Synthese (2011) 183:409–427
DOI 10.1007/s11229-011-9870-3

# Mechanisms revisited

**James Woodward**

Received: 7 October 2010 / Accepted: 3 January 2011 / Published online: 12 January 2011


**Abstract** This paper defends an interventionist treatment of mechanisms and contrasts this with Waskan (forthcoming). Interventionism embodies a difference-making conception of causation. I contrast such conceptions with geometrical/mechanical or "actualist" conceptions, associating Waskan's proposals with the latter. It is argued that geometrical/mechanical conceptions of causation cannot replace difference-making conceptions in characterizing the behavior of mechanisms, but that some of the intuitions behind the geometrical/mechanical approach can be captured by thinking in terms of spatio-temporally organized difference-making information.

**Keywords** Mechanism · Interventionist theory of causation · Difference-making · Perception of causation

## 1 Introduction

I'm pleased to have this opportunity to discuss Jonathan Waskan's "Mechanistic Explanation at the Limit". In what follows I explore some issues dividing the interventionist treatment of mechanisms in Woodward, 2004 from the alternative suggestions in Waskan's essay. Sections 1–4 lay out the basic ideas of the interventionist framework and the contrast between two different conceptions of causation, which I call difference-making and geometrical/mechanical. Sections 5–8 then comment on Waskan's ideas.

It is widely agreed that in many areas of science information about "mechanisms" plays a central role: the identification of mechanisms is a major goal of theory construction, information about mechanisms is important in causal explanation and so

J. Woodward (✉)
University of Pittsburgh, Pittsburgh, PA, USA
e-mail: jfw@pitt.edu

on. I take this to be common ground among virtually all who discuss the role of mechanisms in science, myself included.

Disagreement ensues, however, when one asks how the notion of mechanism is best understood or elucidated. One problem is the notion is used in a variety of different ways or at least has amorphous boundaries. At one end of a spectrum, there is the notion that is in play when one thinks of "mechanical" reasoning involving certain simple machines (gears, pulleys, levers) and which is associated with such "mechanical" philosophers as Galileo and Descartes. Here causal influence is conceptualized in terms of "contact" forces[1] that require something like spatio-temporal contiguity and the focus is on such properties as shape, weight, rigidity, impenetrability, position, energy and momentum. At the other end of the spectrum, identifying "mechanisms" seems to mean little more than "explaining in terms of whatever entities and causal relationships are taken to be fundamental in the relevant domain of investigation, perhaps with the expectation that this often involves the identification of parts or components in some larger structure, and an account of how their interaction produces (or explains) some overall input/output relation". Discussions of "mechanism design" in economics arguably illustrate this latter notion of mechanism—the theorist looks for rules or institutions (e.g., governing different sorts of auctions) that will achieve outcomes that are normatively optimal and where the explanation of the behavior of these institutions is in terms of the interactions of the individuals (the parts) that make them up. In addition, we find writers talking about the "mechanism" of natural selection, where even the notion of decomposition into parts may not have an obvious application. Needless to say, whatever is meant by "mechanism" in this context, it probably has little to do with the identification of structures involving contact forces. As still another possibility, chemists regularly talk about chemical reactions in terms of the mechanisms, where the forces are associated with these reactions, represented in terms of various claims about chemical bonds, are at bottom electromagnetic in nature, but where intuitions about spatio-temporal contact can be misplaced if taken literally.

Given this diversity, I believe it is a mistake to look for an account of "mechanism" that will cover all possible interpretations or applications of this notion. A better strategy is to try to elucidate some core elements in some applications of the concept and to understand better why and in what way those elements matter for the goals of theory construction and explanation. In what follows I pursue this strategy, focusing, among other considerations, on the role of spatio-temporal information in the characterization of mechanisms, but without any suggestion that such information is crucial to all mechanism talk.

I begin with a contrast between two different ways of thinking about causation which I label Difference-making (DM) and Geometrical/Mechanical (GM). "Interventionist" treatments are one possible variety of DM account and I believe Waskan's judgment that such treatments fail to do justice to the GM aspect of causation underlies his resistance to them. After explicating the GM vs. DM contrast and explaining how

[1] It is worth emphasizing, though, that the modeling of such machines often proceeds by radically abstracting away from the details of the operation of contact forces, which are both complex and often unknown.

it figures in debates over the notion of mechanism, I turn to a look at some of Waskan's suggestions and then to some positive proposals of my own.

## 2 Difference-Making versus Geometrical/Mechanical Accounts of Causation

### 2.1 Difference-making accounts of causation

A guiding intuition of many treatments of causation is that causes make a difference to their effects. There are many ways of spelling out this idea; in terms of INUS conditions, in terms of the idea that the probability of the effect conditional on the occurrence of the cause is different from the probability of the effect conditional on the absence of the cause when other factors are appropriately controlled for, and in terms of counterfactuals. Interventionist accounts are a version of a difference-making account in which the notion of difference-making is spelled out by means of counterfactuals concerning what would happen to an effect if certain interventions on the cause were to occur.

DM accounts share several common commitments. One is that causal claims involve a *comparison* of some kind between what happens in a situation in which a cause is present and alternative situations (which may be actual or merely possible) in which the cause is absent or different. For example, when ingestion of a drug is said to cause recovery from a disease, DM treatments take this to imply a comparison between what happens to subjects with the disease who ingest the drug and otherwise similar subjects who do not ingest the drug. Both groups may actually exist, as would be the case if an experimental trial were actually performed with a treatment and control group. But on at least some formulations of the DM approach, including the interventionist version I favor, the existence of both groups is not necessary for it to be true that the drug is efficacious in promoting recovery. On such views, if, say, patients are given the drug and recover, then it will be true that the drug causes recovery as long as it is also true that if (perhaps contrary to actual fact) otherwise similar patients *were* given the drug, they would recover. In both these cases, though, the claim that $C$ causes $E$ is taken to "point beyond" the local, actual situation in which $C$ and $E$ occur and has implications for what happens or would happen in alternative situations in which $C$ does not occur.

