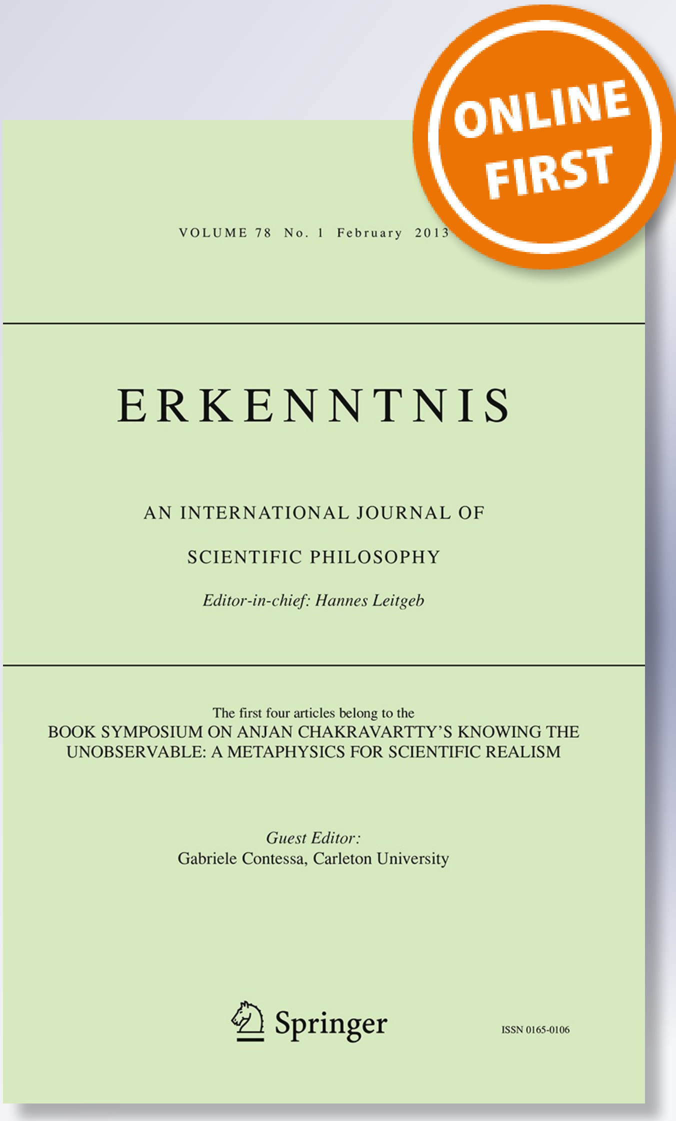
Functions and Cognitive Bases for the Concept of Actual Causation

David Danks

Erkenntnis
An International Journal of Scientific Philosophy

ISSN 0165-0106

Erkenn
DOI 10.1007/s10670-013-9439-2



Erkenn
DOI 10.1007/s10670-013-9439-2

Functions and Cognitive Bases for the Concept of Actual Causation

David Danks

Received: 23 January 2013 / Accepted: 23 January 2013

Abstract Our concept of actual causation plays a deep, ever-present role in our experiences. I first argue that traditional philosophical methods for understanding this concept are unlikely to be successful. I contend that we should instead use functional analyses and an understanding of the cognitive bases of causal cognition to gain insight into the concept of actual causation. I additionally provide initial, programmatic steps towards carrying out such analyses. The characterization of the concept of actual causation that results is quite different from many standard views: it is graded, context-sensitive, and extrinsic.

1 Finding the Concept of Actual Causation

One can usefully distinguish, in a quasi-Humean way, between two different analysis targets about the nature of actual causation—that is, the nature of the causal relations (if any) that obtain between particulars. First, one might have in mind some objective relation in the world, such as the transfer of conserved quantities (Dowe 2000) or reliable changes in probability distributions after interventions (Woodward 2003). Roughly, these analyses aim to explain actual causation as something that is independent of human experience, capacities, constraints, and so forth. In practice, these accounts are often tied closely to our best scientific understanding of the nature of the world. A second target for analysis is our concept of actual causation, by which I mean that thing which is a crucial (perhaps even necessary) and ubiquitous part of our everyday reasoning, practice, discourse, and

D. Danks (✉)
Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA
e-mail: ddanks@cmu.edu

D. Danks
Institute for Human and Machine Cognition, Pensacola, USA

phenomenology. My understanding of this analysis target can be understood as almost pseudo-Kantian: actual causation pervades our experiences and reasoning about the world, and is a foundational part of our cognition. In particular, neither the content nor the use of this concept need be introspectively accessible. It is not exhausted simply by the sorts of cases that people identify as "causal," or the sentences to which people will naively and rapidly assent.[1]

This paper focuses on our concept of actual causation, but not in the traditional manner of trying to find a formal or informal theory that accounts for the salient intuitions of philosophers about which cases "count" as causal. Instead, I will first argue that we must use different methods to develop a theory of our concept of actual causation. In particular, I suggest that so-called "functional analyses" (i.e., determining the role that the concept plays in our cognitive lives) and investigations into the cognitive bases of actual causation judgment and reasoning can provide us with better routes to understanding the nature of this concept. I then provide (in Sects. 2, 3) initial, admittedly programmatic attempts to apply these two methods. Each yields interesting claims about the concept of actual causation, such as actual causation being deeply context-sensitive and being closely connected with our actions. To be clear from the outset, this paper is not intended to provide a full theory of the concept of actual causation that predicts and explains all intuitive judgments about various philosophical test cases. Rather, the position underlying the (programmatic) proposal is precisely that our focus on predicting or explaining test cases has been misguided; to understand this critical concept of actual causation, we need to employ other methods.

