stop to look at a store window.
- You are a reading a book and you stop in the middle to make tea.
- A computer executes a series of code, comes to a bug in the code, and stops compiling the program.
- After hours of rain, the clouds dissipate and the rain finally stops.

None of these cases ought to be thought in terms of prevention. You do not prevent yourself from walking. You simply stop walking and come to a halt. When the clouds dissipate, this does not prevent it from raining.

To solidify intuitions further, consider a minor variation to the event C. Instead of simply releasing your grip, you very lightly throw the ball upwards. The ball then falls back down, passing through exactly the same point as in the original case where you merely let go of the ball and hits the ground at exactly the same point as before.

\[ C': \text{ You lightly throw the ball up} \]

\[ D: \text{ The ball reaches an apex and begins accelerating downwards} \]

\[ E: \text{ The ball hits the ground} \]

We wouldn’t describe \( C' \) as causing \( E \) by double prevention. \( C' \) is an action that causes the ball to accelerate upward at first. \( C' \) causes \( D \) directly, which in turn causes \( E \). We have a simple chain of events where one event causes the next directly. The difference between releasing and giving
a slight upward thrust, between C and C’, isn’t significant enough to alter the causal structure of the case. Both cases involve direct production and not disconnection or double prevention.

The same kind of causation by cessation is involved in the trigger mechanism of the gun. Pulling the trigger causes the sear to cease holding the coil in its compressed state. Pulling the trigger thereby causes the potential elastic energy to convert into kinetic energy, moving the hammer in the barrel to strike the bullet.

I want to suggest now that Schaffer’s analysis of disconnection as involving double prevention is misleading. It is misleading because the neuron diagram applied to both double prevention and disconnection obscures an important difference that is revealed by applying the dispositional theory of causation advanced in chapter 2.

According to the dispositional theory of causation, causes are stimuli and effects are manifestations of dispositions. Sometimes, the removal of a force can count as a stimulus of a disposition. Consider again the simple case of releasing a ball. We can say that the ball has the disposition to fall when I let go of it.

Now consider a contrasting case where no dispositional ascription clearly fits. The ball is in mid-fall, already accelerating towards the ground. My hand is outstretched in its path. But before the ball comes into contact with my hand, I move away. In this case, the absence of my hand does not count as a trigger of the ball’s disposition to fall. There is no such disposition.

Again, both examples might well be shoehorned into the double prevention neuron diagram. Both could be described in terms of the locution of double prevention. That is, in the first case, we could say that by releasing the ball, I am preventing my grip on the ball from preventing gravity from pulling the ball down. In the second case, we could say that I am preventing my hand from preventing the ball from hitting the ground. But no real lesson is to be learnt here. The two cases are different. In the one case, the releasing of the ball is a stimulus of a disposition; and in the other, the absence of my hand, a potential preventer, does not stand in a stimulus relation to any disposition of the ball.

We should not be concerned that absences are stimuli or causes. The event of a force being
removed, or a latch being disconnected, or A being pulled away from B, etc., are all positive, concrete events. Removing is no more negative than moving. Releasing the ball is a positive event.

The cases Schaffer calls “disconnection” all seem to involve dispositions with stimuli that involve the removal of some external force. In the gun case, we have a coiled spring that is disposed to uncoil when the sear is removed. Again, this is just like the case where I am holding the ball, which is disposed to fall when I let go of it. Compare this to the case of Suzy on her bombing mission. It is not natural to describe Suzy as having a disposition to bomb the target when enemy jets are shot down. No such disposition exists. Therefore, Billy’s shooting down Enemy does not serve as a trigger for Suzy to cause the bombing.7

These cases of causation by cessation are odd insofar as there is no energy transfer between cause and effect. Removing x causes y to manifest; but there is no biff and no transference of physical energy or momentum from x to y. Insofar as connection involves biff, Schaffer is right to call these cases “disconnection” and right also to argue that they constitute a deep problem for biffy theories of causation that require a transfer of energy or some physical quantity. But the dispositional theory allows us to see that there is a connection between x and y more fundamentally in terms of stimulus and manifestation. In the cases we’ve considered, the manifestation is the process of conversion of potential energy to kinetic energy. The stimulus is the event I’ve called a cessation such as the releasing of a hand’s grip or the displacement of the sear.

Our solution rejects the description of the gun mechanism as involving disconnection. It involves causation by cessation. It is worth criticizing an alternative view that accepts Schaffer’s description of the cases, but suggests that when we judge there to be causation, we are implicitly ignoring the inner-details of the gun. That mechanism is being “black-boxed.” The fact that

---

7 One might think that we still have causation if the coiled spring was already in the process of uncoiling and where the sear doesn’t trigger the uncoiling, but merely would prevent the spring from uncoiling all the way. Pulling the trigger of the gun serves only to move the sear out of the way of the coil. In short, we could imagine a gun mechanism that looks more like the case where the ball is already falling and I move my hand out of the way. But my response to the variant of the gun case is a shift in judgment. It’s not as clear that there is causation when the spring is already uncoiling.
there is double prevention going on in the details is being ignored when we judge, for instance, that Booth shot Lincoln and caused him to die. The difference between describing the case as disconnection or as causal production is one of pragmatics.

This black-boxing strategy is suggested in Hall (2004). Hall describes different “levels of description” at which it is possible to represent events. There is the level of the gun components—the sear, the trigger, the spring, the bullet. The events at that level of description are changes to the states of those components. There is further a higher level of description that involves the gun as a whole, not distinguishing between its components. At this higher level, the events involve changes to the gun itself. For instance, the event of the gun firing would be a higher level event. The event of the sear unhinging would be a lower level event.

Figure 3.4: At the high level of description: a stubborn neuron

Figure 3.5: Inside the stubborn neuron C: the lower level

Hall gives the example in fig. 3.4. The neuron C in fig. 3.4 is constituted by a set of neurons, the ones inside the box in fig. 3.5. The box is a neuron at one level of description; the neurons that compose it are indviduated at a lower level.

The box neuron (C) is a “stubborn” neuron that needs two stimulatory signals to fire. Looking at fig. 3.4—treating C as a black-box—it is clear that B helps cause C to fire and so helps produce E. (Apply this to the case of the gun: take C to be the gun, A to be the act of cocking and loading the gun, and B to be the pulling of the trigger. So pulling the trigger causes the gun to fire.)
But what should we say when looking at fig. 3.5 and we see that B does not produce E but merely allows A to do so? Does B produce E or not? One option, following Hall, is to say both: it produces E relative to one level of description, viz. the higher level that ignores the details, and it does not produce E relative to another, viz. the lower level that supposedly shows disconnection. Causation is a relativized relation, relative to the level of description. This resolution has the virtue of capturing our conflicting intuitions about cases of disconnection. But it is problematic since it is not clear what rules apply in deciding which level of description and which carving of events is salient.

Hall (2004) leaves the things open and unanswered, saying, that it is a “major piece of unfinished business is to spell out the relevant notion of ‘levels of description,’ and to explain exactly how such levels find their way into the contexts in which we make our causal judgments” (Hall, 2004, 274). A simple suggestion due to Moore (2009) is unworkable. Moore suggests that when two events “are close enough... we should say that to do an act causing the first is to do an act causing the second” (Moore, 2009, 461). According to this proposal, since, for instance, the displacement of the sear is spatially close to the firing bullet, then we should treat causing the sear to move as also causing the bullet to fire. But proximity is not sufficient for causation. If I throw paint at the gun just as it fires, then I cause something near a bullet firing. Surely, however, I don’t cause the bullet to fire.

3.4 Causal Explanation

We noted that some philosophers have thought that if C helps explain E, then C is a cause of E. This is one main reason why some have thought that double prevention is a type of causation. Deny this seems to generate what I called the Relevance Problem. How do double preventers explain if not by causation? But there is a theory of causal explanation on which the inference from causal explanation to causation is not valid. Things can be causally explanatory without being causes. This is David Lewis’s (1986a) theory that causal explanation is information about
the causal history.

On Lewis’s view, a causal explanation of some event provides information about the causal history of that event. The complete causal history of an event is the full set of its causes all the way back. The quality of an explanation is a function of the conversational relevance and the strength of the information about the causal history. On this view, not every adequate explanation needs to be a proposition of the form “C is a cause of E” and not every explanans is explanatory in virtue of being itself a part of the causal history, that is, in virtue of being a cause. Some events are explanatory and provide information about the causes in an indirect way. Sometimes an event that is not part of the causal history is relevant to help us identify another event which is part of the causal history. I can tell you that Columbus Avenue is the road parallel to Amsterdam Avenue, helping locating one by referring to another.

One type of indirect information about the causal history is counterfactual and modal. For instance, it is information about the causal history of E that, had it not been for D, C would not have caused E. Counterfactuals are relevant to causation insofar as they can be effective ways to communicate explanatory information about the causal history. But counterfactuals do not constitute the causal relation. The full force of this claim, therefore, requires giving up the sufficiency of counterfactual dependence for causation. Otherwise, ¬C ⊢ ¬E entails that C causes E. But while denying this entailment, the overall view also provides us with a new way of understanding the connection between causation and counterfactuals. The connection is between counterfactuals and causal explanation.

