Cognitive Psychology 61 (2010) 303–332

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions

Tania Lombrozo *

Department of Psychology, UC Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Accepted 17 May 2010

Keywords:
Explanation
Causation
Teleological explanation
Functional explanation
Intentions
Functions
Causal mechanism
Counterfactual dependence
Double prevention
Late preemption

ABSTRACT

Both philosophers and psychologists have argued for the existence of distinct kinds of explanations, including teleological explanations that cite functions or goals, and mechanistic explanations that cite causal mechanisms. Theories of causation, in contrast, have generally been unitary, with dominant theories focusing either on counterfactual dependence or on physical connections. This paper argues that both approaches to causation are psychologically real, with different modes of explanation promoting judgments more or less consistent with each approach. Two sets of experiments isolate the contributions of counterfactual dependence and physical connections in causal ascriptions involving events with people, artifacts, or biological traits, and manipulate whether the events are construed teleologically or mechanistically. The findings suggest that when events are construed teleologically, causal ascriptions are sensitive to counterfactual dependence and relatively insensitive to the presence of physical connections, but when events are construed mechanistically, causal ascriptions are sensitive to both counterfactual dependence and physical connections. The conclusion introduces an account of causation, an "exportable dependence theory," that provides a way to understand the contributions of physical connections and teleology in terms of the functions of causal ascriptions.

1. Introduction

Both philosophers and psychologists have argued for the existence of distinct kinds of explanations. For example, a knife's sharpness can be explained by appeal to the process that sharpened it ("it is

* Fax: +1 510 642 5293.
E-mail address: lombrozo@berkeley.edu

0010-0285/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2010.05.002

sharp because it was honed"), and also by appeal to the knife's function ("it is sharp because it's for cutting"). This basic distinction between backwards-looking explanations that cite causal mechanisms, which I refer to as "mechanistic," and forwards-looking explanations that cite functions or goals, which I refer to as "teleological," has proved fruitful in characterizing different ways of understanding properties and events (e.g. Keil, 2006; Lombrozo, 2006; Lombrozo & Carey, 2006). Moreover, recent findings attest to the cognitive significance of the distinction, with explanation type having a reliable impact on explanatory preferences (e.g. Kelemen, 1999a; Lombrozo, Kelemen, & Zaitchik, 2007) and on judgments about category membership (Lombrozo, 2009).

In contrast to this pluralist approach to explanation, accounts of causation have tended to be unitary, with both philosophers and psychologists striving to characterize "the" concept of causation. This paper explores the hypothesis that different kinds of explanation involve different criteria for causal ascription, introducing the possibility of a causal pluralism to mirror explanatory pluralism. The sections that follow motivate pluralism about explanation (1.1) and causation (1.2) before presenting hypotheses about the influence of different kinds of explanations on causal ascriptions (1.3), and providing an overview of the experiments that follow (1.4).

1.1. Explanatory pluralism

Aristotle famously identified four *aitia*, translated as "causes" or "modes of explanation," including the efficient cause, which brings something about, and the final cause, "that for the sake of which" something is the case (Falcon, 2008). For example, a piano can be explained by appeal to the manufacturing process that generated it, its efficient cause, or by appeal to its function in producing music, its final cause. This distinction has been picked up and refined by subsequent scholars, including the philosopher Daniel Dennett and the psychologist Frank Keil.

Daniel Dennett takes the notion of distinct explanations beyond answers to why-questions. According to Dennett, there are distinct "stances" that one can adopt in explaining and predicting the behavior of a system, including the physical stance and the design stance (Dennett, 1971, 1987). When adopting the physical stance, one reasons on the basis of underlying mechanisms. When adopting the design stance, one reasons on the basis of functions and goals. Naturally, these stances will be more or less appropriate, and more or less useful, for different systems and judgments. Typically, the design stance is applied to artifacts and biological organisms that exhibit actual or apparent design, and the physical stance to other physical phenomena.

Frank Keil has suggested a variant on Aristotle's and Dennett's ideas as a hypothesis about innate cognitive structure (Keil, 1992, 1994, 1995). Keil proposes that infants' cognitive toolbox includes a number of "modes of construal," among them a physical mode and a teleological mode (see also Gergely & Csibra, 2003; Kelemen, 1999a). Like Dennett's stances, these modes of construal influence how a system is conceptualized, and will be more or less useful for different systems. Keil suggests that these modes of construal become associated with different domains early in development, such that different modes are preferentially employed for different domains.

In this paper, I propose a distinction between mechanistic explanations that cite causal mechanisms (akin to Aristotle's efficient cause, Dennett's physical stance, and Keil's physical mode of construal), and teleological explanations that cite functions or goals (akin to Aristotle's final cause, Dennett's design stance, and Keil's teleological mode of construal). While these distinctions have a long history, there has been remarkably little research investigating the cognitive bases and consequences of different modes of explanation. In particular, do mechanistic and teleological explanations involve different representations or processes? If so, recognizing the impact of explanatory modes on reasoning will be central to a full understanding of cognition.

The psychological reality and significance of the distinction between mechanistic and teleological explanations is supported by recent work on explanatory preferences. In particular, children and adults seem to prefer teleological explanations to mechanistic alternatives, although teleological explanations are not endorsed uniformly across all contexts. The most striking evidence of a preference for teleological explanations comes from studies by Kelemen and colleagues with young children (Kelemen, 1999a). Given the option of explaining a rock's pointiness by appeal to a physical process ("bits of stuff piled up") or a function ("to prevent animals from sitting on them"), the majority of

first- and second-grade children prefer explanations like the latter (Kelemen, 1999b). This suggests that young children are already sensitive to the difference between teleological and non-teleological explanations, if only implicitly. Moreover, the preference for teleological explanations can persist in subtler forms into adulthood (Casler & Kelemen, 2008; Kelemen & Rosset, 2009; Lombrozo et al., 2007). Despite this general preference, both children and adults are more likely to accept and prefer teleological explanations for designed features of artifacts and for biological adaptations than for other kinds of properties and objects (Kelemen, 1999a; Lombrozo & Carey, 2006), again suggesting a sensitivity to the form of explanations. In tasks that involve soliciting explanations, young children may even restrict teleology to some domains (Greif, Kemler-Nelson, Keil, & Guiterrez, 2006), such as biological traits, mirroring adults on equivalent tasks.

A second source of evidence for the reality and significance of the distinction between mechanistic and teleological explanations comes from judgments of category membership. In Lombrozo (2009), participants learned about the relationship between three features of novel artifact or biological categories. For example, one category was a kind of animal that ate blueberries, which caused blue feathers, which attracted mates. Participants were asked to explain why members of the category have blue feathers, which is an ambiguous explanation request – one could answer mechanistically, and appeal to the blueberries, teleologically, and appeal to their role in mate attraction, or both. Critically, responses to the ambiguous why-question predicted later categorization judgments, with participants who provided a teleological explanation underweighting the importance of a feature like "eats blueberries" relative to those who provided only a mechanistic explanation. This effect was sufficiently strong to outweigh previously documented preferences for features that are earlier in a causal chain in categorization (Ahn & Kim, 2000; see also Ahn, 1998). These findings suggest that explanatory mode may play a role in structuring conceptual representations.

The evidence from explanatory preferences and from categorization provides initial reason to suspect that mechanistic and teleological explanations reflect truly distinct explanatory modes, and that these modes have consequences that go beyond the surface structure of explanations. But do these consequences extend to causal reasoning and representation? In particular, do mechanistic and teleological explanatory modes involve distinct criteria for causal ascription? I return to the relationship between explanatory pluralism and causal ascription below, after first introducing a related distinction from philosophy and associated evidence from psychology that will be useful in motivating the experiments that follow.

1.2. Causal pluralism

The most straightforward way to characterize the way in which explanatory mode might impact causal ascriptions is to posit two distinct kinds of causal representation, one corresponding to each explanatory mode. This is almost certainly an oversimplification, but provides a useful starting point. Candidate causal representations come from debates in philosophy and psychology that contrast two very different ways of thinking about causation, which I will refer to as *dependence* theories and *transference* theories (see Godfrey-Smith, 2010; Hall, 2004; Wolff, 2007; for related taxonomies). In this section I briefly characterize each family of theories. The section that follows returns to the relationship between dependence and transference theories, on the one hand, and mechanistic and teleological explanatory modes, on the other.

According to dependence theories, a factor C is a cause of an effect E if E *depends* upon C in the appropriate way. Theories differ in how they formulate the specific dependence relationship required. For example, classic counterfactual theories (e.g. Lewis, 1973; Menzies, 2008) formulate the relationship in terms of counterfactual claims such as: "C is a cause of E if, had C not occurred, E would not have occurred." Interventionist or "manipulability" theories (e.g. Woodward, 2003, 2008) instead emphasize intervention: C is a cause of E if appropriately intervening on C brings about the occurrence of E. Finally, probabilistic theories (see Hitchcock, 2008) require that C raise the probability of E's occurrence.[1] Classic counterfactual, interventionist, and probabilistic theories differ from each other

[1] Counterfactual, interventionist, and probabilistic theories of causation provide more complex and precise characterizations of the requisite dependence relationship; I only sketch underlying intuitions here. For more detailed formulations of theories, the Stanford Encyclopedia of Philosophy is a useful reference.

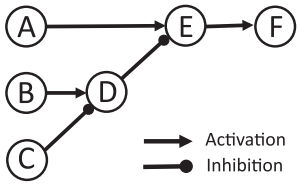

Fig. 1. Representation of a causal structure involving double prevention.

and have numerous variants. However, they share the core notion that the causal relationship is fundamentally about dependence.

Transference theories, in contrast, focus on the physical process by which C brings about E. As a first pass, C is a cause of E if there was physical contact between C and E. Theories differ in how they cash out this intuition about transference (e.g. Dowe, 1992, 2008; Salmon, 1984). Some emphasize the need for a physical connection between cause and effect, others require a transfer of force or some conserved physical quantity, such as momentum. I refer to these theories as *transference* theories to highlight the importance of a transferred quantity between cause and effect. While transference relationships will typically result in dependence relationships, transference theories maintain that a relationship is causal in virtue of the transference, not in virtue of the dependence. Moreover, not all transference relationships require dependence, and many dependence relationships are not mediated by transference.

The differences between dependence and transference theories can be illustrated with well-known examples from the philosophy literature. The examples are intended to isolate dependence from transference, and have been used in arguments for each position. One kind of example involves double prevention (Hall, 2004; see Fig. 1). To illustrate double prevention, consider an almost-thwarted assassination attempt. The gunman fires, but the target's bodyguard rushes to intercept the bullet. However, a bystander accidentally trips the bodyguard, preventing him from preventing the bullet from hitting and killing the target. In this situation, the target's death depends on the gunman: had the gunman not fired the gun, the target wouldn't have died. But the death also depends on the bystander: had the bystander not tripped the bodyguard, the bullet would have been intercepted and the target saved. According to a dependence theory, it seems that both the gunman *and* the bystander are causes of the target's death.[2] However, there was no transference or physical connection between the bystander and the target, as there was between the gunman and the target. Thus according to transference theories, only the gunman is a cause of the target's death.

Research on causal judgments within psychology has generated an intriguing but mixed set of findings about double prevention. Walsh and Sloman (2005, Experiment 3) employed "preempted double prevention": cases in which two agents attempt to remove a barrier to allow a marble to hit a coin and make it land heads. In such cases, the majority of participants judged that neither the agent who succeeded in removing the barrier (the "double-preventer") nor the agent who was attempting to do the same thing (a "preempted double-preventer") caused the coin to land heads. This finding suggests that double preventers are *not* judged causes of outcomes that counterfactually depend upon them. However, Chang (2009, Experiment 1) presented participants with a variety of scenarios that differed in the counterfactual dependence and physical connections involved, and found evidence that double preventers *are* judged as causes. His scenarios involving counterfactual dependence without a physical connection required participants to evaluate the causal status of a double preventer, such as a person (the double preventer) who removed a barrier in front of a train, resulting in the train continuing on to knock over a cardhouse (the outcome). Participants gave quite high causal ratings to the double-preventer, with a mean of over 6 on a 7-point scale. The discrepancy between the results of Walsh and Sloman (2005) and Chang (2009) suggests that double preventers are judged as causes only under some conditions. However, the studies varied in a number of ways that could account for the

[2] Some dependence theories stipulate additional conditions that block factors like the bystander from counting as causes. For example, Ned Hall considers the idea that a cause must bring about the effect intrinsically (Hall, 2004).

difference – only the Walsh and Sloman (2005) scenarios involved preempted double prevention, and causal ratings were solicited differently, with participants in Walsh and Sloman (2005) indicating whether individual factors or pairs of factors caused the outcome, and participants in Chang (2009) rating whether a factor was a cause on a 7-point scale ranging from "definitely no" to "definitely yes."