Consider the question of how we are to go about determining the concept of actual causation. One common position is to think (as in the Canberra plan; e.g., Jackson 1998; Lewis 1986) that our concept of actual causation is whatever fills that role in our best theories. As I noted above, we ought not focus on our folk theories, as the role of actual causation need not be introspectively or consciously accessible. The alternative would be to ask what role causation plays in our proverbial "best sciences," or at least those sciences that regularly use causal notions. The problem is that our best sciences focus on the world, rather than our experiences. It is of course implausible that our concept of actual causation is completely unrelated to actual causation in the world (Danks 2005); the collision of the cue ball presumably *causes* the eight ball to go into the corner pocket, regardless of whether we mean that certain objective relations hold, or that we (necessarily?) experience it that way. At the same time, such correlation is insufficient. Consider by analogy the case of 'solidity.' One of the lessons of quantum mechanics is that (the content of) our deep, ubiquitous concept of solidity is not identical with what our best sciences tell us make something solid (or not), even though applications of the former closely track instantiations of the latter. Similarly, we should not assume that the content of our concept of actual causation simply corresponds to whatever relation(s) are singled

[1] Thus, although I will use cognitive data, I do not equate the concept of actual causation in our cognition and experience with whatever determines the attributions of the "folk" (in contrast with the suggestion of Hitchcock 2007).

out by our best sciences; that should be a conclusion of our investigations (if true), not a premise.

The other major philosophical tradition for understanding the concept of actual causation is through the construction of theories that predict and explain our considered, reflective intuitions about various cases.[2] In this strategy, one begins with some cases that prompt explicit, conscious, and verbal actual causation judgments (positive and negative), develops a theory based on salient features of those cases, and then uses a battery of standard "test cases" to judge the quality of a proposal. Many accounts of actual causation have this structure, though they differ on the substance of the proposal, the exact test cases used for validation, and so forth. In almost all of them, a model of "C causes E in context U" (whether formal or informal) is proposed, where comparison with salient test cases provides both guidance in the development of the theory and grounds for testing the empirical correctness of the account. In many ways, this strategy is just an application of the scientific method, but using canonical intuitions as the empirical data.

There is clearly something quite sensible about this strategy: if the concept of actual causation is a crucial part of our cognition, then any account of it should predict the intuitive judgments on which we all agree. But at the same time, this strategy will ultimately succeed only if there are not too many cases to consider: if there are thousands of distinct, relevant cases, then any survey-based strategy will inevitably fail to capture all of the interesting boundary cases. This problem is one reason why, for example, psychologists typically aim to discover underlying cognitive mechanisms, rather than just capturing observed behavior. Human behavior is simply too complex to try to capture it fully at the level of observable behavior; rather, we need to try to understand the underlying processes that generate that behavior. Moreover, it appears that actual causation presents us a situation in which there are simply too many cases for a survey-based procedure to ultimately be sufficient. Glymour et al. (2010) recently showed that there are an enormous number of distinct situations about which we might need to collect intuitions, even if one takes a highly restricted view of what counts as a 'case' and uses a variety of symmetry and simplicity principles to group together different structures as informationally identical. The underlying challenge is essentially a combinatoric one: there are too many instantiations of the underlying variables, too many ways for the elements of the case to be connected, and too many functional forms for those connections.

One way to block this combinatoric explosion would be to argue that all of the interesting features of actual causation could be explored using simple cases. If three-variable cases suffice to tell us everything we need about the concept of actual causation, then the combinatoric explosion is simply irrelevant. It is hard to conceive, however, of how one could make an argument in the abstract that three-variable cases are sufficient; rather, it seems as though the argument would need to be made on behalf of some particular theory of actual causation. And at least for many of the extant theories, there is no such argument: interesting and novel issues

[2] Many people call this process 'conceptual analysis.' I avoid that term simply in order to sidestep various debates about the proper methodology and exact power of conceptual analysis.

arise as the number of variables grows (Glymour et al. 2010). Alternately, one might wonder whether experimental philosophy might provide a way around this challenge, particularly given recent work suggesting that non-philosophers and professional philosophers have similar intuitions (Hitchcock and Knobe 2009). Although the use of a larger subject population could help with the combinatoric problem, it cannot be a solution as the numbers are simply much too large. We cannot simply mine our intuitions, whether by reflection in the armchair or by empirical surveys, to understand the nature of our concept of actual causation.

I propose that we instead investigate the concept of actual causation using two different, complementary strategies that can (arguably) avoid the combinatoric problem.[3] Both strategies depend on finding ways to characterize whole sets of behaviors, judgments, and actions that involve actual causation with single principles, though they do so from different directions. If one can find such principles or features, then there is no need to survey endless numbers of cases; rather, one gets the cases "for free" since they can be derived from the underlying principles. The first such strategy is to engage in a functional analysis of the concept of 'actual causation' in our experiences. That is, one can ask: What cognitive (or other) functions are fulfilled by this concept, and which functions require such a concept? A functional analysis is not an optimality or evolutionary account of some concept: it asks only what the concept of actual causation provides us in our current experiential lives, rather than asking either (a) whether it provides the optimal satisfaction of those functions, or (b) what functions it fulfilled in our evolutionary ancestors. One could think of this as a "top-down" strategy: first figure out what actual causation is good for, and thereby understand (something about) its content.

A second approach is to determine the cognitive representations and processes that underlie actual causation. That is, one can try to understand how, when, and why our cognition leads to experiences of actual causation. This latter route is explicitly *not* a call for more surveys of intuitions about particular cases, as that would clearly not solve the combinatoric problem. Put more pointedly, this route is not "experimental philosophy" as it is standardly performed today, even though it will involve both experiments and philosophy. Instead, one aims to use knowledge of the underlying mechanisms and representations to determine the scope and application of the concept without fighting the combinatorial battle. As an analogy, understanding addition as a set of ordered triples to which people will assent (e.g., <2, 2, 4>, <4, 4, 8>, <17, 5, 22>, etc.) is a remarkably inefficient way to determine the concept of 'addition' (if it can succeed at all, *contra* the arguments of Kripke 1984). But if one understands the underlying process that generates these acceptable triples, then one *can* fully characterize it. Of course, the development of a cognitive theory depends on empirical data as well, so one might worry that the "too many

[3] These two approaches are not intended to be exhaustive. Another strategy would be to analyze the meanings of causal terms in everyday language (i.e., not simply whether people would assent to apply them in particular cases). This strategy underlies the "force dynamics"-based model of causal language advocated by Talmy (1988), Wolff (2007) and Wolff and Song (2003). That account appears to share substantial overlap with the account advocated here, particularly my focus on causal perception. A unification or reconciliation of causal perception and force dynamics-based thought is an open issue at the current time (though see White 2006, 2009).

cases" problem reemerges. The difference is that a cognitive theory can be developed and tested against multiple sources of data about the underlying cognitive processes (i.e., not just verbal responses to vignettes), including response times, behavioral responses, and impacts on other aspects of cognition (e.g., object perception). By expanding the space of (possible) empirical constraints, one can arguably improve the odds of finding a theory that captures and explains our concept of actual causation.