I am following the overall strategy deployed by Beebee (2004) and Varzi (2007). Beebee and Varzi are primarily concerned with cases of omissions, cases where some non-occurrence of an event is said to cause something. For example, my failure to water the plant can be said to cause it to die. Both authors want to show that we are not committed to the existence of negative causation and negative events when we say that not-C causes E and argue that there is no causation by omission. They argue that “not-C causes E” conveys information about the causal history of E without implying the negative event not-C, or the event of C’s absence, as a
part of that causal history. Part of the strategy, following Davidson (1967), it is the distinction of idioms of causation and explanation. As Varzi puts it, “we often speak loosely” when we use causal language; “C causes E” is ambiguous between “C explains E” and “C produced or caused E.” Claims of omissive causation such as “the absence of C caused E” should be read in the explanatory idiom as “the absence of C explains E.”

How do reports of omissions—statements of the form that “the failure of C caused E”, and understood in the explanatory idiom of “cause”—convey information about the actual causal history if they do not imply that such-and-such negative event or absence is one of the causes? One suggestion is that absences are relevant insofar as they are negative existential claims about the causal history. Varzi (2007, 7), for instances, suggests that in giving an explanation by omission, “we are just saying what (sort of) events are missing from that history.” On this proposal, we should read “not-C causes E” as “it is not the case that C causes E.”

But there are many things that are not causes of any particular event and not all of them are relevant. It is not a cause of the plant’s death that it rained in Hong Kong on Monday, or that a bank robbery occurred in San Diego. Those events are hardly relevant.

Lewis gave examples of how negative existential claims about causal history may constitute good explanations in certain contexts. “Why was the CIA man there when His Excellency dropped dead?—Just a coincidence, believe it or not” Lewis (1986a, 220). The negative report that the CIA man is not a cause of the death is salient here because there is a presupposition that it is false. We suspect, from what we know about international espionage, that the CIA man had a hand in the death. So it seems that bare reports of non-occurrence are only really informative when we presuppose that, to the contrary, the event did occur and was relevant. More generally, the claim that “C is not a cause of E” is explanatory when it is supposed that C is a cause of E.

This formula does not work in the case where Flora’s failing to water is said to cause the plant to die. That proposition must first be translated to, “Flora’s watering the plant did not cause the plant to die” or ”Flora’s watering is not among the causes of the plant’s death.” This is explanatory if we are to suppose that Flora does water the plant. However, there is no reason to suppose that
amongst the causes of the orchids’ death is Flora’s watering them. Why would anyone have thought that? We might have thought that Flora should have watered the plants, and in that sense we were expecting her to water the plants. But this is not the supposition that by doing so, she is amongst the causes of the death. On the contrary, the expectation is that Flora should have watered the plants to prevent their death, not to cause it. By contrast, when we presuppose, in Lewis’s case, that the CIA man was involved, we presuppose that his being there was a cause of the death.

There is a different analysis of the relevance of negative existentials. According to Beebee:

... when I say that the orchids’ death occurred because Flora failed to water them, you learn something minimal about the death’s causal history: that it did not include an event of Flora’s water the orchids. But you also learn something about the causal structure of nearby worlds where Flora didn’t fail to water the orchids—namely, that the causal process that ensued at those world did not cause (or perhaps might not have caused) the orchids’ death (Beebee, 2004, 305).

As well as the information about the actual causal history, namely that it does not include such-and-such events, according to Beebee, absences are relevant for describing nearby worlds. We learn, for instance, that had Flora watered, the plant would have survived.

This counterfactual information is prima facie relevant. But what both Beebee and Varzi have to say about how counterfactual reports are relevant is somewhat confusing. Directly after the passage quoted above, Beebee says, “None of this, of course, is information about what causal processes there were in the actual world; it is information about what causal processes there are in the closest worlds where the actually absent event occurs” Beebee (2004, 306). Varzi similarly says, “Causal statements concern events that feature in the history of this world; counterfactuals concern what goes on in other worlds” (Varzi, 2007, 7). But then, why should information about a non-actual causal history of some altogether different event be explanatory of the actual event? The working theory of explanation considers only information about the actual causal history of the explanandum to be explanatory.

An answer to this might be available from Lewis. Lewis tells us that some explanations serve
to answer contrastive why-questions: Why A rather than B? What is wanted to answer such questions is distinguishing information. Such distinguishing information might be that some event is part of the causal history of A but not that of B. The contrast might well be between an actual causal history and a non-actual one. Why did Lewis visit Melbourne in 1979 rather than Oxford? “Because Monash University invited me... and if I had gone to one of the other places, presumably that would not have been part of the causal history of my going there” Lewis (1986a, 230). So if some claim C is relevant to explain why A rather than B, it is because if B rather than A had occurred, C would not have been a cause of B.

Perhaps, then, it is the backtracking counterfactual claim that if the plants were to have survived, they would have survived thanks to Flora watering them. Thus the event of Flora watering the plants would have distinguished two possible causal histories: the actual one in which the plants die, and the counterfactual one in which the plant survived.

How we ought to evaluate backtracking counterfactuals is far from clear. It is not clear that if the plants were to have survived, it would have been because Flora watered them. Perhaps Flora wanted them to die and in that sense willfully avoided watering them. If there had survived, it would have been thanks to Jim, the neighbor who happened to have a spare key and saw the orchids in the window.

Counterfactual information is relevant. The mistake that Beebee and Varzi make is to think that information in the form of a counterfactual conditional statement is information about other possible worlds. Counterfactual facts are actual facts about actual things, and not hypothetical facts about hypothetical things. It is true of the actual causes of the death of the orchids, for instance, that some of them were preventable by Flora’s watering. That is, it is true of some of the causes of the plants’ deaths that had Flora watered the plant, the causes would not have occurred or been able to bring about the death of the plant. This is a fact about the actual world and the actual causal history of the dead orchids.

Consider this in more detail. Suppose the orchids died because of desiccation: ambient heat caused the water in the cells to evaporate to a point where ordinary living processes cease. So, a
part of the causal history is that the plants dried up. Now, it is true of this event, the drying up, that it would not have occurred had Flora watered. Being preventable by watering is a particular and identifying fact about some events and not others. So this property, conveyed more in terms of the counterfactual predicate, being an event E such that C \( \Box \neg E \), helps to rule out certain events and hence to narrow the list of possible causal histories. Counterfactuals, in this way, provide information about actual events.

This strategy can be applied to double prevention cases as well as to cases of omissions. It is true of Suzy that had Billy not acted, certain events necessary for Suzy to complete her mission, ones that feature in the causal history of the bombing, would not have occurred. We identify parts of the causal history of the bombing by their distinguishing feature of depending on Billy’s action. In that way we explain the bombing by citing Billy’s action.

It does not follow from the fact that Billy explains the bombing that Billy is also a cause. Any intuitive judgement that “Billy causes the bombing” is not indicative of causation. In this case, “cause” is used in the idiom of explanation and not causation. Billy’s action can help explain merely in virtue of being implicated in certain counterfactual claims about Suzy.

Hall (2004, 253) argued that “counterfactual dependence is too causation” because “it is in part because the first happened that the second happened, that the first event is partly responsible for the second event, that the occurrence of the first event helps to explain why the second event happened, and so on.” Hall’s argument is that when E depends on C, C explains E; and if C explains E then C causes E. So if E depends on C, C causes E. This is an argument for the sufficiency of counterfactual dependence for causation. But, as I have argued, the second premise is mistaken: explanation does not entail causation.

It seems to me that if you accept the Lewisian line on causal explanation and the suggestion that I made that counterfactual conditionals provide information about actual events, then the sufficiency of dependence can go. It is a platitude about causation that causation has something to do with counterfactuals. The view espoused here is that the link between counterfactuals and causation exists at the level of causal explanation. Counterfactuals are often useful to give information
about the causal history.
Chapter 4

VALUE AND RESPONSIBILITY

4.1 Introduction

There is an argument to the conclusion that absences can be causes from the notion of responsibility. Indeed, I think this is one of the main motivations for the view that there are causal absences. The argument goes like this. Responsibility is a causal notion. In particular, if one is responsible for some outcome, there must be a causal relation that links the agent to the outcome. People are in fact often held responsible for what they fail to do. For instance, one might be blamed or punished for failing to save a drowning child. Therefore, agents can cause certain outcomes (e.g. the drowning) by the omission of some appropriate action. Call this the “Responsibility Argument” henceforth.

There are good reasons to reject the conclusion. In Chapter 3, I gave an argument that the explanatory role of absences is different from that of presences. Absences function as conditions on causation, rather than as causes themselves. Moreover, omissive causation leads to normative encroachment of metaphysics, and this is another reason to resist causal absences. As has been shown by Beebee (2004) and McGrath (2005), what counts as a causally-relevant omission importantly depends on consideration of what is acceptable or normal. The classic example to support this shows that, while my gardener failing to water my plant may be said to be a cause of its death, the Queen of England’s failing to water my plant cannot be said to be a cause. The only difference seems to be that my gardener is supposed to water the plant whereas no such norm or
expectation is attached to the Queen.