DM accounts emphasize the role covariational or contingency information plays as evidence for causal claims—that is, information having to do with patterns of co-occurrence or correlations among different values of the variables that are candidates for causal relata, whether these are detected through passive observation or generated via experimental interventions. Such information is relevant in an obvious way to whether one variable plays a difference-making role with respect to another. In part for this reason, DM accounts seem to apply most directly to the characterization of so-called *type*-causal relationships—claims that some repeatable type of property or magnitude (variable) is causally relevant to another. It is of course possible to provide difference-making accounts of *token*-causal relationships, as in Lewis' well-known account, but in my view such theories rely, explicitly or tacitly, on type-level difference-making claims.

As remarked above, the interventionist treatment of causation in Woodward (2002, 2003) is a difference-making account. The basic idea is simple:

> (**M**) Suppose $X$ and $Y$ are variables. Then, in the simplest case, $X$ is causally relevant (at the type level) to $Y$ if and only if there is a possible intervention on $X$ such that if such an intervention were to occur, the value of $Y$ or the probability distribution of $Y$ would change—in other words, some interventions on $X$ make a difference for $Y$.

Heuristically, an intervention can be thought of as an unconfounded manipulation of the sort that might be realized in an "ideal" experiment—for details see Woodward (2003). There are many issues about how best to formulate an interventionist theory, but I will ignore most in what follows since they are not germane to the topic of mechanisms. Nonetheless, a few additional observations may be helpful. First, the employment of counterfactuals in (M) contrasts with Waskan's aspiration to characterize causal relationships and mechanisms in a way that does not appeal to counterfactuals and instead has only to do with what is "actual". Second, (M) characterizes a notion of causation that is relatively non-specific and uninformative: (M) will be satisfied as long as some interventions on $X$ in some background circumstances are associated with changes in $Y$. One often wants to know much more than this—for example, exactly which changes in $X$ are associated with exactly which changes in $Y$ and in what circumstances. Within an interventionist framework, this information is spelled out by more specific "interventionist" counterfactuals linking specific changes in $X$ to specific changes in $Y$. One way of accomplishing this is by using functional relationships – writing $Y$ as a function of $X$ and other variables ($Y = F(X, Z)$), with this functional relationship understood along interventionist lines, as telling us how $Y$ will respond to specific changes in $X$ and other variables. Such functional relationships provide difference-making information of a quantitative, rather than purely qualitative sort. Conversely, it seems hard to understand the widespread use of functional relations in science to represent causal relations unless the latter are understood in difference-making terms. I mention this because one motivation for an interest in mechanisms is the observation that if one knows only that $X$ causes $Y$ (e.g., aspirin causes headache relief), understood along the lines of (M) or in some other way, this is uninformative about much else one would like to know—about the circumstances in which $X$ produces $Y$, and so on. The demand for information about mechanisms is often (at least in part) a demand for such more specific and detailed information. As I explain below, there is nothing in a difference-making or interventionist account which is inconsistent with this demand, which can be provided by interventionist counterfactuals that are more detailed and specific than (M).

### 2.2 Geometrical/mechanical theories

Within philosophy, causal process theories such as Salmon (1984), and Dowe (2000) furnish familiar examples of GM theories but many more recent theories that focus on the role of mechanisms also seem largely in this camp. A distinguishing feature is that these approaches do not think of causal claims as implicitly comparative in the way

that DM accounts do. Instead, the underlying intuition about causation is something like this: whether there is a causal relationship between two events *c* and *e* just has to do with whether *c* and *e* occur and whether there is an appropriate "connecting" process (or mechanism) between them; moreover whether there is such a process depends just on what is actually true of the particular occasion of interest, and does not depend on what does or would happen on other occasions (for example, it does not depend on what does or would happen if a *c*-like event had occurred on some other occasion or if *c* had not occurred). In this sense, GM accounts treat causation as a "local" or "intrinsic" relation to be understood in terms of what "actually" happens, and not in terms of modal or counterfactual considerations. (These "actualist" commitments are apparent in Waskan's treatment of mechanisms—indeed, he takes them, not unreasonably, to be the feature distinguishing his treatment from DM accounts like mine.) Just which features of a causal transaction qualify as "actual" in the relevant sense is not entirely straightforward, but presumably these at least include spatio-temporal relationships and other "kinematic" properties, as well as properties like energy and momentum. It is thus unsurprising that these figure prominently in many accounts of "connecting mechanisms".

GM treatments differ from DM treatments in other ways as well. The notion of a connecting process applies most naturally to particular events or interacting entities (as implicitly assumed above when I used lower case *c* and *e* to represent the entities standing in GM causal relationships). In contrast to DM theories, GM theories thus tend to regard token causal claims as primary, although claims relating types of events may be formulated as generalizations over token claims. In addition, unlike DM accounts, GM treatments impose substantive, domain-specific constraints on what it is for one event to stand in a causal relationship to another. Suppose I omit to water the flowers and they die, where my omission makes the difference for whether they die. Suppose also (as seems uncontroversial) there is no spatio-temporally continuous connecting process leading from omission to death. Then, on a GM account requiring such connectedness, my omission cannot not be a cause of the flowers' death. As another possibility, suppose that people typically do not *represent* mental events as located in space or as spatio-temporally connected to behavior. Then, if representation of events as causally related requires representation of events as connected by spatio-temporally continuous processes, people will not represent mental events as causes of behavior.