These two strategies are plausible alternative means to understand the nature of (the concept of) actual causation, but the devil will ultimately be in the details. Many intriguing ideas have nonetheless failed to provide genuine insights. As a "proof of concept," I spend the next two sections on a programmatic "first pass" along each of these paths: a functional analysis of the concept of actual causation, followed by an examination of the cognitive processes underlying causal perception, an aspect of causal cognition that is closely connected with actual causation. Along the way, I will also try to demonstrate the complementary nature of these two strategies: by approaching the problem from different directions, they can provide both different and supporting insights into the concept of actual causation. In many ways, though, the key moral of this paper is not any particular detail of some analysis, but rather that we (as a philosophical community) need to consider alternative ways of coming to understand the concept of actual causation. Instead of mining our own intuitions or performing yet another survey, we should consider the functions and underlying cognitive processes of actual causation.

2 A Functional Analysis of Actual Causation

A functional analysis aims to understand something about the nature of its target by examining the functions that the target is supposed to fulfill. If one understands the functions or uses of *X*, then at the very least, one can infer that *X* must have whatever features, properties, or structure are required for those functions to be satisfied. More colloquially, a functional analysis involves moving from use to nature, though one must be careful not to conflate these two. Functional analyses are particularly useful for targets that have relatively inscrutable "insides," as a careful examination of the target's function can be a useful source of indirect information about its content. Functional analyses are widespread in evolutionary biology (e.g., using a trait's function to infer something about the mechanisms that produce or maintain it) and cognitive science (e.g., finding constraints on learning processes based on the uses of the representations that they generate). At the same time, there are limits to what we can learn from a functional analysis. In particular, if a function is subserved by multiple elements, then we cannot necessarily narrow down exactly which element has which feature; rather, we can only conclude that the complex has those features. Behaviors, including intuitive judgments, typically depend on multiple concepts and multiple cognitive processes, and so one can account for (successful) behavior in multiple ways. In such cases, we will have to look to other sources of information to apportion features further. Various functional analyses (or close relatives) of the concept of actual causation have been proposed in recent years; the basic idea is not novel to this paper. I thus return at the end of this section to discuss how the present analysis differs from them.

The starting point of any functional analysis is an articulation of the function(s) of the target. Even a cursory reading of the cognitive science literature makes it clear that, at least behaviorally (if not always in conscious introspection), causal judgments support explanation, prediction, and control. If we know the causal structure of a system, then we can explain its behavior (e.g., Kim and Keil 2003; Pennington and Hastie 1992), predict what it will do in the future (e.g., Cheng 1997; Shanks 1995; Waldmann 2000), and intervene to control it if we have that ability (e.g., Gopnik et al. 2004; Gopnik, Sobel, Schulz, and Glymour 2001; Sloman and Lagnado 2005). At first glance, this triple—explanation, prediction, and control—seems an odd combination: the first element is backward-looking, while the other two are forward-looking. In terms of the cognitive functions of 'causation,' however, this appearance is misleading: there is substantial empirical evidence that causal explanations are cognitively important not only for making sense of past events, but for doing so in order to predict and control in the future (e.g., Keil 2006; Lombrozo 2011; Lombrozo and Carey 2006). As Lombrozo and Carey (2006) summarize this work: "the function of explanation is to provide the kind of information likely to subserve future intervention and prediction" (p. 195).[4] Thus, any functional analysis should start with the premise that one—possibly, *the*—key function of the concept of actual causation is to support future predictions, actions, and understanding.[5]

This might seem to be a quite banal observation, but it leads to some notable conclusions. Predictive judgments invariably involve gradations: X is more likely than Y; Z has a 90 % chance of occurring; and so forth. A similar observation can be made about control judgments (e.g., action A is more efficacious than action B; action C has a 70 % chance of success) and arguably even explanations (e.g., explanation E is a better explanation of phenomenon P than explanation F). Moreover, in many cases, the graded nature of these judgments can seemingly arise only from gradations in the causal judgments. As a simple example, consider a predictive inference: (a) each ibuprofen tablet causes pain relief; (b) I took some ibuprofen; therefore (c) pain relief will probably follow. Claim (b) is surely not graded in any relevant way, and so claim (a)—the causal judgment about ibuprofen—must be graded in order to explain how the word 'probably' makes its way into the conclusion (c). We here confront a limitation of functional analyses, however, since claim (a) is a complex statement about our beliefs, the causal structure of the world, and so forth. Thus, a functional analysis alone is insufficient to determine whether the gradation in claim (a) arises from (i) the concept of actual causation being graded; (ii) one's degree of belief in a (non-graded) causal claim being graded; or (iii) gradations along both dimensions. We must look to other

[4] Of course, explanations can serve other functions as well, including having some intrinsic value.

[5] One might wonder whether *causal* knowledge is required for these purposes, since learned, non-causal associations (e.g., from classical or instrumental conditioning) have sometimes been suggested to be sufficient (e.g., Shanks 1995). There is, however, no known, purely associative (i.e., non-causal) learning process that can account for people's ability to use prior observations (not actions) to predict the outcomes of their later actions (as in Meder et al. 2010). The full range of human abilities of prediction, explanation, and control do seem to require truly causal knowledge.

sources of data to determine which of these possibilities holds. In the next section, I note some empirical data from research on causal perception that suggest that (ii) is *not* correct,[6] but the functional analysis alone tells us only that causal judgments are graded, not necessarily that the concept of actual causation is graded.