The problem with the Responsibility Argument is that it takes the wrong view about responsibility. We ought to deny its first premise, namely, that if x is responsible for y, then x causes y. The position taken here is analogous to one that was held for the relation between explanation and causation in the previous chapter 3. As was seen, several authors have denied that “E because C” entails “C causes E.” This does not mean, however, that we need to give up on a causal theory of explanation. On Lewis’s view, for instance, explanation is information about causal histories.

In this chapter I sketch a view about responsibility that allows for omissive responsibility without omissive causation. I believe it follows from a broadly consequentialist view about morality according to which right action is that which produces the greatest intrinsic value.

There are two components to a consequentialist view about how actions get their value. The first is a transmission principle that holds that value transmits from effects to actions and that the value of an action is a function of the value of the effect. This motivates the core assumption that responsibility entails causation. The second component is a comparative principle that determines which of a range of possible actions is better or best.¹ I will argue that the comparative principle allows us to hold a person responsible in cases of omission. In the case of letting a child drown, for instance, one is blameworthy not in virtue of the consequences of the omission, but rather because saving the child, an alternative possible action, would have produced the greatest amount of good in the circumstances.

The topic of responsibility is part of a vast literature in ethics and the metaphysics of value, and this chapter will not be able to do justice to all important issues and debates about responsibility. But the goal here is merely to shift the discussion about omissions away from metaphysics.

---

¹Broome (1999) has complained that in doing ethics, philosophers do not tend to think comparatively in terms of “betterness,” but tend rather to think categorically in terms of “good” and “bad.”

Since economists deal so regularly in preferences, they think naturally in comparative terms, and, when they come to think of good, they will naturally ask not what is good, but what is better than what. Philosophers seem not to have this same instinct to think comparatively (Broome, 1999, 9).

Broome’s observation would explain why some metaphysicians have tended to endorse views that admit of causal absences.
and into these other areas. By showing that there is a possible view according to which there can be omissive responsibility without omissive causation, we can best defend the view of causation defended in previous chapters that denies the causality of omissions, absences and mere conditions.

4.2 Does Responsibility Entail Causation?

4.2.1 A BRIEF SURVEY

Our concern is with the second premise of the Responsibility Argument. This is the premise that responsibility entails causation.

\[(R \Rightarrow C)\] If an agent A, by her action C, is responsible for some event in the world E, then C is a cause of E

I shall argue that this principle shouldn’t be accepted. Of course, if you already accept the distinction between causes and conditions I have suggested, then \((R \Rightarrow C)\) is easily rejected. An agent may be responsible for an outcome by being a condition for some process, rather than being a cause. (The bystander in a position to save a drowning child is merely a condition that allows the child to drown, an event that is positively caused by the deadly combination of gravity and water.) Responsibility, then, would not entail causation, but may entail only the disjunction of causation or being a condition. You might, on the other hand, be under the sway of \((R \Rightarrow C)\) and reject the cause/condition distinction instead. So I shall have to work harder to reject the view that responsibility entails causation. My strategy will be to undermine what I think is the best support for the principle.

In a discussion on moral responsibility and causation, Sartorio (2007) helps make the \((R \Rightarrow C)\) view more explicit. She claims that many philosophers would accept \((R \Rightarrow C)\). She also thinks that many philosophers accept the view that causation explains facts about responsibility: reasons for holding someone morally responsible for some outcome would include that they helped cause or bring about that outcome.
Sartorio appeals to the following example to motivate \((R \Rightarrow C)\).

Imagine that someone is killed in a remote part of the world: someone with whom you have never had contact of any sort. Do we consider you to be morally responsible for his death? Obviously not. Why not? Intuitively, because you weren’t causally related to the death in any way: there is no way to link anything you did to the person’s death.

While the view appears to be widespread and it is (allegedly) intuitive, unfortunately clearly stated arguments for the link between responsibility and causation are hard to find.

Now discussion in the literature has often been focused on the connection between responsibility and the notion of control. Arguably, if one is responsible for some eventuality, then one must have had some control over it. Can we argue from this view that responsibility entail causation because control entails causation?

It is far from clear that having control over the occurrence of some event entails that one has brought about that event if it occurs. In fact, we find that control is usually understood in terms of having been in a position to prevent an event rather than being a cause of it. For instance, (van Inwagen, 1978, 205) offers the principle that: “A person is morally responsible for a certain event... only if he could have prevented it.” Does van Inwagen’s principle entail something like \((R \Rightarrow C)\)? Only if being in a position to prevent an event entails that one is a cause. That is, only if “could have prevented” entails “caused.”

As already mentioned, someone who accepts the cause/condition distinction can say that, in at least some cases where an agent could have prevented some outcome, they are merely conditions for causation without thereby being causes. Take Jill: she sees a man across the street mischievously untie a dog leashed outside a store. The dog runs away. Jill is watching across the street and could have tried to catch the dog so that its owner would find it. She could have prevented some anxiety for the dog’s owner. Her inaction is merely a condition on our view. Unlike the man who unleashes the dog, she doesn’t produce the unfortunate series of events. Far from ultimately giving us a reason for holding omissions to be causes, an argument for \((R \Rightarrow C)\) based on the link between control and responsibility presupposes that omissions are causes rather
than conditions. It is therefore a question-begging and not very promising basis for \( R \Rightarrow C \).

The idea that responsibility entails causation via the notion of control might be derived through a difference-making notion of causation. Suppose for instance that having control over an event \( E \) entails that whether \( E \) occurs *depends* on how you act. Now, if dependence is sufficient for causation, as it is on a Lewisian, counterfactual theory of causation, then control entails causation. Yet the sufficiency of dependence for causation is a controversial assumption. It is invalidated, for instance, by the dispositional theory of causation we have defended in chapter 2 of this dissertation; there can be dependence between events that are not linked as stimulus to manifestation of a disposition (or chain of such links).

How might a defender of \( R \Rightarrow C \) establish the principle without begging the question against the dispositional theory of causation? I would like to offer my opponent an argument that doesn’t start by rejecting the cause/condition distinction nor presuppose that counterfactual dependence is sufficient for causation. This is not an argument that I seek to defend. It is rather one that I shall use to expose a weakness in the support for the view that responsibility entails causation. I suggest a way to establish the link between responsibility and causation within a consequentialist moral theory. The strategy is to argue against \( R \Rightarrow C \) by showing that the best argument for it falls short.

4.2.2 THE BEST AVAILABLE ARGUMENT FOR \( R \Rightarrow C \)

Firstly, let’s take moral responsibility to consist in the appropriateness of a set of attitudes that include praise, blame, respect, gratitude, and resentment, amongst others, to actions. Blame, for instance, is appropriate when an agent acts wrongly. Such moral attitudes, moreover, might be self-directed or directed at other’s actions. This allows a very broad construal of what we mean by moral responsibility.
An agent A is blameworthy (or praiseworthy) for some action C only if C was the wrong (or right) action for A to do in the circumstances; moreover, if A’s action C is blameworthy (or praiseworthy) for some event in the world E, then C is wrong (or right) in virtue of E.

Holding a person responsible for things other than their actions, i.e. events, outcomes or states of affairs in the world, is justifiable only if we can explain why their actions are bad in terms of those events or outcomes.

The next step is to hold that what makes an action morally right or wrong are the causal effects of that action. This is the step that requires an appeal to consequentialist view about morality according to which the moral status of action is determined by its outcome or consequences.

If C is morally right or wrong for A to do in the circumstances, then C is morally right or wrong in virtue of C’s outcomes or consequences alone.

The premise (2) leaves open that the action-consequence relation is the cause-effect relation and that by outcomes we mean just those effects that are produced by the action and not those that are merely allowed to occur. Nonetheless, it is a necessary assumption for (R→C) that the relation of action and consequence is a causal relation. To make this explicit, what is required is the following further premise.

E is an outcome or consequence of action C only if C causes E.

Without (3) it would be possible for an agent to be responsible for some outcome without it being the case that the agent caused the outcome; and it would be possible, in particular, to suggest that one is responsible by omission for some outcome without subscribing to the possibility of causation by omission. Omissions might be actions that have outcomes but not causes that have effects if we consider allowing something to happen as an act-consequence relation.
The foregoing principles can work to motivate the Responsibility Argument and the crucial assumption (R⇒C), that a person is responsible for the outcomes of their actions only if they cause those outcomes. To illustrate, consider the following example.

(B1) Billy witnesses a hit-and-run one afternoon. Suzy, who was out riding her bike, is hit by a careless driver who then flees the scene. Billy fails to call for an ambulance.

Commonsense would find Billy’s failure to call for an ambulance morally blameworthy. If we accepted all three principles above, the theory would agree with commonsense only if the omission is considered as a cause of some bad outcome. Firstly, by (1), failing to call an ambulance is blameworthy because it’s wrong. According to (2), the only thing that would make the action wrong is that it gives rise to some outcome that is in itself bad. This might be the state of affairs of Suzy being in protracted pain, on a view that pain makes for moral disvalue. But according to (3) an action can only “give rise to” an outcome by causing it. So, if Billy is to be justifiably blamed and rightly held responsible in the circumstance, then it must be because he caused Suzy to be in a protracted state of pain. One may then point to the claim that Billy’s failing to call the ambulance caused Suzy’s pain as the only reasonable explanation of Billy’s blameworthiness.