While DM accounts assign a central role to contingency information as a source of evidence for causal claims, GM accounts commonly assign a central evidential role to spatio-temporal or geometrical relationships or to facts about the presence or absence of the appropriate sorts of mechanical properties (rigidity, weight etc.). In particular, cases in which (it appears) one can just "read off" which causal relationships are present from geometrical or mechanical properties, without any apparent need to rely on contingency information, play an important role in GM thinking about causation. A paradigmatic example, which I discuss below and which also figures in Waskan's discussion, is the so-called "perception" of causation in "launching" events. When one sees a moving billiard ball strike a stationary one, the second begin to move, and the spatio-temporal parameters governing the collision are right, it appears that one can simply "see" the collision cause the second ball to move—everything one

needs to reach this conclusion seems to be present in the motion of the balls and their spatio-temporal relation. In particular, it appears unnecessary to rely on covariational or counterfactual information about the movement of other balls on other occasions, or about how the second ball would have moved in the absence of a collision with the first in reaching the conclusion that the collision caused the movement of the second ball. Interactions of this sort are thus natural sources for "actualist" or "localist" intuitions about causation. Many other examples seem to exhibit a similarly transparent connection between geometrical structure/mechanical properties and causal relationships—think of seeing that one solid object supports another or that the movement of one gear causes a second, interlocking gear to turn.

What is the relationship between DM and GM approaches to causation? Although, as I suggest below, both capture important strands in how ordinary adults think about causation, they also seem conceptually or logically independent of each other to a remarkable degree. This apparent independence helps to generate many of the conflicting "intuitions" that philosophers have about the role of mechanisms in explanation.

First, if we interpret GM approaches as requiring a spatio-temporally continuous process or spatio-temporal contiguity between $C$ and effect $E$, this is apparently not necessary for $C$ to stand in a difference making relationship to $E$. This is illustrated by the omission example described above as well as by more complex examples involving double prevention, of which more below. As another illustration, if we think of Newtonian gravitational theory as describing a force acting instantaneously at a distance, then this theory describes difference-making relationships with great accuracy, yet without specifying any mediating mechanism. Of course, Newton himself famously thought that there must exist something playing this mediating role, but subsequent generations of physicists, failing to find such any such mediator, became increasing comfortable conceiving of gravity along purely difference-making lines, thus again illustrating the apparent logical or conceptual independence of difference-making from claims about connecting mechanisms.

Somewhat more subtly, numerous examples show that spatio-temporal connectedness (or, more generally, the existence of connecting mechanism between $C$ and $E$) is also not sufficient for $C$ to cause $E$. Suppose (cf. Hitchcock 1995) blue chalk is applied to the tip of a cue stick, which is used to strike the cue ball which in turn strikes the eight ball, which goes into the pocket because of the impact of the cue ball. Portions of the chalk are transmitted to the cue ball, and from there to the eight ball. There is a mechanical relationship in the paradigmatic sense of spatio-temporally continuous transmission of energy/momentum between the presence of the chalk on the cue ($P$) and the falling of the eight ball ($F$), yet $P$ does not cause $F$ in the difference-making sense. (We will see later that this observation creates difficulties for Waskan's proposal to analyze what is for $C$ to cause $E$ in terms of the existence of "some mechanism or other" linking $C$ to $E$.)

Another illustration: Suppose yellow fingers ($Y$) and cough ($C$) are joint effects of a common cause, smoking ($S$). $Y$ and $C$ are correlated but neither causes the other. Are there connecting processes or mechanical relationships between $Y$ and $C$? Sure. Sound waves from $C$ impinge on $Y$, transferring energy and momentum. The slight increase in mass of the fingers because of the deposit of discoloring material exerts a (minute) gravitational influence on the parts of the subject's body involved in coughing. None-

theless, whether or not $Y$ occurs does not make a difference for the occurrence of $C$ or vice-versa.

The problem in both cases is that although there is a mechanical relationship/connecting process between $F$ and $P$, and between $Y$ and $C$, the relationship is not of the right character to mediate or support $F$'s causing $P$ or $Y$'s causing $C$, in the sense of cause associated with difference-making. For example, the gravitational energy transfers associated with $Y$ and $C$ are far too minute and inconsequential to warrant the conclusion that $Y$ makes a difference for $C$. Similarly, what makes a difference for whether $F$ occurs are the details of energy and momentum communicated to the cue ball by the cue and not just whether some non-zero amount of energy/momentum or blue chalk is communicated. Moreover, while these may seem to be artificial examples, similar observations apply to more realistic cases—to anticipate a case discussed below, information about physical connections between neural structures typically does not tell us what sort of difference-making role (if any) these structures play.

In addition to this apparent conceptual distinctness, there is substantial evidence for empirical disassociations between causal claims or judgments based on GM information and those based on DM information. In an experiment discussed by Waskan (described in more detail in Woodward, forthcoming), Schlottman and Shanks (1992) presented subjects both with apparent collisions between moving objects that "looked causal" on a token-by-token basis (because the collisions exhibited spatio-temporal patterns characteristic of launching events) and with other events that did not look causal (e.g., because there was no spatio-temporal contact or the temporal delay in the movement of the second object was too great). The contingency between these events was also varied independently of the spatio-temporal parameters—for example, on some trials, even though the spatio-temporal parameters were such that interactions looked causal, the second object was no more likely to move in the presence of such a collision than in its absence.

Subjects were asked questions designed to probe whether they perceived the interactions as causal (e.g., "Did it look to you as though A hit B and caused it to move?"—Schlottman and Shanks take these to be measures of "causal perception", although they might equally well be described as measures of judgments of perceived causation. Subjects were also asked questions that Schlottman and Shanks took to reflect judgments about the existence of dependencies or contingencies at the type level (e.g. subjects were asked to judge how "necessary" collisions with A were for making B move—Schlottman and Shanks call this "judged causality"). The result was a "dissociation" between these two sets of judgments. Ratings of perceived causality "show[ed] a substantial contiguity effect but no contingency effect whatsoever" (1992, p. 337)—that is, ratings fell with increased temporal delay between collision with A and movement of B, but were uninfluenced by the contingency between collision with A and subsequent movement of B. On the other hand, judgments of causal necessity showed a "large contingency effect and a much smaller contiguity effect" (p. 337).