The vast majority of philosophical work on actual causation assumes that people's intuitive judgments about various cases (i.e., the "data" for our theories of actual causation) are binary judgments—"causes" versus "does not cause." The functional analysis suggests, however, that this is only a special case of more general, graded distinctions that we draw about causation throughout our everyday experiences. The appropriate judgments must at least provide comparative information such as "*C* is a stronger cause of *E* than *D*," and potentially even richer information such as "*C* causes *E* to degree *n*."[7]

Consider now whether a concept of actual causation that is usable for prediction, explanation, and control captures an intrinsic or extrinsic relation (to use the language of Menzies 1996). Roughly, an intrinsic relation between *A* and *B* is one that depends only on features of *A* and *B* (and for causation, the spatiotemporal interval between them), as well as general factors such as the laws of nature and so forth.[8] In contrast, an extrinsic theory of causation holds that "*C* causes *E*" sometimes depends on other specific parts of the world. I contend that the demands of successful prediction, explanation, and control necessitate that our concept of actual causation be extrinsic in practice; that is, judgments that *C* actually causes *E* depend on features of the world (including absences) that are not part of *C*, *E*, or the spatiotemporal interval connecting them.[9] The fundamental problem is that Hume was right about the unobservability of actual causation in the world,[10] and so the application of our concept of actual causation must employ other contextual information to determine whether the appropriate (metaphysical) relations in the world obtain, whatever those might be. As a result, a judgment that *C* actually causes *E* cannot be based purely on factors involving those events.

Consider the simple example of concluding that "my flipping this light switch caused the light to illuminate." At first glance, this judgment appears to be about an intrinsic relation: it does not depend, for example, on my shirt color. Suppose, however, that my daughter flips a switch on the other side of the room at the same

[6] We can also look to the related (though of course, not identical) case of type-level causation. There are various experiments showing that confidence in a causal relation (i.e., the degree of belief) is not identical with causal strength judgments (e.g., Collins and Shanks 2006), and so (ii) cannot explain the empirical data. Thus, to the extent that one thinks that the concepts of type-level and actual causation are related, one should expect that actual causation is similarly graded.

[7] If there are sufficiently many causes in the world (or sufficiently many ways to vary particular causes), then we can potentially even move from "raw" comparative judgments to full cardinality information, analogously to moving from comparative bet acceptances to cardinal utilities.

[8] There is debate about whether the laws of nature or counterfactuals are intrinsic or extrinsic. For both Menzies and myself, that issue is irrelevant to the question of whether causation is intrinsic or extrinsic.

[9] Again, we see the limits of functional analyses: the application of the concept of actual causation must "look" extrinsic, but we cannot draw the more specific conclusion that the *content* of the concept must be of an extrinsic relation.

[10] One might object that we can seemingly directly perceive causation in certain cases (e.g., a collision causing a block to move). The next section focuses on this causal perception.

time. This piece of "extrinsic" information is potentially functionally relevant, since the illumination might be due to her action rather than mine. If I want to be able to predict and control in the future, it is crucial to know which switch was responsible. The functional analysis thus implies that my actual causation judgment in this case should depend on more than just features of the switch flip and light illumination; fine-grained timing information about other events in the environment should matter as well. There are clear similarities with preemption cases that are often used to argue that causation cannot be intrinsic (e.g., Schaffer 2000), but this is subtly different. The "non-intrinsic" information here serves to determine and illuminate the broader causal structure within which the putative cause and the effect are located, including alternative explanations of the occurrence of the actual effect (e.g., my daughter's action). The unobservability of actual causation means that our concept of it must "act like" a concept of an extrinsic relation: its application and use must depend on more than just features of C, E, and the spatiotemporal interval connecting them.

More generally, the functions of prediction, explanation, and control arguably require information about the contexts in which some particular causal relation can be expected to obtain in the future, its likelihood in various conditions, and so forth. The very possibility of successful prediction depends on future conditions being sufficiently similar to past ones that the past can be used as a guide. Similarly, control depends upon future underlying causal structures being relevantly similar to the current structure(s). This similarity has at least two important dimensions: similarity in the putative cause, and in the broader context. As Woodward (2006) extensively argues, causes can be more or less sensitive to variations either in their realizations or in the conditions for their success. For example, whether a switch turns on a light typically is not sensitive to which hand is used to flip the switch, but is sensitive to flipping the switch versus hitting it with a hammer.[11] Successful prediction, control, and explanation all depend on the use of exactly this type of information. If, for example, I want to control the lights in my office, the decision about what action to perform will depend on the sensitivity of the outcome to the various actions at my disposal. Thus, this functional analysis implies that my judgment that "my flipping the switch caused the light to illuminate" is *not* a judgment simply about these two particular events, but rather contains—presumably, implicitly—significant information about the repeatability of this connection in various contexts, under various ways of flipping the switch, and so forth. Actual causation (in our cognitive lives) does *not* seem to be token causation, contrary to the common conflation of the two (unless the relevant token is much broader and different than we typically think). Instead, if we consider the uses of actual causation in prediction, explanation, and control, then we can see that the resulting judgments are graded, about an extrinsic relation, and include information about the sensitivity of both the relation and relata to variations in the environment and the realizations.

[11] This sort of sensitivity is part of a broader issue of how "causal" variables arise in cognition, including the scope and level of the variables. See Woodward (2006) for an extended discussion.