Some options for response are the following. Firstly, we may reject the idea that Billy is blameworthy even if this conflicts with ordinary commonsense. Secondly, we may deny the connection between blame and wrongness and reject some part of principle (1). For instance, we might subscribe to a view about praise and blame according to which Billy is blameworthy not because he did something wrong, but, for instance, because it is overall socially useful to blame Billy. That is, it’s right to blame Billy because doing so might encourage others to call ambulances in similar circumstances. Thirdly, we can reject (2) and hold that actions can be wrong without it being the case that they gave rise to a bad outcome. This is the response that will be pursued presently. Fourthly and finally, of course, we may reject (3) by suggesting that action and consequence is not cause and effect.
4.2.3 THE PROBLEM WITH THE BEST ARGUMENT FOR (R⇒C)

Of particular interest is (2), the consequentialist doctrine about right- and wrong-making. I will accept (3) while noting that it has been challenged by some consequentialists, (for instance, Sosa (1993)).

It is assumption (2) that, I think, is driving Sartorio’s view that (R⇒C) is an intuitive and commonsensical view. Consequentialism is one of the dominant views about ethics. It is no surprise that many would therefore accept that responsibility requires causation. But a closer look reveals that (2) is not a widely accepted statement of consequentialism.

Canonical versions of consequentialism hold that what makes action right or wrong are not the effects of the action alone, but whether that action was the best action that could have been undertaken in the circumstance. G. E. Moore’s *Principia Ethica* offers a classic statement of the view that whether an action is right depends on the effects of alternative possible actions. Moore wrote:

> to assert that a certain line of conduct is, at a given time, absolutely right or obligatory, is obviously to assert that more good or less evil will exist in the world, if it be adopted than if anything else be done instead. But this implies a judgment as to the value both of its consequences and those of any possible alternative (Moore, 1903, 25).

More recently, Sosa (1993) defends the following statement of consequentialism.

(C) It is *right* for S to do A (*S ought to* do A or *S should* do A) iff no total state of affairs that would be a consequence of S’s doing any alternative to A would be better than the total state of affairs that would be a consequence of S’s doing A.

Consequentialism is thus commonly defended alongside a comparative principle that the right action depends on how an action compares against a set of alternative possible actions. It’s what allows us to say, I will argue, that one is wrong (or right) not by producing a bad (or good) outcome, but by the fact there was a *possibility* of producing a good (or bad) outcome. It allows us to say that if an agent does A rather than B, then the omission of B is made right or wrong not
by causation, but because, to use the language of economists, of the “opportunity cost” of doing A.

It should be noted that the argument I am making works only on some versions of the comparative principle. But the argument works on the most plausible comparative principles. There are different types of comparative principles that consequentialists may employ. An optimizing principle says that a right action is one that is at least as good as any other possible action. (Moore, in the quotation above, employs a dominance version of the optimizing principle that requires that the action be strictly better than any other possibility.) A maximizing principle, adopted by Sosa, says that a right action is one that is not worse than any other. The difference between the two principles concerns the possibility that there are actions that cannot be compared. That is, if there are two possible alternative actions A and B such that A is neither worse, better nor as good as B, then, neither A nor B can be the optimal action. By contrast it’s possible for either A or B to be a maximal action if no other action is better than it. (In contexts of noncomparativism where actions are only partially ordered, optimizers would hold that there might be no right action, while maximizers would hold that each maximal element is a right action.)

Now, as well as optimizing and maximizing, some consequentialists may opt for a satisficing principle. Rather than comparing the value of an action against the alternative actions, satisficing works by requiring that a right action have a value greater than some threshold. Whether satisficing counts as a comparative principle, or whether it is compatible with (2) that measures the moral value of an action by the consequences of that action alone, depends on how satisficing is understood. In particular, whether satisficing is to be considered a way of rejecting (2) depends on how the threshold value is determined. On the one hand, the threshold that must be met for an action to be right could be determined relative to each action. That is, the relative satisficer sets a different threshold for each action. To determine the threshold, it seems plausible that the alternatives to the action will be considered. In order to say whether an action is good enough, the relative satisficer will consider the alternatives to that action and decide on a threshold depending on how good all the other alternatives are. The difference between the relative satisficer and either
the optimizer or maximizer is that the relative satisficer won’t necessarily choose only the action that fares best as the right one to do.

But it’s possible to have a version of satisficing that doesn’t require considering the alternative possible actions. Absolute satisficing may say that a right action is one that has some value greater than T, where T is the threshold set for all actions and does not, even implicitly, depend on the alternative possibilities. Perhaps an example of an absolute satisficing consequentialist is a view that an action is right only if it produces some pleasure. Clearly such a version of consequentialism would be problematic. It’s unclear than anyone would defend it. However, I highlight it to note that just because it’s important to consider the value of an action in terms of its betterness, this doesn’t imply that alternative possibilities to that action are implicated. Insofar as either optimizing, maximizing and relative satisficing are the most plausible ways of defending consequentialism, our argument will stand.

The crucial result is that many versions of consequentialism would accept that an action might be wrong for an agent to do because there is a better alternative, and not just because of the goodness of the actual consequences of that action alone. This leads to a denial of (2), that what makes an action right or wrong are its consequences alone. Rather, actions are made right or wrong by their consequences and the consequences of other alternative possible actions. This undermines consequentialist support for (R⇒C).

Consider the following pair of examples.

(J1) Jim is sitting on a park bench enjoying a delicious sandwich. There is nothing troubling going on around him.

(J2) Jim is sitting on a park bench enjoying a delicious sandwich. Suzy asks Jim for his sandwich. She wants to use it to lure her dog out of a hole. Suzy would be very happy to be reunited with her dog. There is no other way to get the dog out in the circumstances. Jim however ignores her request and continues to eat his sandwich and refuses to help.

In (J1), Jim’s eating his sandwich is not morally impermissible. In (J2), by contrast, Jim ought
not to eat his sandwich and ought to instead give it to Suzy to help her get her dog since the value of the outcome of giving the sandwich to Suzy outweighs the value of the outcome of eating it.

In (J2), commonsense morality returns the response that Jim is blameworthy for not giving up his sandwich. The Responsibility Argument would say that here we have an instance of causation by omission. We can only rightly blame Jim if by not giving up his sandwich, he brings about some bad outcome. But the point about comparative judgment is that this is not true. We can rightly blame Jim without it being the case that not giving up his sandwich (continuing to eat the sandwich) brings about a bad outcome and, in particular, without it being the case that eating the sandwich and not giving it to Suzy has, as an action outcome, that Suzy’s dog is hurt. This is because there is something that Jim could have done that would have been better. Continuing to eat his sandwich is wrong because it’s not the best thing he could have done in the circumstance of the (J2) scenario.

We can reformulate the dialectic another way. The issue can be seen to concern the proper sense of “outcome.” We ask, what must count as an outcome in order for the consequentialist theory to agree with commonsense morality in these simple cases? Suppose we need to treat the event of the dog being trapped in (J2) as a consequence of Jim eating his sandwich and refusing to give it to Suzy. Then it seems that the proper sense of outcome must be appropriately wide. (This imposes constraints on the nature of causation. By (3), which identifies the action-outcome relation as the relation of causation, the causal relation must be correspondingly wide. It would be the type of relation that allows for causation by omission.)

By contrast, suppose we hold that outcomes must be construed narrowly according to the relation of causation as production. This is an intrinsic relation that does not allow for causation by omission. Then it seems that the event of the dog being trapped would not count as an outcome of Jim’s not surrendering his sandwich. Jim doesn’t cause the dog to be trapped; so the dog being trapped is not a consequence. Eating the sandwich might have no bad consequences. It just satisfies Jim’s hunger, which, in itself, is a good thing.

In (J2), the commonsense moral judgment holds that Jim acted wrongly by choosing action A,
eating his sandwich, rather than action B, giving up his sandwich. According to consequentialism, action A is wrong in virtue of its outcomes compared with the outcomes of the alternative action B. Supposing for simplicity that B is the only possible alternative to A, then, action A is wrong only if the value of the outcome of action B would have been greater than the value of the actual outcome of action A. So, in order to agree with the commonsense verdict, we need to construe outcomes such that A’s outcomes are outweighed in value by B’s outcomes.

Those who would have us accept that the concept of cause admits of causation by omission can make their case by showing that the narrow construal of outcome would not yield the result that B’s outcomes are better than A’s and that only a more permissive causal relation gives us the right result. But in (J2), it’s easy to see that the narrow construal of outcomes based on a not-too-permissive relation of causation as production is adequate. According to the decision matrix below, action B is better than action A on the narrow outcome construal because the freeing of the dog has greater value than whatever pleasure Jim might derive from eating the sandwich. We do not need to assume a wide outcome notion to get the correct result that action A was the wrong thing for Jim to do.