It is natural to connect these results with the differences between DM and GM approaches to causation.[2] On a DM theory, whether there is causation between the

[2] I should emphasize that this connection with DM and GM conceptions of causation is not suggested by Schlottman and Shanks but is rather my own interpretation.

movement of A and the subsequent movement of B has to do with whether there is contingency or dependence between these two classes of events. This is reflected in the subject's responses to questions about "judged causality". By contrast, on a GM account, whether causation is present in individual interactions between A and B requires only that the appropriate connection—in this case, spatio-temporal contiguity of the right sort—is present. This is reflected in subject's reports about perceived causality. What the Schlottman/Shanks experiments show is that these judgments can dissociate. This possibility should not be surprising in view of the conceptual/logical independence of these two different ways of thinking about causation. As emphasized above, there is apparently nothing in the idea there is a difference-making relation between $C$ and $E$, by itself, that requires that there a connecting process or mechanism between individual instances of $C$ and $E$. Moreover, there seems to be nothing in the concept of such a connecting process being present on particular occasions from which it follows that there is an overall contingency between $C$ and $E$.

## 3 Two Concepts of Causation?

What does all this imply about how we should think about causation? The logical differences between DM and GM treatments of causation and the different sorts of judgments to which they sometimes lead may seem to suggest that we operate with two distinct concepts of causation. This is the conclusion drawn (at one point) by Hall (2004). Hall distinguishes between a concept of causation as "dependence" which corresponds fairly closely to what I have been calling "difference-making" and a concept he calls "production". The latter is broadly similar to although not identical with the notion GM approaches attempt to capture. Hall thinks both concepts figure in ordinary causal thinking, although he regards the production concept as more fundamental.

However, even if Hall is correct in thinking that we operate with two distinct causal concepts, it also seems uncontroversial that these are integrated or intertwined in ordinary adult causal thinking, rather than remaining disconnected and orthogonal to one another. This is apparent in cases of casual perception. When one "perceives" that a collision with ball one has caused a second ball to move, then, at least in the absence of information about some other causal factor that might cause movement of the second ball, most people will find it natural—indeed virtually irresistible—to judge it was the collision that made the difference for the motion of the second ball and that if the collision had not occurred, the second ball would not have moved. In other words, we assume there will be a close relationship between the spatio-temporal relationships and connecting processes on which GM accounts focus and difference-making. This is reflected in the responses of subjects to the Schlottman/Shanks experiment—subjects found the dissociation between what geometrical cues seemed to indicate about causal relationships and what the contingency evidence seemed to indicate unexpected and bewildering. For example, subjects commented that on contingent problems with delay, they were aware that collision was necessary for A to move but that "it just did not look as if it should be." On non-contingent problems with no temporal delay, they commented that "they knew the collision was not necessary for B to move but it looked as if it should be". Similarly when geometrical mechanical information

tells us that one object (e.g. a plate) is supported by another objection which it rests (a table), we immediately also judge that the contact with the table is what makes a difference for whether the plate remains at rest, that if the table was removed, the plate would fall and so on. In short, despite the conceptual distinctness of GM and DM approaches of causation, we also think that in appropriate circumstances the presence of certain spatio-temporal relationships/patterns and mechanical properties can be fallible indicators of the presence of certain kinds of difference-making and that different varieties of difference-making often have characteristic spatio-temporal "signatures". (In ordinary circumstances, one billiard ball does not make a difference for whether another moves if it does not make spatial contact with it.)

## 4 The Role of DM and GM Information in the Characterization of Mechanisms

The rest of my discussion will explore the relevance of these ideas to controversies about how the notion of a mechanism is best characterized and how this notion figures in causal explanation. As I see it, these controversies turn in part on issues about the relationship between the two approaches to causation I have distinguished. Insofar as the emphasis on the role of mechanisms represents a distinctive alternative to DM (and in particular, interventionist) treatments of causation, I believe this is because *mechanistas* think of the GM approach (or at least some sort of non-comparative, "actualist" approach) to causation as more fundamental than DM approaches and as a promising way of capturing what is "really going on" at a "deeper" level when difference-making relations are present. In what follows I express skepticism about this idea but also argue, on a more positive note, that a more adequate characterization of the (or a) notion of mechanism can be found by integrating elements of GM and DM approaches.

## 5 Waskan's Proposals

It should be obvious that one can sometimes tell (for example, on the basis of randomized experiments) that (M) (or other similar conditions having to do with difference-making) are satisfied even if one knows nothing about the "mechanism" (in any intuitive sense of that notion) connecting $X$ to $Y$. In fact, it is quite common to be in this epistemic situation. It was known for many years that ingestion of aspirin causes (in the DM sense) headache relief but only recently has the mechanism of action of aspirin has come to be understood. Similarly, the role of SSRIs in causing alleviation of depression in at least some subjects is generally accepted, but the mechanism by which the drug produces this effect remains controversial. (M) seems to fit these observations in a natural way since for there to be causal relation between $X$ and $Y$, all that is required is the right sort of contingency between $X$ and $Y$ and one can have knowledge that such contingency obtains without knowing anything about the mechanism connecting $X$ and $Y$.

As Waskan notes, these observations seem to tell against one simple proposal about the semantics of causal claims: namely that to *assert* that $C$ causes $E$ is to assert that the specific mechanism that in fact connects $C$ to $E$ holds between $C$ and $E$. The

observations also tell against a related proposal about what it is to (internally) *represent* that there is a causal relationship between $C$ and $E$—namely that this requires that the subject represent whatever specific mechanism as a matter of fact connects $C$ to $E$. Waskan's response is to interpret the claim that $C$ causes $E$ as [making ]" a non-specific assertion to the effect that [$C$] bears *some general sort of mechanical relation"* to $E$ (p. 13). In other words, his contention is that the claim (1) $C$ causes $E$ "asserts" that there exists some mechanism connecting $C$ to $E$, although (1) does not specify what that mechanism is.