This functional analysis is importantly different from at least two prominent attempts at such analyses.[12] Menzies (1996) explicitly proposed a functional definition of causation, analogous with functional definitions of mental states such as pain, belief, and so forth. Specifically, Menzies (1996) argued that 'causation' is whatever fulfills the role of that term in the folk theory of causation, which is defined as (the conjunction of) "the platitudes about causation which are common knowledge among us." (p. 97) He then asserts that the three key platitudes are: (i) causation is a relation between distinct events; (ii) causation is an intrinsic relation; and (iii) except in cases of preemption and overdetermination, causes increase the chances of their effects. The present approach agrees with Menzies's about the importance of use and function, but does not claim that this is the proper way to *define* the concept of causation. One can use a functional analysis to gain insight into a concept without thereby being committed to the claim that there is nothing more to the concept than these functions. The type of functional analysis that I have pursued gives information about the nature, but explicitly does not define that nature by the functions. Moreover, Menzies's approach arguably goes astray by focusing on claims to which all (or most) people would verbally assent. There are many examples of people endorsing statements verbally but then not reflecting those endorsements in their behavior, and so my functional analysis looks to behavior and action, rather than verbal assent. An example of this arises for one of Menzies's "fundamental" features of the concept of causation: he claimed that causation had to be an intrinsic relation, but the present functional analysis provides significant reasons to think that it is an extrinsic one (though Menzies may well be right that many people would say that causation is intrinsic, if asked about only that feature).

The present approach also bears certain similarities to the notion of epistemic causality that has been advanced by Williamson (2005, 2006a, b). At a high level, his epistemic causality can be understood as two separable claims: (a) all there is to causation is mental representation, and so it does not correspond directly to anything in the (non-mental) world; and (b) "*C* causes *E*" is true iff an omniscient, rational agent would (given available evidence) believe it.[13] The present paper shares Williamson's focus on mental states, as I am concerned with the concept of actual causation rather than actual causation "in the world." At the same time, that focus is exactly why this paper is wholly agnostic about claim (a); no arguments about the concept of actual causation can tell us either way whether Williamson is right that "causality is a feature of the way we represent the world, rather than a direct feature of the world itself." (Williamson 2006b, p. 259).

[12] Hitchcock and Knobe (2009) is also framed in terms of the concept of actual causation and the function of resulting judgments, but their paper seems to have the structure: "the concept of causation does not have the content to be used for successful prediction and planning, so it must have some other function." In contrast, I am arguing that it clearly is used for prediction and planning, so its content must be different from what we (*qua* philosophers) have presupposed.

[13] These claims are related, but have no logical dependence. One could believe (a) without (b) by endorsing a kind of relativism. One could endorse (b) without (a) by arguing that it is a contingent feature of causal epistemology (i.e., contingently, an omniscient, rational agent would learn all and only true causal facts), rather than a definitional claim about the nature of causation.

Claim (b) is arguably more relevant for present purposes, as Williamson directly connects what the rational agent should believe with the uses to which those beliefs will be put (e.g., in his 2006a). Specifically, he contends that causal beliefs play both an explanatory and also a predictive/inferential role, and that consideration of these helps us to understand the nature of those beliefs. The former role implies that those beliefs must be consistent with our best understandings of the physical laws and mechanisms that are relevant in a particular situation. The latter implies that those beliefs must carry information about difference-making.[14] The present functional analysis diverges from Williamson's analysis in two key ways. First, it is focused on the functions of the concept of actual causation for creatures like us, rather than omniscient, rational agents.[15] Second, it focuses on aspects of the concept of causation that are (arguably) not easily captured in formal frameworks (e.g., causal graphical models).[16] The present functional analysis and Williamson's epistemic causality thus share many particular claims, but assemble and use them in quite different ways.

3 Actual Causation and Causal Perception

If the preceding fragment of a functional analysis is even approximately correct, then we should be able to find evidence of these features in our cognition about actual causation. Just as the functional analysis focused on our actions and behaviors (e.g., in prediction or explanation), a cognitive analysis should center on the cognitive bases of (uses of) our concept of actual causation, rather than starting with the statements to which people would consciously and verbally assent. In this section, I explore the cognitive bases of causal perception, which provides arguably the most natural and obvious cognitive processing involving actual cause judgments. Recall that the key morals from the functional analysis were that our actual causation judgments appear to be (i) graded and (ii) arising as though causation were an extrinsic relation, precisely because (iii) they are used for prediction, explanation, and control in our everyday lives. Thus, to the extent that causal perception involves such judgments, we should expect it to be (i*) a graded phenomenon that (ii*) is sensitive to relevant features of the broader environment or context, precisely because (iii*) it is closely tied to action and behavior. Moreover, we may be able to better allocate some of these features "closer" to the concept of actual causation by considering causal perception.

[14] And these two together imply that rational causal beliefs should, when possible, be structured as a directed acyclic graph whose parametric component obeys the Markov assumption and is consistent with mechanism information (Williamson 2005).

[15] Williamson's focus seems to be driven by his need for an objective basis for causal claims (which most philosophers believe cannot come from descriptive considerations of humans), which itself arises partly because he denies that there is any causation "out there.".

[16] For example, causal graphical models have great difficulty characterizing causal relations with "thick" spatiotemporal characteristics (e.g., the ability to intervene on the underlying mechanism at arbitrary spacetime points). The present functional analysis has no such restriction. Similarly, there are presently no formal models of causal perception, which is the focus of the next section.

Causal cognition can be roughly divided into causal reasoning and causal learning/judgment, where the latter can be further split into causal inference and causal perception (Danks 2009). 'Causal perception' refers to the relatively direct perception of causality, principally in response to visual stimuli, though other modalities can be relevant (see Rips 2011 or Scholl and Tremoulet 2000 for recent reviews). For example, suppose I observe the cue ball move across the pool table, make contact with the eight ball, followed by the eight ball beginning to roll while the cue ball stops. In this case, I have an immediate and relatively automatic perception that the cue ball hit *caused* the eight ball to roll.[17] Causal perception judgments are arguably much closer to paradigmatic actual causation judgments than other types of causal judgments or cognition.[18]

The classic experimental works on causal perception were Michotte's investigations of the launching effect (summarized in Michotte 1963) and Heider and Simmel's (1944) exploration of perception of social causes. The canonical launching effect is essentially the cue ball/eight ball case, but done with abstract geometrical objects projected onto a screen. One square moves in from the left and makes contact with a square in the middle, at which point the left-hand square stops and the right-hand square begins to move. When the second square's movement is approximately[19] simultaneous with the contact, then people experience the episode as the first square *causing* the second square to move. If there is a significant spatial or temporal gap or overlap in the sequence (e.g., the second square begins moving long after the contact), then the phenomenology of causation disappears. In Heider and Simmel's experiments on social causal perception, people watched movies in which geometrical shapes moved in particular patterns, and they reported an immediate perception of the trajectories as *caused* by the shapes' "personality" traits. For example, given a movie of a triangle moving erratically across the screen with a square smoothly moving behind it, people frequently report that the triangle is "scared" of the square, and the square is "chasing" the triangle.