A related empirical literature on the semantics of causal terms has investigated what might be called "double prevention" at the level of causal types rather than causal tokens. Instead of evaluating the causes of a specific outcome in which a prevention is prevented, participants are presented with general causal relations and asked to draw further inferences. For example, Goldvarg and Johnson-Laird (2001) and Barbey and Wolff (2007) presented participants with information in the form: "A prevents B, B prevents C." Participants were then asked what follows, or to identify the appropriate relationship between A and C (e.g. A causes C, A allows C, or A prevents C). A majority of respondents in Goldvarg and Johnson-Laird indicated "prevents," while the majority in Barbey and Wolff selected "causes" or "allows." The former result would suggest that double preventers are not considered causes, the latter the reverse. It may be that participants in Goldvarg and Johnson-Laird (2001) had difficulty reasoning about negation, or as suggested by Sloman, Barbey, and Hotaling (2009), succumbed to an "atmosphere effect," tending to respond with prevention after premises involving prevention. In any case, it's clear that the current evidence concerning double prevention from psychology is too variable and sparse to support strong conclusions.

Double prevention illustrates the possibility of dependence without transference. Cases of "overdetermination" and "preemption" illustrate the reverse: transference without dependence (Paul, 1998; see Fig. 2). Consider a case of "late preemption" in which two gunmen fire at a target simultaneously. Gunman A's bullet happens to reach the target's heart before Gunman B's bullet, so intuitively Gunman A's actions caused the target's death. However, had Gunman A not pulled the trigger, the target would still have died, as Gunman B's bullet would have hit the target. In this case, the target's death appears *not* to counterfactually depend on Gunman A's actions. However, mirroring the transference relationships, the intuition is that Gunman A caused the death, and that Gunman B did not.

Within philosophy, cases of preemption have provoked a variety of responses. While such cases are typically taken to rule out a very simple dependence theory, dependence theorists have a variety of resources for generating the more intuitive prediction that factors like Gunman A are indeed causes. For example, a dependence theorist could claim that the particular outcome (the target dying at time t) *does* depend on Gunman A, for if his actions were omitted a different outcome (the target dying at time $t + n$) would have ensued. Alternatively, one can claim that the *properties* of the outcome depend on Gunman A, even if the outcome's *occurrence* does not (e.g. Lewis, 2000). Additional responses to late preemption abound (e.g. Halpern & Pearl, 2001; Hitchcock, 2001; Menzies, 2003; Woodward, 2003). For my purposes, the critical point is that late preemption may not succeed in fully isolating dependence from transference; but it does serve as a test case to distinguish the simplest dependence theories from both transference theories and more sophisticated dependence theories.

A few experiments have examined causal judgments in cases of overdetermination or preemption. Mandel (2003) reported several experiments in which participants read vignettes involving two agents who contributed to an outcome, and were then asked to provide causal and counterfactual judgments. For example, in one vignette the target of an assassination plot was given a lethal dose of poison by a first assassin, but was fatally run off the road by a second assassin before the poison could take effect. Participants more often judged the car crash a cause of the death than the poison, but when asked to generate counterfactuals that would undo the outcome, did not "undo" the crash more often than the poison. Mirroring the examples of late preemption considered above, Walsh and

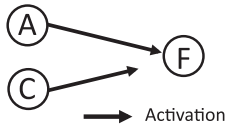

Fig. 2. Representation of a causal structure involving late preemption.

Sloman (2005) presented participants with cases in which two agents simultaneously throw rocks at a bottle or roll marbles to knock over a coin, but one rock or marble reaches the target first. They found that the vast majority of participants identified only the first actor (the one whose rock or marble made physical contact first) as the cause of the outcome. Chang (2009, Experiment 2) reported similar results for late preemption, and additionally considered cases of simultaneous overdetermination in which, say, two rocks hit and break a bottle at the same time. In such cases, causal ratings were middling.

Additional studies have considered cases of "joint causation," in which two causes simultaneously contribute to an outcome, but in which neither alone is sufficient. Spellman and Kincannon (2001) examined situations with multiple sufficient causes as well as situations involving multiple necessary causes to determine the role of counterfactual ("but for") causes in legal decision-making (see also Hart & Honoré, 1985). Participants read about two agents who simultaneously but independently shot a common enemy. One shot went to the head, the other to the heart. In the case of "overdetermination," the coroner determines that either shot alone would have been sufficient to kill the enemy. In the case of "joint causation," the coroner determines that both shots were necessary to kill the enemy. Only in the latter case does the death counterfactually depend on each agent's individual actions – had one of the agents not fired, the enemy wouldn't have died. Nonetheless, participants reported the two agents as causes in both cases, and even assigned higher causal ratings and longer jail sentences to the agents in the overdetermined case, no matter that the effect did not depend on their individual actions. Parallel findings were obtained for situations involving physical forces (e.g. lightning causing a fire).

More recently, Knobe and Fraser (2008) considered cases of joint causation in which an outcome (e.g. a computer crash) required the actions of two agents (e.g. two people logging on), only one of whom was permitted to perform that action. They found that participants judged the agent who wasn't permitted to complete the action more causally responsible than the agent who was. These findings highlight the potential influence of judgments concerning moral responsibility, credit, or blame on judgments of causation.

In sum, research on causal judgments in cases of overdetermination, preemption, and joint causation pose a serious challenge to a simple dependence theory as an account of human causal judgment. However, the current findings do not discriminate between transference theories and a more sophisticated dependence theory.

With some exceptions beyond the empirical work just reviewed (e.g. Gopnik & Schulz, 2007; White, 1990), research in psychology has proceeded almost independently of the literature on dependence and transference theories from philosophy. However, the basic ideas behind dependence and transference theories have been echoed in psychological claims – not about the metaphysics of causation, but about the psychological concepts underlying causal judgments (Danks, 2005; Newsome, 2003). Many approaches to causal learning, for example, focus on co-occurrence data and probabilistic relationships, often with an implicit or explicit commitment to dependence (Gopnik et al., 2004; Griffiths & Tenenbaum, 2005; see Buehner and Cheng (2005) and Sloman (2005) for reviews). Some have argued against such approaches by emphasizing the importance of knowledge of mechanisms in constraining and guiding causal learning and causal attribution (Ahn & Kalish, 2000). There is certainly evidence that children and adults appeal to mechanisms or generative transmission in generating causal judgments (Bullock, 1985; Koslowski, 1996; Schlottman, 1999; Shultz, 1982; White, 1995), and some probabilistic theories have succeeded in accounting for effects of this kind within a dependence-like framework (Griffiths, 2005).

The psychological literature to most closely mirror the distinction between dependence and transference concerns the semantics of causal terms, such as "cause," "enable," and "prevent." In a recent paper, Sloman et al. (2009) developed a *causal model theory* for such terms, which adopts a dependence-like framework known as causal Bayes nets (Pearl, 2000). According to the causal model theory, causal terms express beliefs about the structure of causal models – for example, "A causes B" expresses belief in a dependence relationship between A and B, while "A enables B" additionally implies that A is necessary for B, and that B also has additional causes. An alternative to the causal model theory comes in the form of *force dynamics theory* (Talmy, 1988; Wolff, 2007), a transference theory that analyzes causal terms as distinct patterns of forces. Roughly, "A caused B" implies that A exerted a

force that went counter to the existing tendency of the recipient of the force, while "A enables B" implies that the force contributed to an existing tendency. This approach has recently been augmented to address cases of causation by omission, and by extension double prevention, by recognizing the removal of a pre-existing force as a cause (Wolff, Barbey, & Hausknecht, 2010). The causal model theory and the force dynamics theory have proven difficult to disentangle. To date, most studies investigating the generation or evaluation of causal terms have produced data consistent with both theories (e.g. Barbey & Wolff, in preparation; Sloman et al., 2009).

Given compelling arguments and evidence in favor of both dependence and transference theories, a tempting possibility is to endorse both. In the absence of a workable unified account, it is plausible that dependence and transference theories both reflect psychologically real and distinct ways of thinking about causal relationships. Pluralism about causation has been proposed within philosophy (Godfrey-Smith, 2010; Hall, 2004; Hitchcock, 2003), and the pluralistic intuitions that guide these arguments may well be evidence for pluralism about underlying psychological concepts. The following section considers how a pluralistic picture of causal representations might relate to pluralism about explanation.

1.3. Relating explanatory pluralism and causal pluralism

Why expect a correspondence between explanatory modes and causal concepts? First, explanation and causation are closely related. If explanatory mode influences the representations or processing employed in a given task, it's likely that causal representations underlie or reflect this influence. Second, there are reasons to expect teleological explanations to discourage a transference view of causation. To see why, consider an example from William James, in which he contrasts the behavior of an intentional agent (Romeo) with that of a non-intentional object (iron filings):

> Romeo wants Juliet as the filings want the magnet; and if no obstacles intervene he moves towards her by as straight a line as they. But Romeo and Juliet, if a wall be built between them, do not remain idiotically pressing their faces against its opposite sides like the magnet and the filings [when a card is placed between them]. Romeo soon finds a circuitous way, by scaling the wall or otherwise, of touching Juliet's lips directly. With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed, the path may be modified indefinitely (James, 1890, p. 20).

In this scenario, Romeo's actions exhibit a property referred to as *equifinality*: the same outcome would be achieved despite variation in the means. In contrast, the event involving the iron filings exhibits *multifinality*: variation in the means would generate variation in the outcome. Based on this difference between the behavior of goal-directed agents and most non-intentional systems, the psychologist Fritz Heider (1958) distinguished two kinds of causation. According to Heider, intentional[3] human action – which is goal-directed and hence supports teleological explanations – involves *personal* causation, and accidental human action – which is not goal-directed and supports only mechanistic explanations – involves *impersonal* causation.[4] Subsequent research has confirmed that goal-directed behavior has a special status: causal ratings are typically higher for voluntary and deliberate human actions relative to accidental human actions or to the contributions of physical processes such as wind (e.g. Fincham & Jaspars, 1980; Lagnado & Channon, 2008; Malle, 2004; McClure, Hilton, & Sutton, 2007). However, isolating causal judgments from concerns about moral responsibility or blame can be a challenge when comparing intentional and non-intentional causes.

Neither Heider nor subsequent research has gone on to relate the distinction between personal and impersonal causation to that between dependence and transference. However, personal causation will involve a kind of "mechanism-independence" or "path-invariance" at odds with a transference view of causation. By virtue of the equifinality, situations involving personal causation involve a strong

[3] Throughout this paper, "intentional" is meant as an alternative to "accidental," not as "aboutness" or "at the folk-psychological level," as is sometimes the case in philosophy.

[4] Heider introduced two conditions to distinguish between personal and impersonal causation: equifinality and "local" causation. Only equifinality is discussed here.

dependence relationship that in fact came about in one way, but could just as easily have come about in other ways. The property of mechanism-independence makes goal-directed behavior a good candidate for a dependence concept of causation rather than a transference concept, which emphasizes the specific mechanisms of contact or transmission.

Goal-directed human behavior is a gold standard for equifinality – and is the only case Heider identifies with personal causation – but artifacts and biological adaptations that support teleological explanations may also promote "mechanism-independent" reasoning. Consider a simple device like an alarm clock. Given that it has the function of sounding at a particular time, it's not surprising that many alarm clocks have "back-up" mechanisms: they will draw power from a functioning outlet, but can draw power from batteries should the power go out. Even where an individual artifact does not have contingent mechanisms to achieve a common end, artifacts of a common kind may be unified by virtue of the goals they accomplish, not the mechanism by which they do so. All alarm clocks alert the user at a specified time, but some may be analog, others digital, and so on. Even for biological traits, those that are adaptations may be seen as more likely to obtain a common end despite variation in the means or in the environment (see, for example, Mameli and Bateson's (2006) definitions 9 and 19–22 for the related notion of *innateness*). This mechanism-independent reasoning may be why participants in Lombrozo (2009) who explained category features teleologically were willing to over-ride considerations about a feature's depth in a causal chain in making judgments about category membership.