<table>
<thead>
<tr>
<th>Action</th>
<th>Narrow Outcome</th>
<th>Wide Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Eat sandwich</td>
<td>Enjoys sandwich</td>
<td>Enjoys sandwich &amp; dog is trapped</td>
</tr>
<tr>
<td>B: Give sandwich to Suzy</td>
<td>Dog is rescued</td>
<td>Jim is hungry &amp; dog is rescued</td>
</tr>
</tbody>
</table>

To recap this section, I argued that the view that responsibility entails causation (R⇒C) is best understood in the context of consequentialist ethics. I then pointed out that most consequentialists would appeal to a maximizing, optimizing or a (relative) satisficing notion of moral rightness. What makes an action right or wrong is understood in terms of how its consequences compare with those of the alternative actions. I claimed that this shows that we do not need to assume a wide concept of the action-consequence relation. Instead, the relation can be understood more narrowly in a way that accords with the notion of causal production, a notion that does not admit of causation by omission. I suggested that this would give us an adequate version of consequentialism. If this
is right, then we have undercut the consequentialist support for \((R\Rightarrow C)\): sensible consequentialism does not require causation by omission.

4.2.4 A PROBLEM FOR NARROW ACTION-CONSEQUENCE: PREVENTION

It is necessary to investigate the adequacy of the narrow action-consequence relation further. Do narrow outcomes suffice to give us the right results in all cases? The narrow interpretation of outcome will have trouble with cases which mix in prevention, and where we want to say that a person does good (or harm) by preventing some event. This section poses the problem; the next will present a solution that lets us keep the narrow relation.

The following variant of \((B1)\) above will allow us to see the trouble more clearly.

\((B2)\) Billy is walking to a lunch date. He sees a drunk driver on the street dangerously close to where Suzy is cycling. Billy can shout out to Suzy to watch out and cause her to move out of the way of the car. This would interrupt Suzy’s journey and prevent her serious injury. It would also make Billy ten minutes late for lunch. As it happens, Billy does not intervene. Instead, Billy hurries to his lunch appointment to arrive on time.

The difficulty this case raises for the narrow interpretation of outcome is that shouting out to Suzy would not have brought about—in the narrow sense, i.e. by production—a positive outcome that can be considered to be good in itself. It would have interrupted and delayed Suzy’s journey. This, in itself, isn’t a good outcome. Rather, it’s only valuable insofar as a way to avoid being hit by a drunk driver. To see why, consider that without the reckless driver on the road, the fact that Suzy is off the road isn’t necessarily good.

It’s not clear, then, that Suzy getting off her bike outweighs the value of Billy’s getting lunch on time. If we consider only the outcomes in the narrow sense, it’s not obvious that walking away is morally wrong. This conflicts with our intuitive judgment. By contrast, walking away has the outcome (in the wide sense) of Suzy getting hit by a car insofar as failing to prevent an event is a way of causing it on the wide sense. If this is right, then we have a reason to prefer the
“wide” interpretation of outcomes and a permissive relation of causation. In this case, only the wide outcome interpretation gets the moral judgments correct.

<table>
<thead>
<tr>
<th>Action</th>
<th>“Narrow” Outcome</th>
<th>“Wide” Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Walks away</td>
<td>Gets to lunch on time</td>
<td>Gets to lunch on time &amp; Suzy gets hit by the car</td>
</tr>
<tr>
<td>B: Calls out</td>
<td>Late to lunch &amp; Suzy stops cycling</td>
<td>Late to lunch, Suzy stops cycling &amp; does not get hit by a car</td>
</tr>
</tbody>
</table>

Action A is wrong. But the values of the narrow outcomes have it that A’s outcome is better and not worse than B’s. Getting to lunch on time is better than the alternative. So a consequentialist theory interpreted with a narrow action-consequence relation gets things wrong insofar as it judges action A to have been the right thing to do. By contrast, on the wide/permissive relation, A’s outcomes are worse than B’s. Suzy getting hit by a car is simply worse than the alternative.

The issue at hand concerns the notion of causation by prevention. If we need to include among outcomes of an action what was prevented or what would have been prevented in order to get the right results, then it seems that the appropriate causal relation has to be a permissive one. Causation as production does not allow for causation by prevention for the same reasons it does not allow for causation by omission. Absences cannot be causal relata, neither effects nor causes. Moreover, prevention is not intrinsic. To see that prevention is not intrinsic, consider a duplicate of the interaction between Billy and Suzy in a context where there is no driver on the road. In that case, calling out to Suzy and her subsequent moving out of the way does not constitute prevention of injury.

To solve the problem for the narrow relation, I want to suggest that what was prevented or would have been prevented plays a role in determining the value of the outcomes of an action. We do not need to treat prevented events as outcomes; we only need to see that events prevented by an outcome can make that outcome better or worse. We should say that, even though Billy’s calling out to Suzy to move aside only causes her to stop cycling (in the narrow sense), the value of Suzy interrupting her journey is very high. The value of that interruption takes into account that
the fact that, otherwise, Suzy would have been injured. If this is right, then we can say that A’s outcomes are outweighed by B’s outcomes, even when those outcomes are narrowly construed.

4.3 Counterfactuals and Extrinsic Value

In evaluating an outcome, we are concerned not just with the intrinsic value of an event, but with difference in value between the actual event and what-would-have-been instead. We contrast the event that obtains with the event that would have obtained instead. This is how we can best understand preventative value, as extrinsic value of actual events in light of what would have been in their place. This involves counterfactuals, and a complete account of moral responsibility will rely on counterfactuals.

I will suggest a view on which counterfactuals are relevant in determining the values of outcomes, but in a way that is importantly different to the application of counterfactuals that comes with the counterfactual theory of causation. On my account, the role of the counterfactual is to determine value of an event; I shall defend a counterfactual theory of valuation. By contrast, on a Lewisian, counterfactual theory of causation, counterfactuals are relevant in determining the relations between events, an between outcomes and actions. This view uses counterfactuals to model the causal relation and says what are the consequences of actions prior to evaluating those outcomes. The view I will defend does not deny that counterfactuals are needed to understand moral responsibility. Our disagreement with the Lewisian use of counterfactuals concerns their place: rather than fixing facts about causation, counterfactuals are relevant to whether and to what extent an outcome is good or bad.

The structure of prevention is one where some event x prevents y from occurring. Suppose that y is a bad state of affairs. Then preventing y is a good thing to do; x would be good in virtue of preventing something bad. Now, we can understand the value of x in virtue of preventing y in two ways:

1. There is a causal relation between x and the absence of y. That is, x causes not-y. The causal
relation holds in virtue of the Lewisian counterfactual not-x$\Box\rightarrow$y. The value of a cause is a function of the value of its effect. Therefore, the fact that x is good is explained by the fact that the state of affairs not-y is good.

2. There is a causal relation between x and some positive event z, i.e. x causes z. If z hadn’t occurred, then y would have occurred in place of z. That is, not-z$\Box\rightarrow$y. In this case z does not prevent y; rather, as we shall presently explain, z can be said to preclude y. The fact that x is good is, as before, derived from the value of its effects (i.e. z). The fact that z is good is, in turn, explained in terms of the fact that z is better than y.

On the face of it, the two views give us the same result, namely that it is good to do x in order to prevent some bad outcome y. It might be thought that it is theoretically simpler to adopt a Lewisian view that treats causation, prevention, and omission in the same way in terms of counterfactual dependence. But there are prima facie reasons to prefer the view that counterfactuals are used to determine values only.

Consider the implications of each view for the metaphysics of objects, properties, and relations. On the first view, absences are things that are caused, i.e. x causes an absence not-y. Moreover, absences are assigned a value, i.e. the absence not-y is determined to be good. The second view avoids treating absences as causal relata or as bearers of value. Rather, causation holds between only positive events. And rather than ascribing intrinsic value to an absence not-y, we are ascribing an extrinsic value to the existing entity z.

It seems that we should, all things equal, prefer a metaphysics that does not commit us to treating non-existent entities as things that can stand in relations (such as causation) or capable of having properties (such as value). It creates a conceptual difficulty to accept such a view and requires a non-standard view about property instantiation or causation. For instance, it may require us to think, contrary to commonsense, that the relata of causal relations are facts rather than events. Or, as Lewis (2004) argues, that causation is not (always?) a relation. Or, as has been suggested by many, that the things that have value are states of affairs rather objects, events,
or entities that constitute states of affairs. If option 2 is an adequate way of understand prevention and omission in the context of responsibility, then why would we want to be encumbered by the object/property confusions of the option 1?

4.3.1 BADER ON EXTRINSIC VALUE AND PRECLUSION

Let us turn to consider more precisely option 2, which involves a theory of value that involves the notion of preclusion. The theory follows closely a view advanced by Bader (2015b). It will be helpful to expound some of the details and distinctions of Bader’s framework.

1. Intrinsic Value. The intrinsic value of an object is value that the object has in virtue of how it is in itself. That is, intrinsic value is grounded in the intrinsic properties of an entity.