Waskan adds, "[t]he tricky part… is figuring out how to flesh this idea out into something useful". This is indeed tricky, for at least two reasons. The first issue, which I raise but will not further pursue, is to provide some non-trivial characterization of what "some general sort of mechanical relation" includes—a characterization that is sufficiently general to cover everything that intuitively one wants to classify as a "mechanical" relation and yet at the same time is not so vague as to be entirely toothless and to exclude nothing. For example, this characterization will need to be broader than Salmon/Dowe type characterizations in terms of energy/momentum transfer (in order to include, e.g., neural mechanisms) but also needs to exclude some kinds of DM relationships (perhaps those involving omissions and double prevention?) as not providing information about "mechanisms", if the GM approach is to be a genuine alternative to DM accounts.

Putting this issue aside, suppose we somehow arrive at a satisfactory characterization of what it is for $C$ to bear "some general sort of mechanical relation" to $E$. Is the existence of such a relation between $C$ and $E$ sufficient for $C$ to cause $E$? We have already seen the answer to this question is "no", at least if causation is understood in a broadly difference-making sense. This was the lesson of the cue ball and cough/yellow fingers examples: although there is a mechanical relationship in the paradigmatic sense of spatio-temporally continuous transmission of energy/momentum) between the presence of the chalk on the cue stick ($P$) and the falling of the eight ball ($F$), $P$ does not cause $F$. So "$P$ causes $F$" cannot mean merely that there exists some mechanism connecting $P$ to $F$. Note, moreover, that this difficulty cannot be avoided by looking for a less restrictive notion of "mechanical relationship" than "relationship mediated by energy/momentum transfer". The proposal under consideration says that as long as "there exists" a mechanical relationship between $P$ and $F$, then $P$ causes $F$ and there does exist such a relationship in this case, on any reasonable interpretation of "mechanical".

Even if we cannot appeal to the existence of "some mechanism" between $C$ and $E$ to elucidate what is for $C$ to cause $E$, there remains the interesting and important issue of just what kind of information is conveyed by mechanism claims. Is this just (more detailed) DM information or does it involve something completely different? To further explore this issue, let us briefly consider how one feature of information about mechanisms might be accommodated within an interventionist framework and then contrast this with the alternative that Waskan favors.

Whatever else is meant by "information about the mechanism" connecting $C$ and $E$, virtually everyone agrees that this often involves the provision of information about intermediate or intervening causal relationships that underlie or mediate the overall relationship between $C$ and $E$. As an illustration, consider the mechanism by which

aspirin causes pain relief, which in rough outline is as follows: headaches and various other injuries/disorders cause the production of prostaglandins, which cause swelling of the affected part and are involved in pain signaling. The synthesis of prostaglandins requires COX enzymes, of which there are two types, COX-1 and COX-2, operating on arachadonic acids. Aspirin works by inhibiting the synthesizing action of both enzymes, thus preventing the conversion of arachadonic acid to prostaglandins. In particular, both enzymes have "tunnels", through which the arachadonic acid must pass to reach the active sites of the enzymes. Aspirin blocks these tunnels, although the details of how it does so and the active sites themselves differ from COX-1 to COX-2. These differences are important because the form of prostaglandin made by COX-1 helps to protect the lining of the stomach; it is the inhibition of this form which accounts for the role of aspirin in upset stomach. Potentially, then, it may be possible to create agents which, unlike aspirin, act only on COX-2 and not on COX-1, thus providing some pain/inflammation relief without causing stomach upset.

On one natural way of understanding this mechanism information, it seems to fit well into a difference-making (and interventionist) framework—what is provided includes more detailed, fine grained difference-making information (and relatedly, more detailed, fine-grained possibilities for intervention) that goes well beyond information about the overall difference-making relation between aspirin and pain relief. (Perhaps this information about intervening causes involves more than just more fine-grained difference-making information, but it at least *includes* this.) For example, we are told the presence of prostaglandins makes a difference for whether one experiences pain and inflammation (hence that intervening on whether prostaglandins are present in some other way besides aspirin ingestion would also make a difference for pain and inflammation), that the synthetic activity of COX-1 and 2 makes a difference for whether prostaglandins are present (so that intervening on this can affect whether there is pain or inflammation, and that whether aspirin is present makes a difference for whether this synthetic activity occurs). Similarly, that the active site of COX-1 is different from the active site of COX-2 suggests the possibility of an intervention that inhibits only the former.

This is by no means the only way in which an interventionist account might attempt to capture some aspects of the notion of information about mechanisms (see below for some additional suggestions) but even at this point, it is at odds with Waskan's treatment. I have suggested that at least some features of mechanistic information are just "more of the same" when viewed from an interventionist perspective—more fine-grained and detailed information about interventionist counterfactuals than is provided by the original causal claim.

By contrast, Waskan sees information about connecting mechanisms as different in kind from the information about overall dependence provided by the original causal claim; it is not difference-making or interventionist information at all. If I have understood him correctly, Waskan's idea is that mechanism information differs from difference-making information in that the former is just information about what actually happens (so that in this respect, it incorporates one of the distinctive features of GM information), while the latter has a modal or counterfactual element to it. He writes:

> The driving intuition behind some mechanists' resistance to Woodward's account of the contents of causal claims is that causal claims make assertions about what *actually* happens rather than about what would happen…. In particular, some contend that causal claims make assertions about actual productive *mechanisms* connecting cause and effect.