The basic launching effect phenomenon—perceiving causality in straightforward cases—seems to emerge as early as 6 months of age (Leslie 1982), with more sophisticated launching effect phenomena emerging over subsequent months (e.g., Oakes and Cohen 1990). Social causal perception emerges by nine months of age (Csibra et al. 1999). Causal perception thus arises relatively early in development, and is largely automatic, phenomenologically instantaneous, and not particularly susceptible to top-down control (Blakemore et al. 2001). These features have prompted a significant debate about whether causal perception is a classically

[17] There is a large philosophical literature about whether perception is conceptualized, but those debates are almost exclusively about whether all perception is *necessarily* conceptualized. All I require is the relatively uncontroversial claim that this particular type of perception (in normally functioning humans of at least roughly 2 years of age) is conceptualized.

[18] In fact, some authors seem to have in mind the view that causal perception is the only source of token-level or singular causal judgments. (Something like this idea is expressed in Armstrong 2004, but the general sentiment occurs elsewhere.) I reject this idea for multiple reasons, but it does provide further reason to focus on causal perception.

[19] Interestingly, it turns out that the phenomenological experience of launching and causation is maximal when there is a slight delay before movement onset (Schlottmann and Anderson 1993).

modular process (e.g., Scholl and Tremoulet 2000) or not (e.g., Schlottmann 2000); for our purposes, it suffices to see that causal perception is a vital and ubiquitous part of human cognition.

Given this background on the nature of causal perception, we can turn to the question of whether it exhibits the three properties listed at the start of this section: graded, extrinsic (more properly, referring to an extrinsic relation), and directly tied to actions, both current and future. The first is easy: multiple experimental paradigms have demonstrated that causal perception judgments smoothly vary from "clearly causal" to "clearly non-causal" (e.g., Schlottmann and Anderson 1993; Scholl and Nakayama 2002). Some stimuli are perceived as only "somewhat" causal. Moreover, there is no evidence that these judgments reflect only uncertainty about whether a causal relation obtains, which casts further doubt on the possibility (raised earlier) that actual cause judgments are graded solely because one has degrees of belief that a non-graded causal relation obtains.

The second property is somewhat trickier, as we must ask whether causal perception is sensitive to other aspects of the environment (i.e., not just C, E, and the spatiotemporal interval connecting them) that signal that the causal structure in the world is such that C really did cause E. An example of such a context effect is the phenomenon of *causal capture* (Scholl and Nakayama 2002), in which perception of an ambiguous stimulus shifts from non-causal to causal depending on whether a causal event occurs elsewhere in the environment at approximately the same time. In particular, consider a red square moving in from the left side towards a blue square in the center of the screen, and suppose that the red square comes to completely cover the blue square, at which point the red square stops and the blue square begins to move along the same trajectory as the red square was following. In isolation, this sequence of events is typically perceived as the same square moving smoothly all the way across the screen, but with the squares (non-causally) changing colors at the moment of overlap. If a launching effect event occurs elsewhere on the screen at the same time as the overlap, however, then the exact same sequence suddenly shifts to being perceived causally as a standard launch (albeit, one that occurs when there is overlap rather than initial contact). That is, causal capture is a case in which the perception of "C causes E" (i.e., the first block causes the second block's movement) depends on more than just C, E, and the spatiotemporal interval connecting them. The broader environment matters.

Causal capture is just one of many context effects in causal perception. The spatial arrangement of percepts on the screen—in particular, whether the perceiver groups the percepts into connected (explicitly or implicitly) sets—makes a significant difference for whether an ambiguous stimulus is perceived causally (Choi and Scholl 2004). Causal perception can also be influenced by the timing of various events on the screen, including events that happen *after* the ambiguous stimulus (Choi and Scholl 2006). That is, the perception of some event as causal depends not only on contextual features at that moment, but also contextual features immediately after the event. Moreover, these context effects are not restricted to standard launching effect phenomena in adult humans; they are also found for social causal perception, such as perception of "chasing" (Gao et al. 2009) and arguably even in some non-human primates (Matsuno and Tomonaga 2005). Importantly, causal perception in these cases does not seem to be any less robust than "standard" causal perception. No phenomenological differences are reported, and these context-driven causal perceptions can significantly influence other aspects of the perceived scene. To take one notable example, Scholl and Nakayama (2004) examined situations that were very similar to causal capture cases, but instead of asking participants to judge causality, they asked them to judge the extent of overlap. When the sequence is perceived non-causally, participants should be able to recognize the overlap as complete, and they easily do so. In contrast, if the sequence is perceived causally, then the perceiver should "expect" (implicitly, of course) that the overlap is less than complete. Except in very special circumstances, a perception of a collision event involves only partial occlusion, not complete occlusion. As expected, Scholl and Nakayama (2004) found that participants who came to perceive the ambiguous sequence causally substantially underestimated the extent of overlap, even in the extreme case in which the overlap was actually 100 % at some point in time. There is thus no reason to think that the context-driven causal perception is any different from straightforward instances of causal perception.