2. Extrinsic Value (Local). Extrinsic value is not fully grounded in how things are in themselves. There are two kinds of sources of extrinsic value. There is extrinsic value in virtue of some positive relation, i.e. when the value of x is grounded in some relation xRy, where both x and y exist. Such extrinsic value is determined “locally” in terms of the facts about x and y. It does not depend on “global” facts, i.e. facts about the world where x is located. Put differently, the value of x is extrinsic because it is derived from an extrinsic object (i.e. y) and it is local because it is derived by considering only the local relation R. An example of local extrinsic value is instrumental value, the kind of value actions have in virtue of their outcomes. Where y is good, and x produces y, x is instrumentally good. Instrumental value is the value that an object has in virtue of causing something valuable to exist.

3. Extrinsic Value (Global). The extrinsic value of an object might be global if it depends on facts about the entire world and not some specific relation between parts of the world. For instance, an object may be valuable because it is rare or the only one of its kind. What makes something rare or unique is not a matter of standing in some relation to another object. Rather it concerns what there is in the world, an existential claim quantifying over an entire domain.

95
Extrinsic value of the local kind gives rise to a way of computing value on the model of what Bader calls “value transmission.” This is what is at the heart of consequentialist reliance on causation to ascertain the moral status of actions. Explaining value transmission, Bader writes:

The value of one thing gets transmitted to another thing to which it stands in some relevant asymmetric relation R, whereby both intrinsic and extrinsic value can be transmitted.

\[
x R y \rightarrow [V(y) \Rightarrow EV_R(x)]
\]

(Bader, 2015b, 1).

According to the value transmission formula, if x stands in some relation R to y, then whatever value y has accrues to x as a type of extrinsic value. That is, if y is valued in some way, then x ought to be valued accordingly insofar as x stands in the relevant relation to y. The key example of causation helps us to understand the general value transmission formula. If x causes y, then x has instrumental value proportionate to the intrinsic or extrinsic value of its effect y. If y is intrinsically good, then x is instrumentally good. If y is intrinsically bad, then x is instrumentally bad.

The value transmission model is general in that it applies to other kinds of asymmetric relations besides causation such as relations of signification (x is a sign of y) or composition (the xx’s compose to make y).

Things can have value as a cause (= instrumental value), value as a sign (= signifi-
catory value), value as something that composes (= contributory value), etc. (Bader, 2015b, 2).

For Bader, prevention is a complex relation whose components are a relation of causation and a relation of “preclusion.” Where x prevents y, this is to be understood as x causes some z which precludes y. Whereas causation is a local relation between two existing entities (i.e. two events in the actual world), preclusion by contrast is not a local relation. Rather it holds between an existing entity z and a non-existing entity y that is precluded by z. Whereas x has a local type of extrinsic value, namely instrumental value, in virtue of the causal relation, i.e. in virtue of producing z, z in turn has a non-local extrinsic value, “preclusion value,” in virtue of precluding y. The prevention
value of x is derived in two steps. First z gets its preclusion value from precluding y. Second x gets its instrumental value from causing z. Remember that the instrumental value is equal to whatever kind of value z has, including its preclusion value from precluding y.

What is preclusion?

... preclusion is to be analysed as x exists and y, which is incompatible with x, does not exist but would exist were it not for x. The precluder is, in this way, counterfactually related to that which it precludes and which would have been present had the precluder been absent (Bader, 2015b, 16).

Because of the counterfactual invoked to understand preclusion, the loose talk above of preclusion as a relation is mistaken. Relations hold between things that exist. Counterfactuals are propositions about the actual world. The truth of any counterfactual is determined by a selection function that takes as an argument the actual world. Thus, the counterfactual that if x had not existed, y would have existed, while it may appear to concern only x and y, covertly implicates global facts about the world. The counterfactual is grounded in the way the actual world is as a whole. Preclusion value should, for that reason alone, be classed as a type of extrinsic value that is more like value of things in virtue of their rarity or uniqueness than it is to instrumental value. And, while prevention may seem to be a local relation, properly understood, it is not. Rather, it is grounded in a global fact in virtue of which a counterfactual is evaluated.

The counterfactual relevant for preclusion is different from a counterfactual that is used to analyze causation and prevention on a Lewisian approach. Where x prevents y, the Lewisian approach requires us to evaluate the counterfactual not-x→y. This is a diachronic form of dependence; the antecedent of the counterfactual concerns goings-on that temporally precede those of the consequent. By contrast, Bader’s approach requires us to evaluate the counterfactual not-z→y for some z that is produced by x. This is a synchronic dependence; the two incompatible states z and y occur at the same time, albeit in different possible worlds. As Bader argues, there is a crucial difference between the “if not-x” and the “if not-z” counterfactual in terms of how they handle cases of overdetermination. This is an advantage of understanding prevention in terms of causation and preclusion.
To illustrate the difference consider the following example. A patient is dying from a fast-acting toxin and requires an immediate injection of antidote X. The antidote works when more than a certain concentration of it is present in the bloodstream, such that it can effectively block the toxin from fatally binding with certain neuroreceptors. Consider now the two following scenarios.

(D1) A doctor administers a full dose of antidote X and prevents the death of the patient.
(D2) Two doctors simultaneously inject the patient, each with a full dose of X. Both prevent the patient from dying such that the prevention is overdetermined.

In (D1), the two accounts agree that the doctor prevented the patient’s death. The Lewisian evaluates the counterfactual: had the doctor not injected the patient with X, the patient would have died. Bader’s approach evaluates a different counterfactual. Preclusion is understood as counterfactual dependence between incompatible events. As we’ve described the case, the presence of the antidote is incompatible with the toxic substances binding with the neuroreceptors. Thus the relevant counterfactual for preclusion is: had there not been a full dose of the antidote in the patient’s bloodstream, the toxin would have blocked the neuroreceptors. Put differently, the antidote precludes the toxin from binding with the neuroreceptors. We have prevention when we add further facts about causation. Firstly, the doctor causes there to be enough antidote in the blood. Second, that the toxin binding with the neuroreceptors would have caused death. Thus, the doctor prevents death by causing something which precludes the toxin from causing death.

In (D2), the Lewisian requires that the patient’s survival be dependent on the actions of each doctor. But there’s no individual dependence. That is, it not true that the patient would have died had either doctor failed to give the full dose. By contrast, the preclusion counterfactual holds in (D2) just as it does in (D1). That is, it is true in (D2) just as it was in (D1) that if there hadn’t been a sufficient level of antidote in the blood, the toxin would have bound with the neuroreceptors. What’s different in the two cases is that in (D2) the presence of a sufficient level of antidote is overdetermined. Each doctor is a cause of the there being a sufficient amount of antidote; each is a producer of a precluder of something which would have caused death; so each is a preventer of
Return now to (B2) and the problem for narrow views of the causal relations. Let’s label the relevant events in (B2) as follows:

A: Billy walks away
B: Billy calls out to Suzy
L: Billy gets to lunch on time
H: Suzy is hit by a speeding car
S: Suzy stops cycling

On the narrow construal of the causal relation, the possible causal relations that arise from Billy’s actions are:

N1) A causes L
N2) B causes S

Value transmits from effects to causes. So the valuation of Billy’s action in this case is defined as follows:

\[ \text{Val}(A) = \text{Val}(L) \]
\[ \text{Val}(B) = \text{Val}(S) \]

On the wide conception of the causal relation, we have, in addition, the following possible causal relationships:

W1) A causes H; so, Val(A)=Val(H)
W2) B causes not-H; so, Val(B)=Val(not-H)
The wide conception gets the values of Billy’s options, A and B, right even if we ignore the notion of preclusion value and consider only the instrumental and intrinsic values of events. Since H, the event of Suzy being hit by a car is intrinsically and instrumentally bad, then A is bad. Moreover, since H is bad, its negation not-H is good. So, B, calling out to warn Suzy is good.

\[
\text{Val}(B) = \text{IntVal}(\text{not-H}) + \text{InstVal}(\text{not-H}) = \text{good}
\]

This gives us the right result: it would be better to do B than A. Billy acts wrongly by walking away and failing to call out to Suzy.

The narrow conception rejects both possible relations recognized by the wide relations. The first is a relation of causation by omission, while the second is causation by prevention. Neither omission nor prevention is bona fide causation. However, we can now see that, by taking into account preclusion value, we do get the result that B is better than A. In particular, we have to consider the preclusion value of S, or the event of Suzy’s journey being interrupted.

Event S precludes H. S and H are exclusive events; and it’s true that had S not occurred, H would have occurred. So, S has preclusion value that depends on the value of H. Since S precludes something bad, it is good in itself. Thus, while being stopped on the side of the road is not intrinsically or instrumentally good, it is preclusively good. Overall, it seems plausible to say that the positive preclusion value of S outweights whatever negative intrinsic or instrumental value that event has.

\[
\text{Val}(B) = \text{IntVal}(S) + \text{InstVal}(S) + \text{PrecVal}(S) = \text{good}
\]

The narrow conception is therefore able to deliver the right results, that B is the best action in the circumstance.