These observations suggest that a teleological mode of explanation should discourage a transference view of causation, and hence generate judgments more consistent with a dependence theory. In contrast, a mechanistic mode of explanation can accommodate a greater role for transference. As mentioned at the outset, however, a clean one-to-one correspondence between modes of explanation and causal concepts may be too simple. A first elaboration is to consider dependence and transference as two factors that contribute to causal judgments in both a teleological and a mechanistic explanatory mode, but with different weights. Specifically, dependence relationships may be weighted more heavily in evaluating causal relationships from a teleological mode than from a mechanistic mode, with the reverse for transference. The general discussion considers other alternatives, including the possibility that a dependence account might subsume the influence of transference.

1.4. Overview of experiments

In the remainder of the paper I present two sets of experiments that explore the relationship between modes of explanation and causal judgments. As a point of departure, I explore the hypothesis that a teleological mode prompts a dependence concept, and a mechanistic mode a transference concept. This simple correspondence will be challenged in the general discussion.

The experiments utilize cases of double prevention (Experiments 1a, 1b, 2) and late preemption (Experiments 1c, 2) to isolate the influence of dependence and transference in causal judgments. As already noted, late preemption isolates transference at best imperfectly; the experiment thus bears only on a simple dependence theory. To manipulate the mode of explanation participants employ, the agents involved in the events behave in either a goal-directed way (supporting a teleological mode) or accidentally (supporting only a mechanistic mode). Experiment 1a tackles the predictions most directly, examining causal judgments in cases of double prevention involving intentional and accidental behaviors. Experiments 1b and 1c aim to support the interpretation of Experiment 1a by ruling out alternative explanations.

Experiment 2 moves away from the domain of human agents to consider whether the effects found in Experiments 1a–c are restricted to goal-directed, human behavior, or generalize to other situations that support teleological explanations. Specifically, the experiments involve components of artifacts or of biological traits that either have a particular function (and hence support teleological explanations) or generate a consequence incidentally (and hence support only mechanistic explanations). Like Experiments 1a–c, these experiments consider both double prevention and late preemption.

Experiment 2 not only tests the generality of the findings in Experiments 1a–c, but also addresses the worry that findings about causal judgments involving human agents could be clouded by concerns about moral responsibility or blame. To minimize this potential concern the scenarios in Experiments

1a–c involve morally neutral outcomes, unlike the cases with assassins and gunmen considered so far. But by employing artifacts and biological traits Experiment 2 can more decisively sidestep the concern that evaluations of moral responsibility are responsible for the patterns of causal ratings in these experiments.

2. Experiment 1a: intentions and double prevention

Experiment 1a explores the hypothesis that transference relationships play a relatively greater role in causal judgments when those judgments are mode in a mechanistic mode than when they are made in a teleological mode. To isolate the role of transference, Experiment 1a employs scenarios like the following, which involve double prevention:

> Alice, Bob, and Carol have spent the afternoon juggling and listening to music. At the moment, Alice is juggling and the music is not playing. Alice wants to listen to music, so she deliberately throws a juggling ball, which heads straight for the stereo's 'on' button. But while Alice's ball is in the air, Bob starts pulling on the power cord connecting the stereo to the outlet. If Bob unplugs the cord, it will prevent Alice's ball from turning on the stereo and starting the music. However, Carol wants the music to play, so she deliberately steps on the power cord just before Alice's ball hits the 'on' button, preventing Bob's pull from unplugging the stereo. As a result of these events, the music starts to play.

Note that in this scenario, the outcome (the music starting) depends on both the actions of Alice (the "transference" cause) and on the actions of Carol (the "dependence" cause). Thus according to a dependence theory, both Alice and Carol can appropriately be judged causes of the outcome. However, only Alice's actions transfer a force or quantity to the stereo. On a transference theory, Alice is the only cause of the outcome.

To manipulate whether participants construe this event in teleological or mechanistic terms, the agents' behavior can be described as either intentional or accidental. When it is intentional (as in the sample above), the events are naturally construed in terms of goals, and the agents' actions support teleological explanations (e.g. Alice threw the ball to make the music start). When the actions are accidental (Alice accidentally throws the ball; Carol accidentally steps on the power chord), they are not goal-directed, and can only be explained mechanistically.

The prediction is that when Alice and Carol act intentionally, participants will apply a dependence criterion for causation, and both Alice and Carol will be judged causes of the outcome. But when Alice and Carol act accidentally, participants will apply a transference criterion, and judge Alice a more appropriate cause of the outcome than Carol.

2.1. Methods

2.1.1. Participants

Sixty-four undergraduate students and members of the Berkeley community (47% women, mean age 21, range 18–56) participated in exchange for either course credit or monetary compensation.

2.1.2. Materials

The experiment employed two scenarios featuring double prevention. The first (ABC), included above, involved three characters (Alice, Bob, and Carol) who engaged in physical interactions that resulted in music starting to play. The second (DEF) involved three characters (Doug, Eunice, Frank) who used an office supply ordering program to order pens:

> Doug, Eunice, and Frank are work colleagues learning how to use a complicated new system for ordering office supplies. The new system attempts to streamline the ordering process by presenting multiple options with graphical icons rather than with text. While exploring the system, Doug deliberately submits an order for pens. Later that day Eunice is also exploring the system and hits a button that cancels current orders, which would normally cancel Doug's order. However, earlier that day Frank noticed the order for pens and wanted it to go through, so he deliberately hit

a button that disables order cancellations, thereby preventing Eunice's actions from canceling Doug's order. The next morning, Doug, Eunice, and Frank arrive at work to discover that the office supply company has sent pens.

After seeing a scenario, participants were asked to evaluate causal claims. The judgments for the DEF scenario are presented below:

How appropriate is it to make each of the following claims?
Doug caused the office supply company to send pens.
Eunice caused the office supply company to send pens.
Frank caused the office supply company to send pens.

Ratings were made on a 6-point scale of "appropriateness," with each point labeled (1 = completely inappropriate, 2 = mostly inappropriate, 3 = somewhat inappropriate, 4 = somewhat appropriate, 5 = mostly appropriate, 6 = completely appropriate). To ensure that participants understood the (admittedly complex) scenarios, they were also asked to evaluate the truth of 16 statements, 9 of which were false. Below are sample statements for the DEF scenario[5]:

Doug submitted an order for pens. (T)
Doug canceled an order for pens. (F)
Eunice hit a button that cancels current orders. (T)
Eunice submitted an order for pencils. (F)
Frank hit a button that disables order cancellations. (T)
Frank submitted a new order for pens. (F)

The order of the causal evaluation task and the true/false task was counterbalanced, and the true/false statements were presented in one of two random orders.

The ABC and DEF scenarios were deliberately constructed to exhibit double prevention, but also to satisfy an additional constraint: in order to manipulate participants' explanatory mode of construal, the actions had to be plausible as goal-directed, intentional behaviors and also as accidental behaviors. Each scenario had four versions: one in which both the transference and dependence causes acted intentionally, one in which they both acted accidentally, and two in which only one or the other acted intentionally.

Although the ABC and DEF scenarios share these key features in common, they also vary in notable ways. ABC involves observable, physical actions while DEF involves less observable, electronic interactions, and in ABC the first preventer (B) acts before the dependence cause (C), while in DEF the reverse is true. This variation was included to ensure that an effect, if found, was not due to the specific properties or temporal unfolding of the given event.

2.1.3. Procedure

Each participant read and evaluated a single version of a single scenario. Participants were randomly assigned to one of 32 conditions, the result of crossing two scenarios (ABC, DEF) with two values for the intentional status of the transference cause (intentional, accidental) and two values for the intentional status of the dependence cause (intentional, accidental), two possible orders for the tasks (causal evaluations first, true/false statements first) and two random orders for the true/false statements. The questionnaires were completed as part of a series of experiments, no other of which involved evaluating causal claims.

2.2. Results

The primary predictions concern the influence of intentional status (whether an action was performed intentionally or accidentally) on causal ratings for the transference and dependence causes. However, before examining these predictions three things are worth noting. First, performance on the true/false task averaged 94% (SD = 8.89, min = 62.5%, max = 100%), was well above chance

[5] Complete stimulus materials for all experiments are available from the author upon request.

responding (one-sample t-test, $t(63) = 39.9$, $p < .01$), and did not vary as a function of condition. This suggests that participants understood the causal structure of these complex scenarios, and that differences across conditions are unlikely to be due to differences in comprehension. Second, task order (causal evaluations first or true/false first) did not yield significant main effects or interact with other variables when included in ANOVAs with the variables of interest, suggesting that completing the true/false task did not influence causal judgments. Finally, participants gave higher causal ratings to the candidate transference and dependence causes (Alice and Carol, Doug and Frank) than to the "control" causes (the first preventers: Bob, Eunice), which were included for comparison; the means were 5.08 (SD = 1.10), 4.66 (SD = 1.30), and 1.41 (SD = .988), respectively. All means differed reliably from each other (transference versus dependence: $t(63) = 2.03$, $p < .05$, first preventer versus dependence, $t(63) = -16.10$, $p < .01$), first preventer versus transference, $t(63) = -19.00$, $p < .01$.).

The mean ratings for the transference cause (Alice, Doug) and the dependence cause (Carol, Frank) are indicated in Fig. 3 as a function of intentional status. Consistent with the prediction that intentional actions will prompt a teleological construal that in turn supports a dependence criterion for causation, participants in the condition in which both agents acted intentionally did not provide reliably different ratings for the two causes (paired-samples t-test, $t(15) = .22$, $p = .83$). And consistent with the prediction that accidental actions will prompt a mechanistic construal that supports a transference criterion for causation, participants in the condition in which both agents acted accidentally provided significantly higher causal ratings for the transference cause than for the dependence cause (paired-samples t-test, $t(15) = 2.70$, $p < .05$). To compare these effects, a mixed ANOVA was performed for just those conditions in which both agents acted intentionally or accidentally, with the cause being evaluated (transference, dependence) as a within-subjects factor, intentional status (both intentional, both accidental) as a between-subjects factor, as well as scenario (ABC, DEF), task order (Causal ratings first, true/false first), and true/false question order. Critically, the interaction between the cause being evaluated (transference, dependence) and intentional status was significant, $F(1, 16) = 4.9$, $p < .05$, and did not interact with scenario, $F(1, 16) = 2.5$, $p = .13$.

For the cases in which a single agent acted intentionally, two trends are worth noting. First, the single, intentional actor was rated a more appropriate cause, whether that actor was the transference cause or the dependence cause. Second, this benefit was asymmetrical, with a greater relative boost when the transference cause was the only intentional actor. To analyze the data from all four intentional status conditions more systematically, a difference score was computed for each participant, consisting of the rating for the transference cause minus that for the dependence cause. A $2 \times 2 \times 2$ ANOVA with the intentional status of the transference cause (intentional, accidental), the intentional status of the dependence cause (intentional, accidental), and scenario (ABC, DEF) as between-subjects

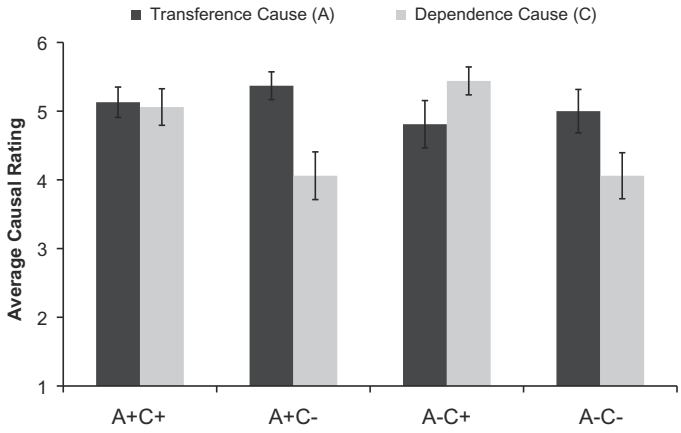

Fig. 3. Average causal ratings for the transference and dependence causes in Experiment 1a as a function of intentional status: when both were intentional (A+C+), only the transference cause was intentional (A+C−), only the dependence cause was intentional (A−C+), or both were unintentional (A−C−). Error bars indicate one standard error of the mean in each direction.

factors and difference score as the dependent measure revealed a single, significant effect: the intentional status of the dependence cause, $F(1, 56) = 13.64$, $p < .01$. The intentional status of the dependence cause influenced the relative ratings participants provided for the two candidate causes, with greater differences between ratings when the dependence cause was accidental than when it was intentional. Interestingly, the intentional status of the transference cause did not have a comparable effect, $F(1, 56) = 1.95$, $p = .17$.