These phenomena, as well as various *absences* of context effects (e.g., lack of causal capture when an object elsewhere on the screen changes color at the moment of overlap), raise the questions of what 'context' means here and *why* it should matter. At the current time, the best explanation seems to be that our cognitive systems—that is, the processes that generate our experiences—are biased to understand the visual scene in terms of internally coherent, but (relatively) mutually independent, causal structures (Scholl and Nakayama 2002, 2004). Thus, 'context' refers to those aspects of one's environment that are informative about the existence and operation of such structures. Consider the case of causal capture. In this situation, the visual system is, in some sense, presented with a "choice," though of course not a conscious one. One possibility is that the overlap event is independent of the collision event elsewhere on the screen and so need not be perceived causally, but then the simultaneity of the two events (the overlap and the collision) goes unexplained. The other possibility is that the overlap and collision events are part of the same causal structure and so the simultaneity is completely unsurprising, but that implies that the overlap must be perceived as a launching event. Given that accidental simultaneity is exceedingly rare in our visual experiences, the explanation for causal capture is that the visual system "opts" for the second option, and so perceives the overlap event as a causal event. Scholl and Nakayama (2002) are quite explicit about this way of thinking about why context effects arise:

> It is as if the visual system functioned in accord with the following reasoning: If the test event [the overlap event] is completely noncausal, then it is a coincidence that the moment of overlap occurred at precisely the same moment as the impact in the unambiguous context collision. Given that such coincidences are unlikely, the test event must have been a causal launch too. (p. 497)

Similar explanations can be provided for other occurrences and absences of context effects. For example, the effect of attentional grouping arises because the visual system "expects" (again, unconsciously) that groups will exhibit relative internal coherence. Thus, if one member of the group participates in a causal interaction at a moment in time, then we should (and do) perceive ambiguous stimuli about other group members as causal as well. At the same time, we should be (and are) less likely to perceive causally an ambiguous stimulus about an out-group member, since simultaneous causal events in distinct groups would be quite surprising (Choi and Scholl, 2006). Absences of context effects (e.g., simultaneous color change does not prompt causal capture) are similarly explainable. Color change and motion initiation rarely co-occur in causal structures,[20] and so there is no unexplained coincidence that "forces" the visual system to regard the ambiguous stimulus as causal.

This explanation presupposes that the visual system implicitly encodes the relevant statistics about, for example, the likelihood of simultaneous causal change in different causal structures. One might worry that that this is a fatal flaw in the explanation, as one might think that young infants, who do exhibit causal perception, are not capable of learning the relevant statistics. There are several observations to make in response. First, there is substantial empirical evidence that very young infants actually perform quite sophisticated statistical learning in multiple modalities (e.g., Saffran et al. 1996). Second, we do not know at what point infants exhibit these more sophisticated context effects in causal perception. Various, more complicated forms of causal perception emerge only gradually throughout development (Oakes 1994; Schlottmann 1999), and so context effects might be relatively late-emerging. Finally, there is neural evidence that sophisticated statistical learning in the visual domain can occur entirely unconsciously, thereby showing that conscious processing is not required to learn the necessary statistical regularities (Turk-Browne et al. 2009). These observations suggest that there are not yet substantive reasons to conclude that context effects could not depend in the necessary way on learned statistical (visual) regularities.

The third and final aspect that the functional analysis identified for our concept of actual causation is that it should be closely tied to future actions (see Wu 2011 for a related idea). It turns out that this feature is tricky to determine for causal perception. Perception of a causal relation between C and E clearly provides adults with information that can be used for future actions, but that does not tell us that the "future-directedness" is an integral part of the causal perception. The concern is that some causal perception might be entirely about a particular case, and then we use rich causal reasoning to draw conclusions from that observation. There is indirect evidence against this concern: White (2009) showed that people easily characterize launching effect events in terms of force and resistance—that is, exactly the information needed to manipulate the objects (see also White 2006). More generally, we can consider connections between the development of an infant's causal perception and her subsequent actions. To the extent that these are tightly coupled, we have (defeasible) reason to think that causal perception is closely connected to one's ability to perform sensible future actions, as very young

[20] In fact, the orthogonality of these two types of changes was used to help tease apart causal inference and causal perception (Schlottmann and Shanks 1992).

infants presumably do not yet have sophisticated causal reasoning abilities and are even still developing the relevant causal perception inputs for those abilities.

There is substantial ongoing research about the connection between the developments of causal/action perception and action production, but a growing consensus is that they are very tightly coupled (Rakison and Woodward 2008; Sommerville 2007). Three-month-old infants are typically unable to produce many actions; in particular, they have difficulty in performing directed reaching-and-grasping behaviors. Similarly, they do not (seem to) perceive others' actions as target-directed reaching, rather than simply arbitrary physical motions. Sommerville et al. (2005) examined the connection between these two *in*abilities by placing Velcro-covered gloves on a group of three-month-old infants. When in an environment with Velcro-covered toys, these infants could now perform reaching-and-grasping actions, as the toys would stick to the gloves. The infants who had these experiences subsequently perceived adult motions as target-directed, while a control group of infants with no such action experiences continued to perceive the adult motions as just arbitrary physical movements. We thus have evidence that self-actions can influence other-perceptions (see also Meltzoff and Brooks 2008; Rakison and Krogh 2012; Sommerville et al. 2008), which would be quite surprising unless those perceptions contained information relevant to the self-actions. Of course, as acknowledged earlier, these data do not establish that causal perception has a future-directed component. Nonetheless, they make it much less plausible that causal perception is not conveying, though its actual causation judgments, information about future actions.

4 Conclusion

Much of the actual causation literature is either ambiguous about its target, or focuses quite explicitly on actual causation in the world. I have aimed to focus instead on our concept of actual causation, though using different methods than one typically finds in a conceptual analysis. Empirical data are arguably crucial to understanding this concept,[21] but the standard methods of argument-by-intuition and argument-from-canonical-cases are simply insufficient. Enumeration of cases on which there is universal agreement (and many philosophical chestnuts do *not* fall into this category) does not provide enough constraints to fix the nature of actual causation in our experiences. Instead, we must turn to other techniques.