The difference between the wide and narrow calculations is this. The wide view says that B is good because of the intrinsic or instrumental value of what B prevents. The narrow views says that B is good because of the preclusion value of what B causes.
Imagine a world where squirrels are an endangered species because of some disease, or a world with a thriving population of pandas. Suppose both are true. Then I think we would value an individual squirrel over an individual panda. Scarcity and abundance affects value. Yet surely no individual squirrel would have derived additional value from the near extinction of its species. Nor would pandas have lost some value because of their abundance. This is not to illustrate merely that the value of the rare squirrel is not intrinsic, although this is true. The point is that it’s an extrinsic value that doesn’t seem to be “transmitted” from anywhere. In that respect, scarcity is similar to preclusion value.

The model of value transmission gives us a way to understand the computation of extrinsic value of the local variety. But not all extrinsic value fits this model. What is it, for instance, about something being rare that endows it with value? How does the absence of other things of its kind make it more valuable?

Preclusion and scarcity may, somewhat surprisingly, be understood in the very same way in terms of difference-making and counterfactuals. This mitigates the concern that, in the notion of preclusion value we’ve introduced ad hoc machinery to treat prevention and preventative value. Preclusion value is a species of a more general way in which we think about the species of value.

One way to understand how scarcity affects value is in terms of the notion of diminishing marginal utility. That is, in a situation of abundance of a particular kind of thing, the marginal value of an extra individual of that kind is often lower than the marginal value of an extra individual when it is rare. There are two ways to illustrate marginal value. The first is familiar. It involves a sequential incrementation. Suppose you are trying to sell John 10 oranges, but you can only offer one at a time. John may be willing to pay $1 for the first orange, $0.75 for the second orange, and so on. It’s obvious what explains John’s diminishing willingness to pay for an additional orange. The satisfaction or pleasure extractable from an orange is not constant. The first orange tastes good. The second might be too filling. John knows that he won’t be able to eat a
fourth or fifth orange now. There’s a risk that the sixth orange will go bad. In other words, in the
sequential case, diminishing marginal value corresponds to diminishing expected *instrumental
value*. The fifth orange simply might not produce as much pleasure as the first.

The phenomenon of diminishing marginal value can also be understood non-sequentially. You
want to rent a storage unit in a local self-storage facility to store your unused sports equipment.
Your snowshoes, golf clubs and surfboard will fit easily into a Small or a Medium unit. Both
the Small and the Medium unit will do the job. However there is only a slight marginal value to
the few extra cubic feet of space provided in a Medium. The way we might of think of this is
that both Small and Medium have the very same instrumental value insofar as they would both
produce the desired storage results. But the Medium unit “overdetermines” that result. It is more
than sufficient for storing the equipment. One might therefore be willing to trade-off the extra
space. (One might be willing to pay $1/ft\(^3\) to get 10ft\(^3\), but only $0.75/ft\(^3\) to get 15ft\(^3\).)

Something similar to the second case of diminishing marginal value in the context of overde-
termination may serve to explain changes in value with changes in scarcity. Consider again the
squirrel. Suppose that in general our interest in having squirrels in the environment has to do
with the contribution a squirrel makes to biodiversity. Let’s say then the value of a squirrel is
derived from its contribution to a biodiverse environment. Now suppose that 100 squirrels in the
neighborhood contribute to local biodiversity equally as would 1,000 squirrels. The additional
900 squirrels, let’s suppose, won’t harm biodiversity or having any other side effect. Call the
two possible populations of squirrels C (where there only 100) and M (where there are 1,000). A
random squirrel from the population C would have the very same contribution value as a random
squirrel from population M. The reason, however, why a random squirrel from M has a lower
marginal value than one in population C has to do with the fact that M overdetermines biodiver-
sity. We might be more willing to tradeoff the life of one M-squirrel than we are the life of a
C-squirrel. While a C-squirrel has the same contribution value as an M-squirrel, we wouldn’t risk
losing much by way of biodiversity if we lost an M-squirrel whereas perhaps there is such a risk
in the case of losing a random C-squirrel.
What exactly does all this have to do with preclusion value? Both involve counterfactual difference-making notions of value. When considering tradeoffs and marginal value, we are thinking counterfactually. We ask, for instance, what would happen if we lost that one squirrel. As economists might say: what is the opportunity cost to preserving the life of each squirrel? What is the opportunity cost to paying for extra unneeded storage space? Similarly, when considering the preclusion value of some event, we are considering the opportunity cost of its occurrence. What would be brought about instead and what would be the value of that alternative?

To summarize, I think that preclusion is not a special case where we invoke counterfactuals. It fits the general schema for computing a type of value that’s familiar from economics—marginal value, tradeoffs and opportunity cost. In other words, the proper distinctions in the type of extrinsic values is not between (i) transmitted value and (ii) preclusion value. That would be smack of an ad hoc distinction. Rather the proper categorization is between (i) transmitted value and (ii) marginal value. Preclusion is an instance of the latter, which is understood in terms of difference-making.

4.3.4 PRECLUSION AND PREEMPTIVE PREVENTION

In section 4.3.1 it was argued that the synchronic counterfactual implicated in preclusion does not suffer from the overdetermination problems that plague the standard, diachronic counterfactual dependence. In this section I want to propose a way in which preclusion may shed light on the problem of preemptive prevention raised by Collins (2000).

Collins discusses the following two cases:
(C1) A ball is thrown at the window. Unimpeded, the ball would have caused the window to break. As it happens, you catch the ball. But I also leapt to catch the ball and I would have done so if you had not preemptively the caught it before me.

(C2) As in (C1), a ball is thrown at the window and you catch it. This time there is a solid brick wall between you and the window rather than a second, potential catcher. If you had not caught the ball, it would have bounced off the brick wall.

Collins writes:

When I stand between your catch and the window, I am happy to say that you prevented the window from breaking. It is far less clear to me that that is the right thing to say when it is not just me but a solid brick wall that stands as the second line of defense... I am very reluctant to say that your catch prevented the ball from breaking the window (Collins, 2000, 224).

It would be good to have an explanation for the apparent asymmetry in (C1) and (C2) with respect to the status of the first catcher as someone who prevents the window from breaking and is thereby responsible for protecting the window. Standard counterfactual theories have trouble explaining the difference since the pattern of counterfactual dependence in the two cases are similar. We are forced to say that the catch constitutes prevention either in both cases or in neither case.

In terms of our account, the asymmetry may be explained in terms of the difference in the preclusion value of catching the ball. What catching the ball precludes in both cases is the ball in being motion in the air, with a certain momentum and velocity. Let’s call this state B, i.e. the state of ball, intrinsically characterized, as it would have been at the time of the actual catch if the catcher had not caught it. In both (C1) and (C2), B is precluded by catching the ball.

The difference in the two cases is that B has a greater chance of breaking the window where there’s a second catcher than in the case where there is a brick wall. The second catcher is a
human being and not as failsafe as a brick wall. What’s precluded in (C1), therefore, is an event which had some probability of destroying the window. In (C2), the precluded event had almost no chance.

Again, in (C1) what’s precluded isn’t an event that would have for sure been a cause of the window breaking. It’s simply not true that if B (i.e. the ball carrying on) had occurred, that B would have caused the window to break. So the precluded event B would not have instrumental value as a cause. However it would have value as a threat. Preclusion value in cases of preemption would be due to precluding threats and not due to precluding causes.

In a sense I’m suggesting we apply a notion of expected instrumental value, which is a generalization of instrumental value to allow for probabilistic cases. The preclusion value of catching the ball in both cases can be determined by the following formulae:

\[
\text{PrecVal(Catch)} = -\text{ExpVal(B)}
\]

\[
\text{ExpVal(B)} = \text{prob. of } B \text{ causing } E \text{ (if } B \text{ were to obtain)} \times \text{Val}(E)
\]

The probability that B causes E (in this case, E = window breaking) is greater with a second catcher than with a brick wall. Human beings are far less reliable than brick walls in stopping things. Something may have gone wrong the expected value of B in (C1) is greater than the expected value of B in (C2). So the preclusion value of the catch in (C1) is greater than in its in (C2); and therefore the catcher accrues more value in (C1) than in (C2). It is in that sense that we are inclined to praise the catcher for the action in (C1) but not in (C2).

Notice also that neither the second catcher nor the brick wall bear much, if any responsibility, for protecting the window. The backup catcher does not catch the ball. Leaping through air alone does not preclude any potential threat to the window. Similarly, the brick wall does not stop the motion of the ball and does not produce anything that precludes a threat to the window. So there is no concern on our account that the backups will be treated as responsible events. Only the actual (first) catcher precludes the ball from traveling towards the window in both cases.
4.4 Responsibility without Causation by Omission

I gave a way of understanding responsibility by omission in terms of whether an action was better or worse than the other available actions. The problem of page 90 was that the two possible narrow outcomes seem to suggest that Billy is right not to warn Suzy of the drunk driver. This was because if we looked at the outcomes narrowly, we find that the outcome of calling out to Suzy brings about only that Suzy is on the sidewalk. There is no reason to prefer the event of Suzy being on the sidewalk over, say, Billy having lunch. Clearly, Suzy’s being on the sidewalk is not intrinsically important in itself. It’s merely good insofar as it prevents her from being run over. Since the narrow outcomes thus generate the wrong results by prescribing indifference or perhaps even not calling out and walking away, it is arguable then that the required understanding of “outcome” is one that follows from a “wide” action-consequence relation that admits of absences as outcomes.