To illustrate this idea Waskan refers to a mechanistic account of movement disorders in late stage Huntington's and Parkinson's put forward by Mink (1996). In Waskan's summary:

> To vastly oversimplify, [Mink] claims that motor-planning systems of the cortex initiate movement by focally exciting neurons in the striatum and globally exciting neurons of the *substantia nigra*. The latter in turn send inhibitory signals to each of the endogenously active neural triggers ….. in the cerebellum for familiar motor patterns (e.g., raising one's arm, clenching one's fist, etc.). At the same time, local activity in the striatum leads to selective inhibition of the *substantia nigra*, which has the effect of selectively *dis*inhibiting specific motor-pattern triggers. It is easy to envision how, on this setup, degeneration of striatum might produce movement-execution deficits (i.e., no brake would ever be disabled) and how degeneration of SN would lead to co-contraction rigidity (i.e., many brakes would be disabled simultaneously).

To illustrate what is distinctive about this information, Waskan asks us to consider the following causal claim, which I re-label as [2]:

> [2] Degeneration of dopaminergic striatal neurons (S) causes movement-execution difficulties (E).

He writes:

> There are at least two distinct ways in which we might conclude that [2] is the case. On the one hand, we might monitor the relative values of S and E and come to believe on this basis, and finally to claim, that S causes E. Let us call this a *superficial* justification…
> [On the other hand] one might, in principle, also reach the conclusion that [2] is true by learning about the relevant physical parts and processes of the nervous system …. For instance, one might discover that Mink's general proposals about the mechanisms connecting motor planning centers and motor pattern triggers are correct and subsequently determine that an implication of this setup is that increasing or decreasing S will result – that is, absent such impediments as the introduction of reuptake inhibitors, electrocortical stimulation, etc – in an increase or decrease in E.

In what sense is the information Mink provides just about "actual" mechanical connections or "physical parts and processes", in contrast to information about what would happen under interventions? One thing that might be meant by "actual connections" etc. are the anatomical or physical connections involving neurons, synapses etc linking different neural structures. These do figure prominently in Mink's discussion, since he is concerned to show that the various neural connections required

by his hypotheses about motor control are in fact present. But it is generally agreed one cannot just read off from facts about these anatomical connections, what the causal role (in the sense of difference-making) of structures like the *substantia nigra* or striatum are in motor behavior. One reflection of this is the contrast neurobiologists draw between "anatomical connectivity" and what they call "functional connectivity"—the latter being a difference-making notion having to do with whether various changes or signals originating in one neural structure cause changes in other structures. The existence of the right sort of anatomical connections is necessary for one neural area to influence another, but the existence of such connections does not in itself tell us what the pattern of functional connectivity between the two areas is or the details of how what happens in one area depends on what happens in the other.[3] Still less does it yield anything like a computational theory of how normal movement is produced or how damage to the circuits and structures involved in normal movement leads to movement disorders. For these latter purposes, covariational information of some kind seems required, which may derive either from experimental interventions or passive observation. In fact such evidence plays a central role in Mink's account, as is reflected in his use of information about the impact of surgically and chemically induced lesions of various sorts on movement disorders to elucidate the mechanism underlying [2]. For example, Mink notes the hypothesis that parkinsonism is associated with decreased activity of certain structures and increased activity of others [GPi and STN] can be tested with focal lesions of the latter structures in normal (animal) subjects. This in effect involves the use of experimental interventions to show how the normal activity of these structures makes a difference of various sorts for movement control. Mink also describes evidence involving recording from various structures when normal subjects perform tasks that involve different sorts of movements (for example, tasks involving sequences of movements versus single movements). This provides information that different structures are differentially involved in single versus sequential movement tasks and that damage to different structures leads to different sorts of movement disorders. In all of these cases, it looks as though what is represented or elucidated is the difference-making role of these different structures, and the different connections (excitatory or inhibitory) between them. In addition, Mink's review article (as well as his subsequent papers on movement disorders) contains several purported "circuit diagrams" (e.g., Figs. 14 and 15) showing (in a schematic way) how various sorts of activity, excitatory or inhibitory, in certain neural structures influence responses in others, and in turn influence movement. The diagrams also suggest how damage or disruption of certain of these circuit components leads to abnormal movement. Again it is natural to think of these diagrams as representations of difference-making relations, showing us that, e.g., increased activity in the striatum leads via a direct pathway to

[3] Would it help the actualist analysis if we included, within the class of "actual connections", such properties and processes as the discharge of axons, diffusion of ions across synaptic clefts, and the like? No. We still face the same fundamental difficulty—if we want to understand the causes (in the difference-making sense) of normal and abnormal movement, we need to identify the difference-making role of those properties/processes which make a difference for these outcomes. The mere occurrence of, say, a certain pattern of neural spiking followed by normal movement does not tell us what (if anything) in that pattern made a difference for whether the movement was normal. For that we need information about what would happen if the spiking pattern were different.

inhibition of GPi etc. What seems distinctive about these diagrams as representations of mechanisms is not that they eschew difference-making information, but rather the way in which they *integrate* it with other sorts information—for example, information about the physical connections and spatial relationships along which, so to speak, the difference-making relations operate.

I thus see this example as illustrating the same general point as the chalk transmission example. It is true enough that if the impact of the cue stick causes the eight ball to go into the pocket, there must be a "connection" of some appropriate kind between the stick and the ball. But the existence of this connection does not by itself pin down what it is about the motion of the cue stick that makes a difference for whether the ball goes into the pocket. Similarly for the role of various neural circuits in the control of movement.

## 6 More on Mechanisms and Interventionist Counterfactuals

So far, I have focused on one way in which an interventionist account might try to capture aspects of mechanistic information—this having to do with information about intermediate causal links. I turn now to some other ways in which interventionist accounts might capture other features of mechanisms.