In this paper, I have provided an admittedly programmatic start to two such alternative methods: a (non-definitional) functional analysis of the concept of actual causation, and an examination of the cognitive bases of one source of actual causation judgment. These two different analyses provide convergent reasons to suspect that actual causation judgments must have contextual and forward-directed aspects. We care about actual causation relations because we want to know not simply what just happened, but also what will or could happen in the future. That is, the suggestion here is that an important part of the content of an actual causation

[21] A fully Kantian analysis could presumably explore the concept of actual causation without appeal to any particular empirical facts about us or the world.

judgment is information about the reliability or reproducibility of that relation, and not simply information about this particular case. Such information is inevitably graded in nature, context-sensitive, and dependent on other features of the environment (i.e., appears extrinsic).

Testing any claim about the content of a judgment is exceptionally challenging, but actual causation judgments are sensitive to context features and elements that are informative about the broader causal structure(s) in which *C* and *E* obtain. The concept of actual causation is not a concept of an intrinsic relation, or at least is not applied in any "intrinsic" manner. Elements of the environment besides simply *C*, *E*, and the spatiotemporal interval connecting them should be (according to the functional analysis) and are (based on causal perception research) relevant to our judgments of actual causation. Moreover, the evidence about the joint development of causal perception and action production in infancy is intriguing: perception of causal relations and the ability to use those relations seem to be closely connected in the development of human cognition. In general, the concept of actual causation seems to be a complex creature that is unlikely to be captured by a biconditional of the form "*C* causes *E* iff …" Thankfully, however, we have many different methods to help us come to understand the nature of actual causation in our experience.

Acknowledgments Initial versions of these ideas were presented at the 2010 Konstanz workshop on actual causation. Many thanks to the participants at that workshop—particularly Ned Hall, Chris Hitchcock, and Laurie Paul—for their helpful comments, questions, and criticisms. Three anonymous referees and two editors provided valuable feedback on an earlier version of this paper. Thanks also to Clark Glymour for discussions about the issues in Sect. 2. The author was partially supported by a James S. McDonnell Foundation Scholar Award.

References

Armstrong, D. M. (2004). Going through the open door again: Counterfactual versus singularist theories of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 445–457). Cambridge, MA: The MIT Press.

Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., et al. (2001). How the brain perceives causality: An event-related fMRI study. *Neuro Report, 12*, 3741–3746.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Choi, H., & Scholl, B. J. (2004). Effects of grouping and attention on the perception of causality. *Perception and Psychophysics, 66*, 926–942.

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception, 35*, 385–399.

Collins, D. J., & Shanks, D. R. (2006). Conformity to the power PC theory of causal induction depends on the type of probe question. *The Quarterly Journal of Experimental Psychology, 59*, 225–232.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition, 72*, 237–267.

Danks, D. (2005). The supposed competition between theories of human causal inference. *Philosophical Psychology, 18*, 259–272.

Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation* (pp. 447–470). Oxford: Oxford University Press.

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

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology, 59*, 154–179.

Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., et al. (2010). Actual causation: A stone soup essay. *Synthese, 175*, 169–192.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3–32.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629.

Heider, F., & Simmel, M.-A. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–249.

Hitchcock, C. (2007). Three concepts of causation. *Philosophy Compass, 2*, 508–516.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 106*, 587–612.

Jackson, F. (1998). *From metaphysics to ethics: A defense of conceptual analysis*. Oxford: Clarendon Press.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory and Cognition, 31*, 155–165.

Kripke, S. A. (1984). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

Leslie, A. M. (1982). The perception of causality in infants. *Perception, 11*, 173–186.

Lewis, D. (1986). *Philosophical papers* (Vol. II). Oxford: Oxford University Press.

Lombrozo, T. (2011). The instrumental value of explanations. *Philosophy Compass, 6*, 539–551.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.

Matsuno T., & Tomonaga, M. (2005). Effects of visual context on the perception of collision events in chimpanzees (*pan troglodytes*). *Perception, 34*, ECVP Abstract Supplement.

Meder, B., Gerstenberg, T., Hagmayer, Y., & Waldmann, M. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *The Open Psychology Journal, 3*, 119–135.

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology, 44*, 1257–1265.

Menzies, Peter. (1996). Probabilistic causation and the pre-emption problem. *Mind, 105*, 85–117.

Michotte, A. (1963). *The perception of causality*. London: Methuen.

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology, 30*, 869–879.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5*, 193–207.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189–206.

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science, 15*, 43–53.

Rakison, D. H., & Woodward, A. L. (2008). New perspectives on the effects of action on perceptual and cognitive development. *Developmental Psychology, 44*, 1209–1213.

Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science, 6*, 77–97.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Schaffer, J. (2000). Trumping preemption. *The Journal of Philosophy, 97*, 165–181.

Schlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology, 35*, 303–317.

Schlottmann, A. (2000). Is perception of causality modular? *Trends in Cognitive Sciences, 4*, 441–442.

Schlottmann, A., & Anderson, N. H. (1993). An information integration approach to phenomenal causality. *Memory and Cognition, 21*, 785–801.

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

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science, 13*, 493–498.

Scholl, B. J., & Nakayama, K. (2004). Illusory causal crescents: Misperceived spatial relations due to perceived causality. *Perception, 33*, 455–469.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299–309.

Shanks, D. R. (1995). Is human learning rational? *The Quarterly Journal of Experimental Psychology, 48A*, 257–279.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science, 29*, 5–39.

Sommerville, J. A. (2007). Detecting causal structure: The role of interventions in infants' understanding of psychological and physical causal relations. In A. Gopnik & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation* (pp. 48–57). New York: Oxford University Press.

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology, 44*, 1249–1256.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*, B1–B11.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934–1945.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*, 53–76.

White, P. A. (2006). The role of activity in visual impressions of causality. *Acta Psychologica, 123*, 166–185.

White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review, 116*, 580–601.

Williamson, J. (2005). *Bayesian nets and causality: philosophical and computational foundations*. Oxford: Oxford University Press.

Williamson, J. (2006a). Causal pluralism versus epistemic causality. *Philosophica, 77*, 69–96.

Williamson, J. (2006b). Dispositional versus epistemic causality. *Minds and Machines, 16*, 259–276.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276–332.

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

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

Wu, W. (2011). Attention as selection for action. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays* (pp. 97–116). Oxford: Oxford University Press.