The notion of preclusion value resolves the problem in favor of the narrow outcome view. It is not correct to value Suzy being on the sidewalk purely in terms of the intrinsic value of that state of affairs. We should also take into account what is precluded by Suzy being on the sidewalk. Once we consider the preclusion value, it will turn out that that outcome is better, after all, than Billy having lunch.

Moreover, there is reason to prefer a purely causal notion of the action-consequence relation together with a notion of value that admits of a counterfactual analysis of preclusion value rather than a counterfactual notion of action-consequence and a merely intrinsic notion of value. This is because, as we have seen, Bader’s account gets cases of overdetermination right, which are counterexamples to the alternative view.

I conclude that responsibility does not entail causation. We can make sense of moral responsibility, even in cases where one is held responsible for an omission of an action or where one is responsible for what the prevent, without a matching notion of causal responsibility. The Responsibility Argument fails.
Chapter 5

CONCLUSION

There’s an intuitive plausibility to idea that causation is a matter of “difference making.” This may explain why, despite all the signs that the counterfactual theory of causation is a degenerating research project, many philosophers continue to think that counterfactuals are a guide to causation. For instance, Menzies (1996), speaking more generally in terms of counterfactual probability-raising (i.e, if C had not occurred, the chance of E occurring would have been lower than it actually is) rather than counterfactual dependence (if C had not occurred, E would not have occurred), writes:

The third crucial platitude... relates the causal relation to the relation that exists between two distinct events when one event increases the chance of the other... [T]he causal relation does not always coincide with this relation of increase of chance between distinct events. Nonetheless, it is a striking fact that the two relations coincide for the most part. Aside from cases involving pre-emption and overdetermination, one event causes another event just when the two events are distinct and the first event increases the chance of the second event (Menzies, 1996, 100).

The dissertation offers a perspective for a reappraisal of the relation between counterfactuals and causation. Most counterexamples in the literature on counterfactual theories of causation almost always concern the necessity of counterfactual dependence, that if two events are causally related, then had the cause not occurred, the effect would not have occurred. In cases of pre-emption, for instance, we have causation without dependence. Had the preemptive cause C1 not occurred, the effect E would have still have occurred thanks to the preempted event C2. Similarly,
in cases of overdetermination, had one of the overdetermining causes C1 not occurred, the effect E would have obtained thanks to the presence of the other overdetermining cause C2. In these cases, we have causation without dependence.

In the course of the preceding chapters, I argued that we should surrender commitment to the sufficiency of counterfactual dependence for causation. We should reject the following principle:

*Sufficiency of Counterfactual Dependence*: Necessarily, two events C and E are causally related if E would not have occurred had C not occurred.

In chapter 2 (sec. 2.6), I argued that giving up the thesis would allow us to resist the arguments of Prior et al. (1982) that every disposition has a causal base. In chapters 3 and 4, I argued that the accepting Sufficiency conflates causes, on the one hand, and enablers, threat-cancelers, bystanders, and omissions on the other.

We do often argue that an event is not a cause because it doesn’t make a difference. Consider, for instance, the case of switching train tracks due to Hall (2000).

“The Engineer”: An engineer is standing by a switch in the railroad tracks. A train approaches in the distance. She flips the switch, so that the train travels down the right-hand track, instead of the left. Since the tracks reconverge up ahead, the train arrives at its destination all the same; let us further suppose that the time and manner of its arrival are exactly as they would have been, had she not flipped the switch (Hall, 2000, 205).

As Hall explains, there is a strong temptation to react by saying that the engineer does not cause the train to arrive at the destination:

“After all,” goes this reaction, “isn’t it clear that the switching event makes no difference to whether the train arrives, but merely determines the route by which it arrives?” *(ibid.)*

But it is a mystery why the same argument doesn’t work in cases of preemption and overdetermination. Suzy’s rock reaches the window before Billy’s. She causes the window to break. We are not then inclined to argue: “Suzy doesn’t cause the window to break. After all, the event
makes no difference to whether the window breaks but merely to whether her rock or Billy’s destroys the window.”

I made the case in chapter 3 that counterfactuals play a role in the auxiliary notion of causal explanation but not necessarily in the notion of causation itself. It is worth distinguishing three senses of “difference making.” This I think will allows us to see why arguments from difference-making are effective in cases of switching, like Hall’s “The Engineer,” but not in cases of overdetermination and preemption.

There is, in the first place, a sense of difference making as it relates to the causal relata. At the level of causal relata, difference making serves as a way of individuating events. What is caused, say, by Suzy’s throwing a rock at the window is that the window breaks as opposed to being intact and safe. The difference in question between a broken window and an unbroken window is relevant only to understanding what the effect of Suzy’s action was. There is difference making, in this sense, even in cases of overdetermination. That is, suppose that Suzy and Billy each throw a rock that hits the window at exactly the same moment. Each may be said to cause the window to break, and each may be said to have made a difference to the breaking of the window. This is so even though the breaking of the window does not depend on either Suzy or Billy throwing their rocks. They each cause the window to be broken as opposed to intact: the difference in question concerns those two possible states of the window.

Typically, however, difference making is taken to correspond to the causal relation rather than the individuation of the relata. The relation between cause and effect is one such that without the cause, the effect would not have occur. In this sense, “causation as difference making” corresponds to a conception of the relation as a kind of dependence that may be understood by contrasting the actual situation to hypothetical alternatives. The causal relation is understood by considering what would happen to the effect if we changed the initial causal conditions. Lewis’s influence theory may be construed as still operating within the bounds of a difference making notion of the causal relation but dropping the commitment to a difference making notion of the causal relata. This is because the truth-conditions of the claim that C causes E involve contrasting
actual and non-actual variations.

Cases of preemption are typically counterexamples to causation as difference making in the second sense relating to the relation. Preemption cases turn on the possibility that the relevant variant scenarios might not involve any variation to the effect. If Suzy hadn’t thrown her rock, the very same effect would occur. More sophisticated cases of preemption might make the case more robust by showing that there is no variation to how or when Suzy throws her rock would result in any difference to the effect.

Keeping clearly distinct the two foregoing sense of “difference making” is important to see exactly how so-called “productive” conceptions of causation contrasts with “dependence” conception of causation. Both views might well agree on the individuation of the relata in terms of the contrast between the event occurring rather than not occurring. Views of causation as production are best understood as denying the second sense of the difference making conception of causation but not necessarily the first.

I want to suggest there is a third level of difference making as it relates to causal histories. An event might be relevant to whether the effect has the causal history that it actually does. If C had not occurred, then E would not have had such-and-such events in its causal history. It is this kind of difference making which I think is at play in the argument that the engineer’s hitting the switch does not cause the train to arrive at its destination. It also explains why a similar argument doesn’t work in cases of preemption or overdetermination.

In the Engineer case, had the engineer not hit the switch, then the only difference there would have been to the causal history of the train arriving at its destination concerns a small segment of that event’s causal history. It would have meant that the train travels down the left hand track as opposed to the right hand track. But almost everything else that is actually in the causal history, such as the train driver waking up on time and drinking his coffee, the train making its way through all the stations that came before the track-adjustment, etc., remains part of the causal history in the counterfactual scenario where the switch is not hit.

Compare with the preemption case. Had Suzy not thrown the rock, it would have made a much
bigger difference to the causal history of the window breaking. Not throwing the rock wouldn’t make a difference to whether the window breaks. But it changes the causal history of the event from being one that’s produced by Suzy, to being one that’s produced by Billy. Given that Suzy throwing the rock itself has a different set of causes from the event of Billy throwing his rock, this will result in a much bigger change in the causes of the window breaking if Suzy doesn’t throw her rock. Suppose for instance, that Suzy is in the location where she throws the rock because she cycled to meet Billy, whereas Billy walked there. Then if Billy’s rock had hit window instead of Suzy’s, this would have switched the event of Suzy cycling out for the event of Billy walking as one of the causes of the window breaking. Thus, if Suzy had not thrown her rock, a swathe of events that are part of the actual causal history, which are mediated by Suzy’s throwing the rock, would be causally disconnected from part of the causal history of the broken window in the counterfactual scenario.

In short, we can see how in cases of preemption, the supposition that the preemptive cause does not occur will have broad implications in terms of the causal history. But in cases of switching, the “switch” event makes a relatively little difference to the causal history. Insofar as difference making is an important part of our everyday causal reasoning, it is the notion of difference making to causal histories and not the notion of difference making as a causal relation that is relevant. Far from being a constitutive part of the causal relation, difference making in terms of causal histories presupposes the network of causal relations that constitutes a causal history. Causation is, therefore, prior to difference making.
Bibliography


Bader, Ralf M. 2015a. “Dispositionality, causality and intrinsicality.” manuscript.

—. 2015b. “Extrinsic Value.” manuscript.