It may be that the intentional status of the transference cause had a minimal effect because the causal ratings for the transference cause were already so high. To help rule out this possibility, a final analysis considered the relative ratings for the two candidate causes, irrespective of the magnitude of those ratings. Participants were classified into three groups: those who provided a higher rating for the transference cause than for the dependence cause, those who provided equal ratings, and those who provided higher ratings for the dependence cause. Fig. 4 illustrates the distribution of participants exhibiting each pattern as a function of intentional status. The distributions differ significantly across conditions, $\chi^2(6) = 14.1$, $p < .05$, and reveal that the pattern of results in Fig. 3 reflects the modal response for each condition.

2.3. Discussion

Experiment 1a confirms the prediction that explanatory mode – as reflected by intentional versus accidental behavior – has an impact on the relative importance of a transference mechanism in evaluating causal claims. When an effect was brought about by two agents who acted intentionally, participants gave comparable causal ratings to both agents, even though one agent shared a transference relationship with the effect, and the other (the double preventer) only a dependence relationship. However, when the two agents acted accidentally, participants continued to give a high causal rating to the transference cause, but gave a lower causal rating to the dependence cause. This suggests that the presence of a transference relationship played a more minor role in evaluating relationships supporting teleological explanations than in those supporting only mechanistic explanations.

While the findings from Experiment 1a are consistent with the prediction that teleological and mechanistic modes of explanation prompt different criteria for causal ascriptions, other interpretations are possible. For example, it could be that the double-preventer's intention to bring about the outcome – not whether a teleological explanation was supported – was responsible for boosting causal ratings for the dependence cause. Experiments 1b and 1c examine two versions of this alternative explanation.

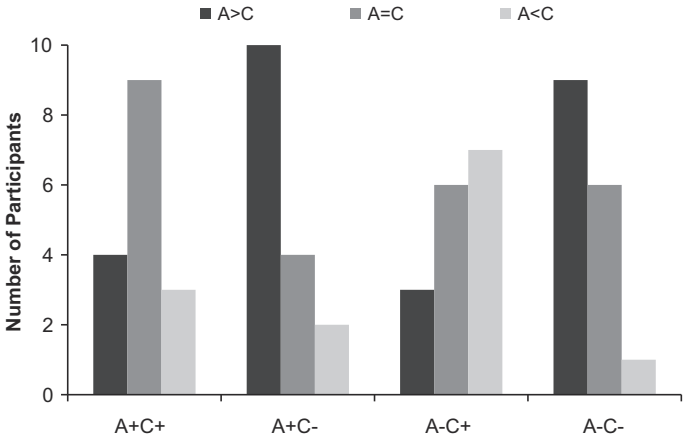

Fig. 4. Participants in Experiment 1a were classified into three categories depending on causal ratings: those who gave a higher rating to the transference cause to than the dependence cause (A > C), those who gave equal ratings to both (A = C), and those who gave higher ratings to the dependence cause than to the transference cause (A < C). The number of participants exhibiting each pattern is indicated as a function of the intentional status of the transference and dependence causes.

3. Experiment 1b: intentions and deviant causation

Experiment 1b aims to rule out a deflationary explanation for the findings in Experiment 1a, namely that the influence of intentions on causal ratings was not mediated by explanatory mode, but by the agents' intentions per se. For example, the double preventer (e.g. Carol) or the relationship between the double preventer and the outcome may simply have become more salient when Carol had the intention to bring about the outcome.[6] If the relationship between Alice and the outcome was already salient, intentions may not have had a comparable effect for causal ratings of Alice. This alternative can be distinguished from the hypothesis that explanatory modes influence causal judgments with cases of so-called "deviant" causation. Consider the following variant of the ABC scenario from Experiment 1a:

> Alice, Bob, and Carol have spent the afternoon juggling and listening to music. At the moment, Alice is juggling and the music is not playing.
>
> Alice wants the music to start, and plans to throw a juggling ball at the stereo's 'on' button. However, the thought of doing so makes her so nervous that she loses a juggling ball, which happens to head straight for the stereo's 'on' button.
>
> But while Alice's ball is in the air, Bob starts pulling on the power cord connecting the stereo to the outlet. If Bob successfully unplugs the cord, it will prevent Alice's ball from turning on the stereo and starting the music.
>
> However, Carol also wants the music to start, and decides she'll step on the power cord, thereby preventing Bob's pull from unplugging the stereo. The thought of doing so makes her so nervous that she loses her balance, but she happens to land with a foot on the power cord, preventing Bob from pulling out the power cord after all.
>
> As a result of these events, the ball reaches the on button, the stereo turns on, and the music begins to play.

This scenario preserves the counterfactual dependence relationships that characterize double prevention, and also involves intentions that match those from the intentional condition in Experiment 1a.[7] However, cases of deviant causation may influence the acceptability of teleological explanations (e.g. Peacocke, 1979). Consider whether it is appropriate to claim that Alice threw a juggling ball *because* she wanted the music to start, or that Carol stepped on the power cord *because* she wanted the music to start. If cases of deviant causation do not support a teleological construal, then causal ratings for deviant cases should mirror those for accidental actions, with ratings for the transference cause (e.g. Alice) higher than those for the dependence cause (e.g. Carol). In contrast, if the findings from Experiment 1a are due to salience or other consequences of the double-preventer's intention to bring about the outcome, deviant cases should mirror intentional cases, with both causes given comparable ratings.

3.1. Methods

3.1.1. Participants

Two-hundred and thirty-nine participants (58% women, mean age = 36, range 18–81) were recruited from an on-line platform (Amazon Mechanical Turk) and participated in exchange for monetary compensation.

3.1.2. Materials

The experiment employed two scenarios involving double prevention, modeled after the ABC and DEF scenarios in Experiment 1a. The causal relationships between the transference and dependence

[6] I thank an anonymous reviewer for raising this possibility.

[7] In cases of deviant causation, it's important to distinguish an intention to bring about an outcome from having brought an outcome about *intentionally* (see, for example, Wilson, 2009). While Alice had the intention to make the music start, and her intention caused her to throw the ball that made the music start, one might not judge her to have made the music start intentionally.

causes and the outcome were either "normal," mirroring the intentional condition from Experiment 1a, or "deviant" as in the ABC scenario included in the introduction to this experiment.

After reading the scenario, participants were asked to evaluate a causal claim for each agent (e.g. "Alice caused the music to start"), an explanation for each agent's actions (e.g. "Why did Alice throw the ball? Because she wanted the music to start."), and a counterfactual involving each agent's actions (e.g. "If Alice had not thrown the ball, the music wouldn't have started."). Causal ratings and explanation ratings were made on 6-point scales; the counterfactual judgments were presented as true/false questions. As before, the causal rating for Bob served as a comparison, as did the explanation rating ("Why did Bob try to pull the power cord? Because he wanted the music to start.") and counterfactual judgment ("If Bob had pulled the power cord, the music wouldn't have started."). The explanation ratings served not only as a manipulation check to confirm that teleological explanations were better supported in the normal cases than in the deviant cases, but also to examine the relationship between causal ratings and explanation ratings across participants.

3.1.3. Procedure

Participants read and evaluated a single scenario, selected from the 4 options obtained by crossing scenario (2: ABC, DEF) with causal condition (2: normal, deviant). Participants were randomly assigned to condition, with 59 to 61 participants per condition.

3.2. Results

Mean responses for all judgments are indicated as a function of causal condition (normal, deviant) in Table 1. First, it's worth noting that performance on the counterfactual task was high (87–95% correct across conditions), and that deviant conditions had the intended effect of reducing the acceptability of teleological explanations for both the transference cause (4.15 versus 5.39; e.g. "Why did Alice throw the ball? Because she wanted to music to start") and the dependence cause (4.42 versus 4.95; e.g. "Why did Carol step on the power cord? Because she wanted the music to start").

To test the key prediction that transference plays a greater role in the deviant condition – which only moderately supports teleological explanations – than in the normal condition – which strongly supports teleological explanations – causal ratings for the transference and dependence causes were analyzed as the dependent measures in a repeated-measures ANOVA with scenario (2: ABC, DEF) and causal condition (2: normal, deviant) as between-subjects factors (see Fig. 5). This analysis revealed a main effect of the cause being evaluated, $F(1235) = 31.0$, $p < .01$, with the transference cause receiving higher ratings than the dependence cause (4.84 versus 4.23), as well as the predicted interaction between the cause being evaluated and the causal condition, $F(1235) = 5.52$, $p < .05$. In the normal condition, participants provided ratings for the transference cause that were only slightly higher than for the dependence cause (4.94 versus 4.59), but in the deviant condition this difference was significantly larger (4.74 versus 3.88). These effects were qualified by a three-way interaction between the cause being evaluated, the causal condition, and the scenario, $F(1235) = 4.70$, $p < .05$: the predicted

Table 1
Judgments from Experiment 1b as a function of causal condition. Means are followed in parentheses by standard deviations. Causal and explanation ratings were made on 6-point scales, with a 6 indicating that it is appropriate to make a causal claim, or that an explanation is very good. The (∗∗) indicates a significant difference across conditions with an independent-samples *t*-test (all *ps* < .01).

	Normal condition	Deviant condition	Difference
Transference cause rating (e.g. Alice)	4.94 (1.15)	4.75 (1.25)	.20
First preventer rating (e.g. Bob)	1.68 (1.17)	1.78 (1.18)	−.10
Dependence cause rating (e.g. Carol)	4.59 (1.35)	3.88 (1.45)	.71**
Transference cause explanation	5.39 (.94)	4.15 (1.56)	1.25**
First preventer explanations	1.50 (1.07)	1.78 (1.32)	−.28
Dependence cause explanation	4.95 (1.21)	4.42 (1.57)	.53**
Transference cause counterfactual	92% correct	95% correct	−3%
First preventer counterfactual	87% correct	89% correct	−2%
Dependence cause counterfactual	93% correct	88% correct	5%

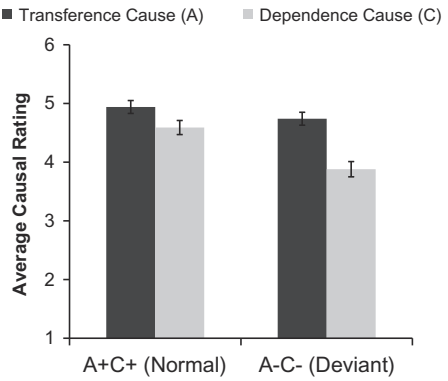

Fig. 5. Average causal ratings for the transference and dependence causes in Experiment 1b as a function of causal condition: normal (A+C+) or deviant (A−C−). Error bars indicate one standard error of the mean in each direction.

interaction was driven by the DEF scenario, although the ABC scenario followed the same numerical trend. For the DEF scenario, the transference cause was rated .10 units (SD = 1.76) higher than the dependence cause in the normal condition, with a significantly greater difference of 1.08 (SD = 1.72) in the deviant condition (post hoc *t*-test, $t(116) = -3.07$, $p < .01$). In the ABC scenario, an equivalent post hoc *t*-test was not statistically significant (normal: $M = .60$, SD = 1.56; deviant: $M = .64$, SD = 1.67; $t(119) = -.13$, $p = .89$). Finally, there was a main effect of causal condition, $F(1235) = 12.62$, $p < .01$, with higher causal ratings in the normal condition than in the deviant condition (4.76 versus 4.31).

Experiment 1b also provides the opportunity to examine the relationship between causal ratings and explanation ratings. If causal judgments are influenced by explanations, causal ratings should be significantly correlated with explanation ratings. This was indeed the case. There were significant correlations between causal ratings and explanation ratings for both the transference cause ($r = .24$, $p < .01$) and the dependence cause ($r = .38$, $p < .01$). These effects were independently observed in the normal condition (transference: $r = .31$, $p < .01$; dependence: $r = .28$, $p < .01$) and in the deviant condition (transference: $r = .19$, $p < .05$; dependence: $r = .41$, $p < .01$). A more subtle prediction is that the relative contribution of transference in causal ascriptions should be lower to the extent a participant adopted a teleological mode of construal. The relative contribution of transference should be reflected in the difference between causal ratings for the transference and dependence causes, and the extent of teleological reasoning in the average explanation ratings for the transference and dependence causes. Correlating these quantities revealed a small but significant relationship in the predicted direction, $r = -.13$, $p < .05$: participants who privileged transference over dependence in causal ratings tended to provide lower ratings for the teleological explanations.

3.3. Discussion

The findings from Experiment 1b support the proposed interpretation of Experiment 1a. When the acceptability of a teleological mode of explanation was manipulated by introducing deviant causal chains, ratings mirrored those in the accidental condition from Experiment 1a, with transference playing a relatively greater role in deviant cases than in normal cases. This suggests that the effects in Experiment 1a were not driven by the agents' intentions per se, as agents in both the normal and deviant cases in Experiment 1b intended to bring about the outcome. Rather, the relative importance of transference in causal ascriptions seems to depend on whether the agents' actions support teleological explanations. This interpretation is bolstered by the reported relationships between causal ratings and explanation ratings.