I noted above the characterization (M) is vague and non-specific and that we often wish to know more details. Even when quantitative information about the $X-Y$ relation is not available, it may be possible to discover more qualitative features of the relationship that go beyond the simple claim that $X$ causes $Y$ and often this information will bear on the nature of the mechanism connecting $X$ to $Y$. For example, if $X$ is a variable that can increase or decrease in magnitude, then it is often of interest to learn whether such changes are associated with increases or decreases in $Y$ (whether there is a "dose–response" relationship), whether there are threshold effects in $X$'s relationship to $Y$ (as in the discharge of an axon) and so on. Again, this is naturally thought of as information about difference-making and it can be useful for the identification of mechanisms even when imprecise. For example, before the molecular mechanisms by which smoking causes lung cancer were elucidated, researchers had, in addition to the observation of overall correlations between smoking and lung cancer, general if vague ideas about the mechanism by which smoking produced this effect: they assumed that substances in tobacco smoke was inhaled during smoking, deposited in the lungs, and that their presence had a carcinogenic effect. This suggested additional hypotheses about the qualitative character of the smoking/lung cancer relationship, understood along interventionist or difference-making lines. For example, one might expect to see evidence of tobacco products in smoker's lungs. Another fallible but not unreasonable inference is that those who smoke more should deposit more carcinogenic material in their lungs and hence have a higher probability of lung cancer. Similarly for those who inhale deeply while smoking in comparison with those who do not. Filtering devices, if effective in removing the carcinogenic material in smoke, should reduce the inference of cancer. And so on. When these patterns were found, this increased confidence that smoking causes lung cancer. As this example illustrates, one reason why information about mechanisms is valuable (in addition to the increased

understanding it brings) is that it can play an important role in testing causal hypotheses. Relying just on the observed correlation between smoking and lung cancer, one might worry this is produced by some unknown confounding factor. However, it is extremely implausible that this factor should happen to be distributed in such a way that all of the patterns described above obtain—more cancer among those who smoke more heavily and so on. Finding such patterns helps to rule out the confounding factor hypothesis.

Yet another feature of information about mechanisms that can be naturally captured within an interventionist framework has to do with stability (discussed in more detail in Woodward 2006, 2010). The stability of a causal or counterfactual relationship has to do with whether that relationship would continue to hold as various other "background" factors change. As noted above, mechanistic explanations often provide information about intermediate causal links "underlying" an overall causal relationship and often these intermediate links will be more stable than the overall relationship itself. A concern with identifying stable relationships thus leads directly to an interest in identifying intermediate links in underlying mechanisms. In addition, as I have attempted to explain elsewhere (2006), relationships of counterfactual dependence that are not mediated by direct physical connections, such as relationships of "double prevention", may differ greatly in their stability. Other things being equal, the more stable such relationships are, the more likely they are to be judged as paradigmatically causal. By appealing to this consideration, an interventionist can provide a partial explanation of why many double prevention relations in biomedical contexts strike us as unproblematically causal,[4] while other double prevention relations, such as those in the Walsh and Sloman experiment described by Waskan, do not.

## 7 Mechanisms and Spatio-temporally Organized Difference-Making Information

Despite these suggestions about how interventionist ideas might be used to elucidate mechanism claims, many *mechanistas* will still think something crucial has been omitted. In this section, I offer a further proposal about what this might be—it is intended as a supplement to rather than a retraction of the interventionist approach.

When one looks at accounts of mechanisms in philosophy as well as examples of mechanistic explanations in science itself (especially the biomedical sciences) one thing that stands out is the role that the spatio-temporal structure or organization of the various components plays in the operation of the mechanism. For example, as emphasized in Bechtel (2006), biochemical mechanisms require that various reaction products be brought together at just the right times and at the right spatial positions with respect to each other for reactions to go forward. One expects there to be systematic relationships between difference-making and spatio-temporal relationships in other respects as well. If a reaction is claimed to proceed via some (causally) intermediate product, then one expects this causal structure will be mirrored in the spatio-temporal

[4] Lombrozo (2010) provides empirical evidence that subjects are more likely to judge stable double prevention relationships as causal than unstable relationships.

relationships that are present: we expect to find the causally intermediate product in a position that is spatially and temporally intermediate between the initial and final steps of the reaction. If not, this is at least prima-facie evidence that the hypothesized reaction mechanism is mistaken. Similarly, accounts of neural mechanisms require not just that there be neuronal connections mediating putative causal influences, but that these respect various temporal and spatial constraints that are characteristic of the brain. For example, the speed at which an animal carries out some perceptual discrimination task sharply constrains the possible neural mechanisms that might carry out that task, given information about the relatively slow velocity with which neural signals propagate. As another illustration, many causal influences in the difference-making sense (fundamental physical forces, transmission of infectious diseases, some social influences) diminish in regular ways with distance, so that nearby objects or regions are more strongly affected than more distant objects and regions.

These are specific illustrations of the more general point noticed in connection with launching experiments: there are often systematic connections between causation in the sense of difference-making and facts having to do with spatio-temporal relatedness and organization, with different difference-making relationships having, so to speak, different characteristic spatio-temporal signatures. One thing many mechanistic explanations do is to provide information about these DM-GM connections; in telling us about more intermediate or fine-grained causal relationships they also tell us how the objects and processes standing in these relationships are spatio-temporally related in such a way that these relationships allow them to exercise their difference-making roles. I emphasize, however, that in my view, unlike Waskan's, this is not a matter of *replacing* "superficial" facts about difference-making with "deeper" explanations specified in purely spatio-temporal or other "actualist" terms. Rather, mechanistic explanation involves, so to speak, difference-making information all the way down, but spatio-temporally organized difference-making information. In thinking about how mechanistic explanations work, we should think in terms of understanding how difference-making and spatio-temporal information work together in an integrated fashion, rather than thinking in terms of the replacement of the former by the latter.

The interrelationships between difference-making and spatio-temporal structure are diverse and defy easy summary. There is a tendency for philosophical treatments of these interrelations to focus on just one possibility—causal relationships in which cause and effect are spatio-temporally contiguous. It is important to recognize, however, that even when difference-making relationships do not involve spatio-temporal contiguity, they may involve some other characteristic spatio-temporal pattern. For example, many drugs and diseases are such that there is a characteristic temporal delay between exposure to the cause and the occurrence of the effect. If an illness occurs immediately after exposure to some putative cause (rather than, e.g., hours or days later), this is reason to doubt that the factor in question was really efficacious. Similarly for the role of many drugs in promoting recovery from disease.