Experiment 1c further examines the influence of intentions in causal judgments by considering non-deviant cases of intentional action, but with late preemption rather than double prevention.

4. Experiment 1c: intentions and late preemption

Experiment 1a and 1b involve scenarios with double prevention to tease apart the relative contributions of transference and dependence relationships in causal ascriptions. Specifically, the double preventer shares a dependence relationship with the effect, but not a transference relationship. Cases of preemption likewise aim to distinguish transference from dependence, but do so by attempting to isolate transference. Consider a scenario in which Alice (the first cause) and Carol (the second cause) both throw juggling balls that head straight for the "on" button of a stereo. Alice's ball happens to reach the button first, and the music starts to play. But had Alice's ball not been there, the music would still have started to play, as Carol's ball would have hit the button. In a case like this, Alice's actions and the music starting share a transference relationship, and it seems natural to say that Alice caused the outcome. But arguably, the music does not *depend* on Alice. For this reason, cases of preemption have been posed as a challenge for simple dependence theories of causation. However, both transference theories and more complex dependence theories can accommodate these intuitions.

Examining the influence of intentions in cases of preemption can nonetheless clarify the interpretation of the findings from Experiments 1a and 1b. While Experiment 1b suggests that merely adding an intention to achieve the outcome is insufficient to boost causal ratings for the dependence cause, it remains a possibility that intentions do play a role for non-deviant causes. Cases of late preemption can potentially address this concern: if intentions indiscriminately boost causal ratings for any non-deviant cause, then participants should give higher ratings to the first and second causes when intentional than when accidental.

Experiment 1c can address an additional, alternative explanation for Experiments 1a: that participants have a transference concept of causation, but with intentions treated as (metaphorical) mechanisms of causal transmission. This alternative would predict the current pattern of findings, with either "physical" transference or "psychological" transference (i.e. intentional actions) generating high causal ratings. Consistent with this possibility, Wolff (2003) reports that the relationship between the start and end of a causal chain is more likely to be perceived as a single event involving "direct causation" if the causal chain is initiated intentionally. If participants treat intentions as metaphorical mechanisms of causal transmission – even in the absence of corresponding physical mechanisms – then the second cause should receive higher ratings when intentional, as the intention should partially compensate for the absence of a completed, physical transmission process.[8]

In sum, examining the influence of intentions on causal ascriptions in scenarios involving late preemption can clarify the interpretation of Experiments 1a and 1b by examining whether intentions indiscriminately boost causal ratings for borderline causes, or act as metaphorical mechanisms for causal transmission even in the absence of physical forces.

4.1. Methods

4.1.1. Participants

Sixty-four Berkeley undergraduates and members of the Berkeley community (67% women, mean age = 20, range 18–34) participated in exchange for course credit or monetary compensation.

4.1.2. Materials

The experiment employed two scenarios involving late preemption, modeled after the ABC and DEF scenarios in Experiment 1a. Below is a sample scenario:

> Alice, Bob, and Carol have spent the afternoon juggling and listening to music. At the moment, Alice and Carol are juggling and the music is not playing. Alice wants to listen to music, so she deliberately throws a juggling ball, which heads straight for the stereo's 'on' button. Meanwhile, Bob switches on his laptop to check his e-mail. And at the same time, Carol decides she wants to listen

[8] Not all views that treat intentions as metaphorical forces need generate this prediction. Wolff (personal communication) points out that the force dynamics model does not make a clear prediction in this case. The theory has not been applied to situations in which an "intentional" force is not accompanied by a (possibly indirect) physical force.

to music, so she also deliberately throws a juggling ball, which heads straight for the stereo's 'on' button. Alice happens to throw first and her juggling ball reaches the stereo's 'on' button before Carol's ball does. As a result of these events, the music starts to play.

After reading the scenario, participants were asked to evaluate causal claims and 16 true/false statements as in Experiment 1a.

Also as in Experiment 1a, the agents' actions were designed to be plausible as either intentional or accidental actions, and intentional status varied across conditions.

4.1.3. Procedure

Participants read and evaluated a single scenario, selected from the 32 options obtained by crossing the scenario (2: ABC, DEF), the intention status of the first cause (2: intentional, accidental), the intentional status of the second cause (2: intentional, accidental), the task order (2: causal ratings first, true/false first), and the order of the true/false questions (2 random orders). Participants were randomly assigned to condition, and completed the questionnaire as part of a series of experiments, no other of which involved causal ratings.

4.2. Results

Before proceeding to the causal ratings, it is worth noting that performance on the true/false comprehension task was high (mean = 95.3%, SD = 6.6, range 75–100%), and significantly better than chance performance (one-sample *t*-test, $t(63) = 55.02$, $p < .01$). Performance was slightly higher for the DEF scenario than for the ABC scenario, $F(1, 32) = 4.46$, $p = .043$, 97.3% correct (SD = 6.1) versus 93.6% (SD = 6.8), but did not otherwise vary as a function of condition when included in an ANOVA with the variables of interest (*p*s > .15). It is also worth noting that the order of the true false questions and the order of the tasks (true/false first or causal ratings first) did not generate main effects or interact with any variables when included in an ANOVA with the variables of interest; they are thus excluded from further analyses.

Collapsing across all between-subjects variables, there was a significant effect of the candidate cause (e.g. Alice, Bob, Carol) being rated, $F(2126) = 314.9$, $p < .01$. The mean ratings were 5.61 of 6 (SD = .83) for the first cause, 2.28 (SD = 1.46) for the second cause, and 1.25 (SD = .62) for the "control cause" (e.g. Bob, Eunice). All pair-wise differences were significant (paired-samples *t*-tests, *p*s < .01).

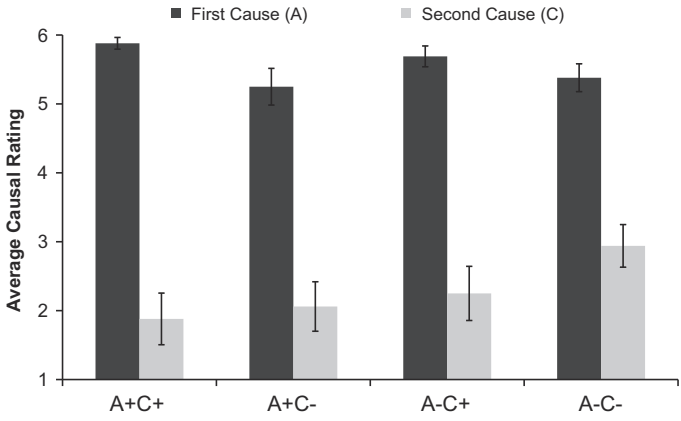

Fig. 6. Average causal ratings for the first and second causes in Experiment 1c a function of intentional status: when both were intentional (A+C+), only the first cause was intentional (A+C−), only the second cause was intentional (A−C+), or both were unintentional (A−C−). Error bars indicate one standard error of the mean in each direction.

To examine the influence of intentions on causal ratings (see Fig. 6), the data were analyzed with a mixed ANOVA with causal ratings as a dependent measure, the cause being evaluated (2: first cause, second cause) as a within-subjects factor, and three between-subjects factors: the intentional status of the first cause (2: intentional, accidental), the intentional status of the second cause (2: intentional, accidental), and the scenario (2: ABC, DEF). This analysis revealed a main effect of the cause being evaluated, $F(1, 56) = 2.61$, $p < .01$, with higher ratings for the first cause than for the second cause, as well as an interaction between the cause being evaluated and the intentional status of the second cause, $F(1, 56) = 5.03$, $p < .05$. Curiously, the intentional status of the second cause did not have a significant impact on causal ratings for the second cause (one-way ANOVA, $F(1, 62) = 1.44$, $p = .24$), but did have a significant impact on causal ratings for the first cause (one-way ANOVA, $F(1, 62) = 6.35$, $p < .05$): the causal ratings for the first cause (e.g. Alice) were higher when the second cause (e.g. Carol) was intentional than when the second cause was accidental, 5.78 of 6 (SD = .49) versus 5.31 (SD = .93). This effect was not predicted; it may be that the second cause's failure to achieve an intended outcome highlighted the first cause's efficacy by establishing a competitive context. There were no effects of scenario version.

Overall, these findings are consistent with a transference theory, but can potentially be accommodated by a complex dependence theory. The latter possibility is supported by a follow-up experiment in which 32 participants read the stimulus materials, and were asked to evaluate two claims: a claim about dependence (e.g. "The outcome (the music starting) depended on Alice.") and a claim about influence (e.g. "The outcome (the music starting) would have occurred a bit differently if not for Alice."). Twenty of the 32 participants judged the dependence claim appropriate, and 18 judged the influence claim appropriate. In fact, only 5 of 32 participants rejected both claims, suggesting that participants' intuitive notion of dependence is not the simple formulation challenged by late preemption.

4.3. Discussion

Whether the agents acted intentionally or accidentally, participants gave high causal ratings to the first cause – the one that made contact with the effect – and gave low causal ratings to the second cause – the one that would have made contact if not preempted. It's noteworthy that even the low ratings for the second cause were reliably higher than those for the "control" cause – the agent who performed an unrelated action in the story. This suggests that the second cause has an intermediate causal status, and thus serves as a good test case for an alternative explanation for Experiment 1a, namely that intentions boost the causal ratings for any marginal cause. This alternative explanation is not supported by Experiment 1c: the ratings for the second cause were not reliably influenced by the intentional status of the second cause.

The finding that the first cause received uniformly high ratings can be interpreted in two ways. The first possibility is that the scenarios employed genuinely isolated transference from dependence, and thus demonstrate that transference is sufficient for a causal relationship, and dependence unnecessary. Another possibility, consistent with a large literature in philosophy and with the follow-up experiment in which a majority of participants claimed that the outcome *did* depend on the first cause, is that these scenarios do not succeed in fully isolating transference from dependence. If this is the case, the findings are compatible with a sophisticated dependence theory.

While the findings cannot distinguish a sophisticated dependence theory from a transference theory, they do provide a *prima facie* challenge to a transference theory according to which intentions (metaphorically) supply a mechanism of causal transmission, even in the absence of corresponding physical mechanisms. Such a view would predict that the first and second causes should be differentiated more strongly when the agents act accidentally (as only one will have a transference relationship with the effect) than when both act intentionally (as they should both have a transference relationship with the effect, albeit a metaphorical one for the second cause; see footnote 8).

Experiment 2 provides a stronger test of the hypothesis that explanatory mode – and not intentions per se – influence causal ratings by considering other kinds of relationships that support teleological explanations: the relationship between an artifact part and its function, and the relationship between a biological trait and its function.

5. Experiment 2: functions, double prevention and preemption

Experiments 1a–c attempt to isolate the unique contributions of transference and dependence relationships in teleological and mechanistic explanatory modes. This is accomplished by varying whether human behavior supports a teleological explanation or does not. Goal-directed, intentional human behavior is a prototypical case for teleology, but it is not the only kind of event or property that supports teleological explanations. In particular, artifact parts or properties typically support teleological explanations (e.g. the knife is sharp because it is for cutting), as do biological traits (e.g. flamingos have special beaks to filter food from mud and silt). But just as human behaviors can be either intentional or accidental, artifact parts and biological traits can be functional or incidental. For example, a knife might be shiny, but the shininess does not (typically) support a teleological explanation – it is merely a side effect of the material. Similarly, flamingos are typically pink, but this is the result of their diet, and need not be an adaptation.[9]

There are two advantages to extending the predictions from Experiments 1a–c to artifacts and biological traits. First, replicating the findings from these experiments for scenarios in these domains would argue for the generality of the effects, and provide strong support for the hypothesis that explanatory mode *in general* influences causal judgments. In particular, Experiment 2 addresses the concern that teleological explanations for human behavior may be unique because they are also intentional. While a creator's intentions play a central role in how artifacts are conceptualized (e.g. Diesendruck, Markson, & Bloom, 2003; Gelman & Bloom, 2000; Matan & Carey, 2001), the artifact and its properties are themselves non-intentional. And among participants who regard biological traits as the result of natural selection (and not divine creation), intentional explanations should play no role in supporting teleological explanations (Lombrozo & Carey, 2006).[10]

Second, examining cases other than human behavior can address the concern that the findings from Experiments 1a–c reflect something special about causal judgments involving moral agents. Specifically, considerations about agents' moral responsibility for outcomes could influence ratings. Previous research finds that agents are typically judged more harshly when they generate negative outcomes intentionally (e.g. murder versus manslaughter, e.g. Cushman, 2008), and more recent work suggests that similar effects may permeate causal judgments (e.g. Hitchcock & Knobe, 2009; Lagnado & Channon, 2008; McClure et al., 2007). For these reasons, the effects in Experiments 1a–c were designed to be as morally neutral as possible (music starting, pens being ordered), but there is a lingering concern that mechanisms specific to moral evaluation could nonetheless be triggered by any scenario involving human action. If moral evaluations are responsible for the effects in Experiments 1a–c, then manipulating whether artifact parts and biological traits bring about effects functionally or incidentally should not have comparable effects.

Experiment 2 thus examines cases of double prevention and late preemption involving both artifacts and biological traits, where outcomes are brought about to fulfill their function (supporting a teleological explanation) or incidentally (supporting only a mechanistic explanation).

5.1. Methods

5.1.1. Participants

Participants were 192 Berkeley undergraduates, summer school students, and members of the Berkeley community (72% women, mean age = 20, SD = 3, range 18–46) who participated in exchange for course credit or monetary compensation. Six additional participants were replaced for leaving answers blank.

5.1.2. Materials

Participants read scenarios involving double prevention and late preemption, and involving either biological traits or artifact parts. For half of the participants, the candidate causes were functional; for

[9] In fact, it may be that the pink color serves as a signal to potential mates, or has some other adaptive function. I thank a reviewer for raising this possibility.

[10] For participants who regard biological kinds as the product of divine creation, biological kinds should have an equivalent status to artifacts in terms of their relationship to intentional explanations.

the other half, they generated an outcome incidentally. Below is a sample double prevention scenario, with the extra information provided for participants in the functional condition indicated in brackets:

> The diet of a certain kind of Australian shrimp consists of three kinds of foods: alphaplankton, bacterioplankton, and cromplankton. Alphaplankton contain chemical A, which triggers a reaction that changes the shrimp's skin to make it reflect high frequencies of ultraviolet light. Bacterioplankton contain chemical B, which neutralizes chemical A and thereby prevents the shrimp from reflecting high frequencies of ultraviolet light. However, cromplankton contain chemical C, which binds to chemical B and thereby prevents chemical B from preventing chemical A from making the shrimp reflect high frequencies of ultraviolet light. Because of these interactions, a shrimp that has eaten alphaplankton, bacterioplankton, and cromplankton will reflect high frequencies of UV light.
>
> [Reflecting high frequencies of UV light is biologically important, as it aids the shrimp in regulating its temperature by reflecting the frequencies of light with the most energy. In fact, while eating bacterioplankton is important for nutritional reasons, Australian shrimp have evolved to eat alphaplankton and cromplankton because these foods result in the reflection of high frequencies of UV light and thereby improve temperature regulation.]

After reading this initial information, participants completed 24 "type" true/false questions to assess comprehension of the general causal structure, and then read about a specific event for which they provided causal ratings:

> Your friend shows you his aquarium, which contains an Australian shrimp of this type (specimen S). Specimen S has eaten alphaplankton, bacterioplankton, and cromplankton. As usual, chemical C in the cromplankton binds to chemical B in the bacterioplankton, preventing it from neutralizing chemical A in the alphaplankton. After eating these foods specimen S reflects high frequencies of ultraviolet light.
>
> Given the scenario above, how appropriate is it to make each of the following claims?
>
> The alphaplankton caused specimen S to reflect high frequencies of UV light.
> The bacterioplankton caused specimen S to reflect high frequencies of UV light.
> The cromplankton caused specimen S to reflect high frequencies of UV light.

There were an additional 5 "token" true/false comprehension questions for the particular event (e.g. "If specimen S had not eaten alphaplankton, it would not have reflected high frequencies of ultraviolet light.").

The scenarios involving late preemption followed a similar structure, illustrated here with an artifact scenario. Again, the functional information provided to participants in the function condition is indicated in brackets:

> Furniture made of a certain kind of Australian wood has three coatings: alpha compound, beta seal, and crom compound. The alpha compound contains chemical A, which triggers a reaction that makes the wood reflect high frequencies of ultraviolet light. The beta seal contains chemical B, which results in the synthesis of a transparent layer. The crom compound contains chemical C, which also triggers a reaction that makes the wood reflect high frequencies of UV light. Thus using either alpha compound or crom compound is sufficient to make the wood reflect high frequencies of ultraviolet light, while using beta seal results in the synthesis of a transparent layer unrelated to the reflection of high frequencies of UV light.
>
> [The reflection of these frequencies is functionally important, as it prevents furniture made of the wood from warping and discoloration by reflecting the frequencies of light with the most energy. In fact, while beta seal is important for making the furniture stain-resistant, furniture made of this kind of Australian wood is coated with alpha compound and crom compound because these coatings result in the reflection of high frequencies of UV light and thereby protect the wood from warping and discoloration.]
>
> Your friend shows you a dresser he built out of this type of wood. The dresser has just been coated with alpha compound, beta seal, and crom compound. The wood happens to absorb the alpha

compound more quickly than the crom compound, so the chemical A from the alpha compound triggers the reaction that makes the wood reflect high frequencies of ultraviolet light. But had the alpha compound been absent, chemical C from the crom compound would have triggered the reaction that makes the wood reflect high frequencies of ultraviolet light.

The late preemption scenarios also involved true/false comprehension questions and causal ratings.

5.1.3. Procedure

Each participant read and evaluated two scenarios from the same domain, one involving double prevention and the other late preemption, with order counterbalanced. There were two versions of each scenario, so if participants received the double prevention scenario involving light reflection presented above (the light version), for example, they would receive a different version of the preemption scenario (the heat version). There were 64 distinct questionnaires, the result of crossing domain (2: biological, artifact), functional status (2: functional, incidental), scenario version (2: light version, heat version), scenario order (2: double prevention first, preemption first), task order (2: token true/false first, causal ratings first), and true/false question order (2 random orders). Participants were randomly assigned to one of these 64 versions, and completed the task as part of a series of experiments, no other of which involved causal ratings.

5.2. Results

For clarity, the results for double prevention and preemption are reported separately. The pattern of findings did not vary as a function of the order of these tasks.

5.2.1. Double prevention

Participants performed acceptably on the true/false comprehension task, with an average of 94.3% (SD = 8.0, range 33–100%) on the type true/false questions, and 76.6% (SD = 23.0, range 20–100%) on the token true/false questions. Performance on both tasks was well above chance (one-sample *t*-tests, *p*s < .01), and performance on the type true/false task did not vary across conditions. However, for token true/false accuracy, there was a main effect of domain, $F(1, 184) = 4.84$, $p < .05$, a main effect of scenario version, $F(1, 184) = 29.80$, $p < .01$, and a domain by version interaction, $F(1, 184) = 103.7$, $p < .01$. These effects appear to have been driven by especially poor performance for the light version in the biological domain (53.8%, SD = 13.8, versus 73.4–93.8% for other conditions), but it is not clear why performance was discrepant in this condition.

Before examining the influence of functional status on causal ratings for the first cause and the double preventer, the causal ratings were analyzed with a repeated-measures ANOVA with the cause being rated (first cause, first preventer, double preventer) as a within-subjects variable, and collapsing across all between-subjects variables. This analysis revealed a main effect of the cause being evaluated, $F(2, 190) = 1519$, $p < .01$, with the highest ratings for the first cause (5.43, SD = 1.00), intermediate ratings for the double preventer (3.96, SD = 1.53), and low ratings for the first preventer (1.44, SD = 1.13). All pair-wise comparisons were significant (paired-samples *t*-tests, *p*s < .01).

Finally, the hypothesis of interest can be examined by considering causal ratings for the first cause and the double preventer as a function of functional status (see Fig. 7). A mixed ANOVA was conducted with the cause being evaluated as a within-subjects factor (first cause, double preventer) and several between-subjects factors: functional status (functional, incidental), domain (artifact, biological), and scenario version (light, heat). This analysis revealed a main effect of the cause being evaluated, mirroring the previous analysis, as well as the key interaction between the cause being evaluated and functional status, $F(1, 184) = 6.44$, $p < .05$: functional status increased ratings for the double preventer, but not for the first cause. Put differently, the existence of a transference relationship mattered more when the causes brought about their effects incidentally than when they did so to satisfy a function.

There was also a significant main effect of scenario version, $F(1, 184) = 6.99$, $p < .01$, as well as an interaction between the cause being evaluated and scenario version, $F(1, 184) = 18.1$, $p < .01$. Ratings were typically slightly higher for the light version than for the heat version for the double preventer, but not for the first cause.

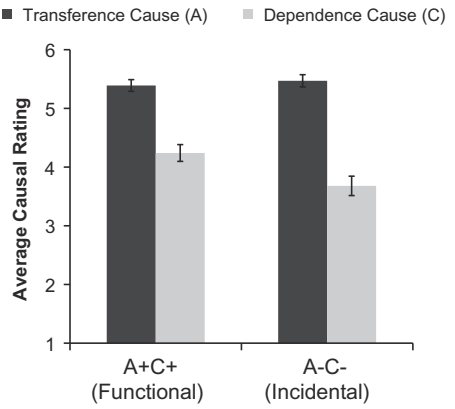

Fig. 7. Average causal ratings for the transference and dependence causes for the double prevention judgments in Experiment 2, broken down by functional status: when both brought about the effect as their function (A+C+) or both did so incidentally (A−C−). Error bars indicate one standard error of the mean in each direction.

5.2.2. Late preemption

Participants performed acceptably on the true/false comprehension task, with an average of 91.0% (SD = 8.0, range 45.8–100%) for the type true/false questions, and 76.6% (SD = 20.8, range 0–100%) for the token true/false questions. Performance on both tasks was well above chance (one-sample *t*-tests, *p*s < .01), and performance on the token true/false questions did not vary across conditions. However, performance on the type true/false questions did vary across conditions: there was a main effect of domain, $F(1, 184) = 6.79$, $p < .05$, an interaction between domain and scenario version, $F(1, 184) = 4.58$, $p < .05$, and an interaction between functional status and scenario version, $F(1184) = 4.92$, $p < .05$. While these effects were significant, they were quite small, with average performance ranging from 87.5% to 95.8% across conditions.

Collapsing across all between-subjects factors, there was a significant effect of the cause being rated, $F(2190) = 1518$, $p < .01$, with high ratings for the first cause (5.61, SD = .77), intermediate ratings for the second cause (3.91, SD = 1.71), and low ratings for the "control cause" (1.19, SD = .60). All pairwise differences were significant (paired-samples *t*-tests, *p*s < .01).

Finally, to examine the effect of functional status on causal ratings (see Fig. 8), the data were analyzed using a mixed ANOVA with the cause being rated (first cause, second cause) as a within-subjects factor, and several between-subjects factors: functional status (functional, incidental), domain (artifact, biological), and scenario version (light, heat). This analysis revealed a significant effect of the cause being rated, $F(1, 184) = 167$, $p < .01$, mirroring the previous analysis, with no other significant effects.

5.3. Discussion

The findings from Experiment 2 parallel those from Experiments 1a and 1c. The most notable difference from the double prevention findings is that an intention thoroughly eliminated the influence of transference in Experiment 1a, but a function merely moderated the influence of transference in Experiment 2. With regard to preemption, Experiment 2 mirrored Experiment 1c in finding uniformly high ratings for the first cause and intermediate ratings for the second cause, but Experiment 2 did not replicate the (unpredicted) finding from Experiment 1c, in which the intentional status of the second cause influenced causal ratings for the first cause.

These findings provide strong support for the hypothesis that explanatory mode influences causal judgments, with effects that range across human behavior, artifacts, and biological traits. In particular, the causal rating for a dependence cause that does not involve transference (a double preventer) is boosted when reasoning teleologically. This suggests that the criteria employed in causal ascription

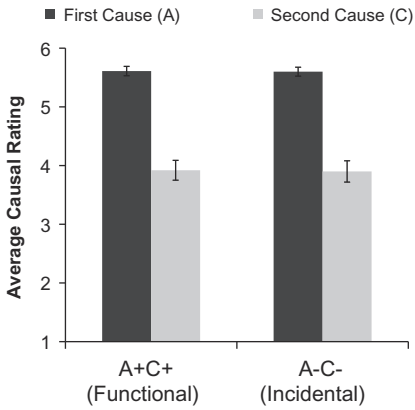

Fig. 8. Average causal ratings for the first and second causes for the preemption judgments in Experiment 2, broken down by functional status: when both brought about the effect as their function (A+C+) or both did so incidentally (A−C−). Error bars indicate one standard error of the mean in each direction.

weight transference more heavily in a mechanistic mode of reasoning than in a teleological mode of reasoning. The findings from Experiment 2 also help rule out the possibility that the effects in Experiments 1a–c were driven exclusively by mechanisms involved in the evaluation of moral responsibility.