A related observation is that there are many other possible relationships between difference-making and spatial organization, besides the obvious one that some difference-making relations require spatial contact. Even when literal spatial contact is not required, many causal agents require some degree of spatial proximity to produce their effects or, as noted above, tend to affect nearby objects more strongly than more dis-

tant objects or tend, if causally efficacious at all, to affect many or all nearby objects rather just a few. It is for this reason that patterns of spatial clustering and spread of diseases can often convey important information about the mechanisms by which they are caused, as illustrated, for example, by Snow's well-known investigations into cholera.[5]

## 8 Causal Perception and the Development of Causal Representation

I conclude with some brief remarks on the development of causal representation—a topic Waskan also addresses. One popular idea goes like this: infants' capacity for causal perception in connection with launching events (and perhaps other sorts of mechanical interactions) emerges very early—perhaps it is "innate". The representations "triggered" in such perception then become available for more general use to represent other sorts of causal relationships and for other kinds of causal reasoning and learning, but the concepts/reasoning patterns themselves undergo at best limited further change and development. As Schlottman and Shanks (1992) summarize the idea, "the mechanism of causal perception… would provide a robust intuitive understanding of the concept of cause". Although Waskan does not endorse the claim that the adult notion of causation derives *only* from causal perception, he is sympathetic to the claim that this plays some role:

> While our adult concept of causation probably does not arise solely from the content of the kind of causal perception associated with launch events, it may well arise from the integrated use of an array of perceptual, or perceptually derived, expectations regarding spatial, temporal, and other properties.

I agree with Waskan that the challenge is to understand how the different elements making up the adult notion of causation are (eventually) integrated or put together. But unlike what may be Waskan's view, I think one of the central elements that needs to be integrated has to do with the role of difference-making considerations. As we have seen, conceptually (and in some respects empirically) these DM considerations are independent of notions having to do with spatio-temporal connectedness and other notions definable in actualist terms (such as the features that Waskan describes as "perceptual"). As a result, it seems unlikely that the representation of causation qua difference-making can somehow be derived from, reduced to, or replaced by a representation characterized just in actualist terms.

The acquisition of a full adult notion of causation is a very complicated matter, but here is a speculative suggestion about *part* of the story, which draws on my remarks about difference-making relations often having characteristic spatio-temporal signatures. Imagine the young infant begins, developmentally, by finding certain spatio-temporal relationships, such as those involved in launching and other contact events

[5] The notion of modularity in Woodward (2003) provides another illustration of interconnections between DM aspects of causation and spatio-temporal relationships. Often different modules correspond to spatio-temporally distinct parts.

particularly salient. At some early point—I believe it is not fully known when[6]—they also notice these relationships are connected to motion (or its absence) in characteristic difference-making ways—balls move when and only when there is spatial contact with other moving objects, objects remain at rest when supported by spatial contact with solid surfaces and fall when unsupported. Infants are sensitive to covariational information in general even when this does not have distinctive spatio-temporal features, but covariational relations are often more easily learned and incorporated into action when they do possess such features (cf. Bonawitz et al. 2010). The infant's experiences with moving balls etc. leads to "expectations" regarding the relationship between covariational and spatio-temporal relationships (for which we can test in looking-time experiments) but these need not have all the features of full-fledged adult causal representations—among other things, the infant may lack full appreciation of their significance for manipulation. In addition, expectations about covariation do not in themselves require the idea that some patterns of covariation are causal and others are not. Moreover, initially infants register only very coarse-grained spatio-temporal features of these experiences—e.g., they register just that support requires there be *some* spatial contact between the object supported and the support. As a result (as experiments show), the infant is initially not surprised by an ecologically unnatural case in which an object fails to fall even though its center of mass is located well off the support. But with additional experience, the infant begins to learn that the *details* of spatio-temporal relationships are also systematically connected to covariational or difference-making relationships: the infant learns whether an object in contact with a surface will be supported depends on how much of the object overlaps the surface, that whether (and how much) an object hit by another will move depends not just on whether there is spatial contact but on the relative sizes, heaviness, and speeds of the objects, and so on. At the same time, the infant also acquires experience of difference-making relationships—for example, those involved in social interactions—that show that difference-making can exist independently of spatio-temporal contact. The result, eventually, is an integrated understanding (and accompanying representation) of how the difference-making aspects of causation are connected to distinctive spatio-temporal relationships in some cases, but not always. In this process, the representations deployed in perception of launching and similar interactions are not retained unaltered from early infancy but are rather enriched and transformed in important ways.

## References

Bechtel, W. (2006). *Discovering cell mechanisms*. Cambridge: Cambridge University Press.

Bonawitz, L., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A., Woodward, J., & Schulz, L. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition, 115*, 104–117.

Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.

[6] For example, it is unclear whether the 6 month-old infants who allegedly perceive causation in launching events would show evidence of violated expectations if they were given information that is discordant between perceptual and covariational cues to causation.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul, *Causation and counterfactuals*. Cambridge, MA: MIT Press.

Hitchcock, C. (1995). Discussion: Salmon on causal relevance. *Philosophy of Science, 62*, 304–320.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology 61*, 303–332.

Mink, J. (1996). The Basal Ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50*, 381–425.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Schlottman, A., & Shanks, D. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology, 44A*, 321–342.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, 366–377.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review, 115*, 1–50.

Woodward, J. (2010). Causation in biology: Stability, specificity and the choice of levels of explanation. *Biology and Philosophy, 25*, 287–318.

Woodward, J. (forthcoming). Causal perception and causal cognition. In J. Roessler (Ed.), *Perception, causation and objectivity: Issues in philosophy and psychology*. New York: Oxford University Press.