6. General discussion

This paper began by motivating the hypothesis that different kinds of explanations have different consequences for cognition. In particular, teleological explanations may promote reasoning on the basis of functions, goals, and design, while mechanistic explanations promote attention to causal mechanisms. If this hypothesis is correct, different explanatory modes could generate judgments that appear more or less consistent with different concepts of causation. Teleological explanations, by virtue of being "mechanism-independent," should encourage a criterion for causation in terms of the dependence of the effect on the cause. In contrast, mechanistic explanations should encourage a criterion for causation that is more sensitive to aspects of the transmission from cause to effect.

In two sets of experiments, dependence and transference were isolated using causal structures common to discussions from philosophy: double prevention (Experiments 1a, 1b, 2) and preemption (Experiments 1c, 2). Within these structures, explanatory mode was manipulated by presenting participants with scenarios in which a candidate cause and effect supported a teleological explanation or did not. For Experiments 1a–c, which involved human behavior, this was accomplished by having the agents in the scenarios act intentionally or accidentally/deviantly. In Experiment 2, which involved artifacts and biological traits, this was accomplished by having the candidate causes bring about effects to fulfill a function, or do so incidentally.

The findings from double prevention support the prediction that teleological reasoning encourages a dependence criterion for causation. When an effect counterfactually depended on two agents who acted intentionally, participants provided very similar causal ratings, no matter that one agent (the double preventer) did not share a transference relationship with the effect. In contrast, when the agents acted accidentally, a majority of participants provided higher causal ratings for the cause that did involve transference, suggesting that transference was weighted more highly in causal ascriptions. The results from Experiment 2 with functions rather than intentions mirrored these findings: transference made a greater contribution to causal judgments when two causal factors brought about an effect incidentally than when they did so to fulfill a function. However, the effect was smaller for functions than for intentions, and transference continued to influence judgments when both candidate causes fulfilled a function.

The findings from preemption are more difficult to interpret, in part because it is difficult to isolate transference from dependence. Whether the candidate causes were agents (Experiment 1c) or artifacts and biological parts (Experiment 2), participants gave high causal ratings to a candidate cause that made contact with the effect, no matter that a "back-up" cause would have brought about the same outcome. Participants also gave uniformly low ratings to the "back-up" cause, which shared neither a transference nor a dependence relationship with the effect. These results help rule out alternative explanations for the double prevention findings, including the possibility that intentions and functions indiscriminately boost the causal ratings of marginal causes.

In the remainder of the general discussion, I consider the interpretation and implications of these findings. Section 6.1 contrasts the current hypothesis of causal–explanatory pluralism with existing accounts of causal ascription in psychology, including the mental model theory (Goldvarg & Johnson-laird, 2001), causal model theory (Sloman et al., 2009) and the force dynamics theory (Talmy, 1988; Wolff, 2007). Section 6.2 considers the functions of causal ascriptions, and introduces the possibility that the current findings can be accommodated within a more unified account of causation, which I refer to as an "exportable dependence" theory.

6.1. Competing theories of causal ascription

The introduction discussed two existing theories of causation within psychology that aim to characterize the conditions under which a relationship is described with the term "cause": the force dynamics theory (Talmy, 1988; Wolff, 2007) – a transference theory – and the causal model theory (Sloman et al., 2009) – a dependence theory. A third theory, which emphasizes logical rather than probabilistic dependence or transference relationships, is mental model theory (Goldvarg & Johnson-Laird, 2001). Mental model theory aims to characterize causation and other causal terms by appeal to the possibilities (mental models) a term licenses. For example, the claim "A causes B" is consistent with the possibility that A and B will both be present, A and B will both be absent, or B will be present and A absent. With additional assumptions motivated from the mental models framework (Johnson-Laird, 1983), the theory makes predictions not only about when a relationship will be described with the term "cause," but also how relationships will compose, e.g. what participants will say about the relationship between A and C given that A causes B and B allows C. For example, the theory predicts that if A prevents B and B prevents C, participants should claim that A prevents C. As noted in Section 1.2 some findings support this prediction (Goldvarg & Johnson-Laird, 2001), while others do not (Barbey & Wolff, 2007).

The findings from Experiments 1a and 2 that causal ratings for a double preventer are significantly higher than for a single preventer suggest that when A prevents B and B prevents C, participants are reasonably willing to claim that A causes C, and hence (presumably) reject the claim that A prevents C. Assuming that the present findings concerning individual, causal events mirror judgments for type-level causal claims, these results challenge the mental model theory, but match the predictions of the force dynamics and causal model theories. It's not clear how the mental model theory could be modified to accommodate the present findings, and in particular the influence of intentions on causal claims.

How might the force dynamics and causal model theories account for the influence of intentions and functions? One possibility is for force dynamics theory to claim that "psychological forces" (e.g. intentions) can be metaphorically treated as mechanisms for causal transmission (e.g. Wolff, 2007), and result in judgments that a causal relationship is more "direct" (Wolff, 2003). Therefore a double preventer should count as a cause when it acts intentionally, no matter that there isn't a mechanism of physical transmission between cause and effect. Force dynamics theory thus has a natural way to accommodate the finding that intentions boost the causal ratings for a double preventer. The theory might also be extended to cover functions (Experiment 2), so long as artifact and biological functions similarly involve a metaphorical force. However, the findings from preemption are potentially troubling for this account: the theory seems to predict that an agent who acts intentionally but is preempted from bringing about an effect should be considered a cause, or at minimum *more* like a cause than a similarly preempted agent who acts accidentally. In contrast to this prediction, participants in Experiments 1c and 2 gave uniformly low ratings to a preempted "cause," regardless of

intentional or functional status. With a more sophisticated treatment of metaphorical forces than the one I have offered, this obstacle could presumably be overcome (see also footnote 8).

The causal model theory does not predict an influence of intentions on causal ascriptions. However, given the theory's use of causal Bayes nets as a representational formalism (e.g. Pearl, 2000), causal model theory can potentially handle such cases by drawing on the extensive literature employing causal Bayes nets in theories of causal learning (e.g. Griffiths & Tenenbaum, 2005; Waldmann & Holyoak, 1992). In particular, one might expect causal ratings to be higher to the extent the relationship between a cause and effect is direct, in the sense that there are few intermediate variables (see also Lombrozo, 2007), and to the extent the cause reliably brings about the effect – that is, to the extent the conditional probability of the effect given the cause is high. In general, it may be that double prevention involves more intermediate variables than a transmitted force, and that intentional actions are more likely to bring about particular consequences than are accidental ones (but see Lagnado & Channon, 2008; McClure et al., 2007, for some evidence that the role of intentions is not reducible to differences in probability). If this is the case, then the causal model theory could plausibly be extended to account for the current data, although the current formulation (Sloman et al., 2009), designed to explain the difference between "cause" and "enable," fails to do so.

6.2. An exportable dependence theory

The findings from Experiments 1 and 2 support a causal–explanatory pluralism according to which both dependence and transference relationships influence causal ascriptions, but with different weights depending on whether the relationship in question is construed teleologically or mechanistically. A pluralistic position is attractive insofar as it explains the diversity of experimental findings, with some results supporting a dependence theory and others a transference theory. Moreover, the proposal is in line with recent evidence that teleological and mechanistic explanations have distinct consequences for reasoning (Lombrozo, 2009), and with recent calls for causal pluralism within philosophy (Godfrey-Smith, 2010; Hall, 2004; Hitchcock, 2003).

Despite these merits, it is worth considering whether the present findings support a more unified treatment. In particular, why might physical contact, intentions, and functions be the sorts of properties that boost causal ratings? To address this question, we should take a step back to consider the function of causal ascriptions. To be sure, causal ascriptions are likely to serve multiple functions: facilitating communication, assigning blame, and so on. But in many cases, representing causal structure is valuable insofar as it allows us to predict and control the world around us. Heider (1958) argued that causal attributions are essential precisely because they relate "offshoot events" to "underlying core-processes or core-structures," thereby helping us "to attain a stable environment" (Heider, 1958, p. 80). Put differently, causal ascriptions are valuable insofar as they identify relationships that are sufficiently stable or invariant across situations to be useful in prediction and intervention. This proposal mirrors a hypothesis in Lombrozo and Carey (2006) about the function of explanation, called "Explanation for Export." According to Explanation for Export, explanations identify factors that are "exportable" in the sense that they are likely to subserve future prediction and intervention. If the function of isolating parts of causal structure, be it in an explanation or a causal claim, is to subserve future prediction and intervention, then "good" causes should be those that reflect stable, invariant relationships that are exportable to relevant situations.

Considering the function of causal ascriptions provides some initial insight into the influence of physical contact and intentions: both may be aspects of causal structure that make for greater exportability. Intuitively, a process of direct transmission will bring about an effect more reliably than double prevention, and an agent is more likely to generate a particular outcome when intending to generate that outcome than when acting accidentally. In the language of a dependence theory, causal relationships involve a kind of dependence, but not all dependence relationships are equal. In particular, dependence relationships featuring direct, physical contact or intentions and functions may, all else being equal, be better.

The idea that not all dependence relationships are equal echoes discussions from philosophy, particularly in the recent work of Woodward (2006) and Campbell (2008). Woodward distinguishes between what he calls "sensitive" and "insensitive" causation. The key idea is that two causal

relationships that in fact obtain may differ in their sensitivity to contingent properties of the events in which they were embedded – that is, in whether they would still have obtained in counterfactually possible situations. A cause is relatively insensitive if it is robust to such counterfactual variation; it is sensitive if counterfactual changes would revoke its causal status. Recalling William James' comparison between Romeo and iron filing, this makes Romeo's intention to reach Juliet a relatively insensitive cause, as the dependence relationship between his intentions and reaching Juliet would have been maintained in a world with a wall. But if Romeo accidentally trips and stumbles into Juliet, his stumble would be a sensitive cause of reaching Juliet, as the world could easily have been different such that the stumble would not have resulted in so fortuitous a destination. In general, processes that are equifinal will be insensitive, because the same end is achieved despite variation in the means.

Campbell emphasizes other properties of dependence relationships that make some factors "better" causes than others. In particular, a cause will be better if it varies with the outcome in a one-to-one way (see also Woodward, 2010), and if parametric variations in the cause generate corresponding variations in the effect, as in a dose–response relationship. For example, Romeo's intentions satisfy these constraints in that an intention to Q corresponds in a (roughly) one-to-one way to Q-ing, and had Romeo intended Q' the outcome would likely have been Q'. But Romeo's accidental stumble does not share this relationship with reaching Juliet. The stumble could easily have generated different outcomes (e.g. falling into someone else), and had the stumble occurred differently (a bit sooner, to the right rather than the left), Romeo's physical actions would have been correspondingly different, but the outcome of reaching Juliet would not be (e.g. he would not have reached Juliet sooner, or reached her on his right rather than on his left).

Both Woodward's and Campbell's proposals suggest that a dependence relationship is not sufficient for a "good" causal relationship. Some dependence relationships will support strong causal ascriptions, while others that are nonetheless counterfactual-supporting may not. The proposals also share the idea that causal evaluations depend not only on the dependence relationships that obtained in the actual event being evaluated, but also on the dependence relationships that would have obtained under different conditions. Responsiveness to non-actual but possible alternatives makes sense if the goal is exportability. I thus refer to a dependence theory that builds in additional constraints about which dependence relationships are exportable as an "exportable dependence theory." If physical contact and intentions do in fact increase exportability, then exportable dependence provides a more unified account of the findings.

Does exportable dependence obviate the need for pluralism when it comes to causal ascriptions? Not necessarily. First, it could be that distinct dependence and transference concepts exist, no matter that both serve the common function of identifying exportable dependence relationships. That is, the ultimate explanation for dependence and transference concepts may appeal to exportable dependence, but the actual psychological processes involved in causal ascription may involve distinct criteria depending on whether a transference or dependence concept is invoked.

Alternatively, it could be that teleological and mechanistic modes of construal influence the *input* to processes of causal ascription, but do not influence the criteria employed in evaluating the quality of potential causal ascriptions. Nonetheless, the difference in input may result in judgments that appear more or less consistent with difference concepts of causation. An attractive version of this possibility is that teleological and mechanistic construals influence the variables over which the criteria for exportable dependence are evaluated. For example, when reasoning about intentional human behavior, it is natural to assess a criterion like Woodward's "sensitivity" by considering whether the relationship between the agent's intentions (a psychological variable) and a goal would be preserved in different situations. When reasoning about an accidental action, the relationship that one evaluates for sensitivity will be formulated in terms of physical variables, like a specific force or movement and the outcome. Similarly, one might assess Campbell's parametric variation for intentional actions by considering whether parametric changes in the agents' mental states (e.g. desiring silence, desiring the radio rather than music) result in parametric changes to the outcome, where the mental states are psychological variables. But when an agent acts accidentally, the relevant parametric variation may range over physical variables, such as the nature of the agent's physical actions (e.g. throwing faster, throwing harder). (See Lien & Cheng, 2000, for a relevant discussion of causal judgments involving variables at multiple levels.)

The view that explanatory construals generate distinct inputs to a common mechanism for causal ascription – one based on exportable dependence – has a number of advantages over the alternative proposals that have been considered in the course of this paper. First, the proposal can explain the finding that intentions have a greater impact on the causal ratings of double preventers than do functions. An intentional agent is a fairly "insensitive" cause of an intended outcome, as the agent can modify behavior in response to environmental circumstances to bring about that outcome. A functional artifact part or biological trait should be less sensitive than an incidental part or trait, but because such parts or traits must be "designed" by people or natural selection in advance of the specific causal event being evaluated, they cannot respond to unforeseen environmental variation to bring about the original end, and are thus more sensitive than intended human actions.

Second, the notion of exportable dependence helps explain why the "back-up" causes in the late preemption scenarios were treated as marginal causes, receiving ratings reliably higher than the "control" causes. While the outcomes did not depend on the back-up causes in the event that in fact unfolded, the outcome *would* have depended on the back-up causes in very similar, counterfactually possible worlds. Thus the back-up causes had an intermediate degree of exportability: less than the cause that proved effective, but more than the control cause.

Finally, the idea of exportable dependence fits nicely with the motivation for explanatory pluralism. Different modes of explanation are useful because they capture different but real regularities in the environment. Reasoning teleologically – in terms of functions and design – makes it possible to capture relationships between behavior and outcomes that would be very difficult to express in terms of purely physical variables. Likewise, reasoning mechanistically – in terms of causal processes – makes it possible to capture relationships that would be very difficult to express in terms of intentional or functional variables. The fact that each mode of explanation employs variables that support particular generalizations suggests that variable selection is itself subject to a criterion of exportability. It is therefore unsurprising that explanatory mode should have consequences for the judged exportability of particular relationships, and hence for causal ascriptions. Moreover, one should expect reasoners to spontaneously adopt the mode of explanation that supports greater exportability, such as a teleological mode when assessing an equifinal system.

6.3. Summary and conclusions

The current studies demonstrate that causal ascriptions are sensitive to a variety of factors, including whether a relationship is intentional or functional, and whether a mechanism involves direct transmission. I began by arguing for a causal–explanatory pluralism according to which different modes of explanation prompt different criteria for causal ascription, with a teleological mode decreasing sensitivity to properties of the physical process relative to a mechanistic mode, and thereby generating judgments more consistent with a dependence theory than with a transference theory. This prediction was generated from the observation that teleological systems exhibit equifinality – the tendency to obtain the same end despite variation in means – and therefore involve a kind of "mechanism-independence." The findings from Experiments 1a and b bore out these predictions, as did – to a more limited extent – those from Experiment 2.

To provide a more complete account of the findings, the general discussion introduced the idea of "exportable dependence" – the hypothesis that causal ascriptions are evaluated with respect to a dependence criterion, but one that evaluates the "exportability" of the dependence relationship to relevant but non-actual situations. Explanatory modes may reflect two different ways to parse the environment into variables that support exportable dependence relationships: one in terms of intentions, functions, and goals; the other in terms of physical processes. Even if causal judgments are ultimately made with respect to a shared criterion of exportable dependence, causal judgments can thus be expected to vary as a function of explanatory construal.

These findings and arguments contribute to a growing body of work supporting causal–explanatory pluralism. The distinction between teleological and mechanistic explanations appears to be psychologically real and cognitively deep, with consequences for how we reason about the very cement of the universe (Mackie, 1980): causation.

Acknowledgments

Sincere thanks to many people for relevant discussions, including John Campbell, Winston Chang, Fiery Cushman, Alison Gopnik, Tom Griffiths, Ned Hall, Chris Hitchcock, Joshua Knobe, Laurie Paul, Steven Sloman, Michael Strevens, Phil Wolff, and James Woodward, as well as the members of the Causation Collaborative funded by the McDonnell Foundation, the participants in the 2009 NEH workshop on Experimental Philosophy, and the members of the Concepts and Cognition lab at UC Berkeley. John Campbell, Winston Chang, Chris Hitchcock, Joshua Knobe, Laurie Paul, Steven Sloman, Phil Wolff, and James Woodward additionally provided helpful feedback on an earlier manuscript. Cleo Barrable, Christina Botros, Brian Christian, and Rosemary Jammal provided valuable input and helped tremendously with data collection. This work was supported by NSF Grant BCS-0819231.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? *Cognition, 69*, 135–178.
Ahn, W., & Kalish, C. (2000). The role of mechanism beliefs in causal reasoning. In R. Wilson & F. Keil (Eds.), *Cognition and explanation* (pp. 199–226). Cambridge, MA: MIT Press.
Ahn, W., & Kim, N. S. (2000). The causal status effect in categorization: An overview. In D. L. Medin (Ed.). *Psychology of learning and motivation* (Vol. 40, pp. 23–65). New York: Academic Press.
Barbey, A. K., & Wolff, P. (in preparation). Composing causal relations in force dynamics.
Barbey, A. K., & Wolff, P. (2007). Learning causal structure from reasoning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 713–718). Mahwah, NJ: Erlbaum.
Buehner, M. J., & Cheng, P. W. (2005). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Handbook of thinking & reasoning* (pp. 143–168). New York, NY: Cambridge University press.
Bullock, M. (1985). Causal reasoning and developmental change over the preschool years. *Human Development, 28*, 169–191.
Campbell, J. (2008). Interventionism, control variables and causation in the qualitative world. *Philosophical Issues, 18*, 424–443.
Casler, K., & Kelemen, D. (2008). Developmental continuity in the teleo-functional explanation: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development, 9*, 340–362.
Chang, W. (2009). Connecting counterfactual and physical causation. In *Proceedings of the 31th annual conference of the cognitive science society* (pp. 1983–1987). Austin, TX: Cognitive Science Society.
Cushman, F. A. (2008). Crime and Punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353–380.
Danks, D. (2005). The supposed competition between theories of human causal inference. *Philosophical Psychology, 18*(2), 259–272.
Dennett, D. (1971). Intentional systems. *Journal of Philosophy, 68*, 87–106.
Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on the creator's intent in extending names for artifacts. *Psychological Science, 14*, 164–168.
Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science, 59*, 195–216.
Dowe, P. (2008). Causal processes. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2008 Edition)*. <http://plato.stanford.edu/archives/fall2008/entries/causation-process/>.
Falcon, A. (2008). Aristotle on causality. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2008 Edition)*. <http://plato.stanford.edu/archives/fall2008/entries/aristotle-causality/>.
Fincham, F. D., & Jaspars, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 13). New York, NY: Academic Press.
Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76*, 91–103.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*, 287–292.
Godfrey-Smith, P. (2010). Causal pluralism. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.) *Oxford handbook of causation* (pp. 326–337).
Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25*, 565–610.
Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning: Psychology, philosophy, computation*. New York: Oxford University Press.
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*, 1–30.
Greif, M., Kemler-Nelson, D., Keil, F. C., & Guiterrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17*, 455–459.
Griffiths, T. L. (2005). *Causes, coincidences, and theories*. Stanford, CA: Unpublished doctoral dissertation, Stanford University.
Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354–384.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge, MA: MIT Press.
Halpern, J. Y., & Pearl, J. (2001). Causes and explanations: A structural-model approach – Part I: Causes. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 194–202). San Francisco, CA: Morgan Kaufmann.
Hart, H. L. A., & Honoré, T. (1985). *Causation in the law* (2nd ed.). New York, NY: Oxford University Press.

Heider, F. (1958). *The psychology of interpersonal relations*. Lawrence Erlbaum Associates.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy, 98*, 273–299.
Hitchcock, C. (2003). Of humean bondage. *British Journal for the Philosophy of Science, 54*, 1–25.
Hitchcock, C. (2008). Probabilistic causation. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2008 Edition)*. <http://plato.stanford.edu/archives/fall2008/entries/causation-probabilistic/>.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 11*, 587–612.
James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.
Johnson-Laird, P. (1983). *Mental models*. Cambridge, MA: Harvard University press.
Keil, F. C. (1992). The origins of an autonomous biology. In *Modularity and constraints in language and cognition*. In M. R Gunnar & M. Maratsos (Eds.). *Minnesota symposium on child psychology* (Vol. 25, pp. 103–138). Hillsdale, NJ: Earlbaum.
Keil, F. C. (1994). The birth and nurturance concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain Specificity in cognition and culture* (pp. 234–254). Cambridge, England: Cambridge University Press.
Keil, F. C. (1995). The growth of causal understanding of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multi-disciplinary debate* (pp. 234–262). Oxford, England: Clarendon Press.
Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254.
Kelemen, D. (1999a). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Science, 3*(12), 461–468.
Kelemen, D. (1999b). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology, 35*, 1440–1452.
Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*, 138–143.
Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 441–448). Cambridge, MA: MIT Press.
Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The influence of intentionality and foreseeability. *Cognition, 108*, 754–770.
Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.
Lewis, D. (2000). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. Cambridge: MIT Press.
Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology, 40*, 87–137.
Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464–470.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232–257.
Lombrozo, T. (2009). Explanation and categorization: How "why?" informs "what?". *Cognition, 110*, 248–253.
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.
Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science, 18*, 999–1006.
Mackie, J. L. (1980). *The cement of the universe: A study of causation*. New York, NY: Oxford University Press.
Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
Mameli, M., & Bateson, P. (2006). Innateness and the sciences. *Biology and Philosophy, 21*(2), 155–188.
Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General, 132*, 419–434.
Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition, 78*, 1–26.
McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology, 37*, 879–901.
Menzies, P. (2003). Counterfactual theories of causation. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2009 Edition)*. <http://plato.stanford.edu/archives/fall2009/entries/causation-counterfactual/>.
Menzies, P. (2008). Counterfactual theories of causation. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Winter 2008 Edition)*. <http://plato.stanford.edu/archives/win2008/entries/causation-counterfactual/>.
Newsome, G. L. (2003). The debate between current versions of covariation and mechanism approaches to causal inference. *Philosophical Psychology, 16*, 87–107.
Paul, L. (1998). Problems with late preemption. *Analysis, 58*, 48–53.
Peacocke, C. A. B. (1979). *Holistic explanation: Action, space, interpretation*. New York, NY: Clarendon Press/Oxford University Press.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
Schlottman, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and knowledge of underlying mechanisms. *Developmental Psychology, 35*, 303–317.
Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, 1–51.
Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University press.
Sloman, S. A., Barbey, A. K., & Hotaling, J. (2009). A causal model theory of the meaning of "cause," "enable," and "prevent." *Cognitive Science, 33*, 21–50.
Spellman, B. A., & Kincannon, A. (2001). The relation between counterfactual ("but for") and causal reasoning: Experimental findings and implications for jurors' decisions. *Law and Contemporary Problems: Causation in Law and Science, 64*, 241–264.
Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100.
Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222–236.
Walsh, C. R., & Sloman, S. A. (2005). The meaning of cause and prevent: The role of causal mechanism. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 2331–2336). Mahwah, NJ: Lawrence Erlbaum Associates.

White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin, 108*, 3–18.
White, P. A. (1995). *Understanding of causation and the production if action: From infancy to adulthood*. Hillsdale, NJ: Erlbaum.
Wilson, G. (2009). Action. In E.N. Zalta (Ed.) *The stanford encyclopedia of philosophy (Fall 2009 Edition)*. <http://plato.stanford.edu/archives/fall2009/entries/action/>.
Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition, 88*, 1–48.
Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111.
Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General*.
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.
Woodward, J. (2008). Causation and manipulability. In Edward N. Zalta (Ed.) *The stanford encyclopedia of philosophy (Winter 2008 Edition)*. <http://plato.stanford.edu/archives/win2008/entries/causation-mani/>.
Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy, 25*, 287–318.