2

Theories of causation should inform linguistic theory and vice versa

BRIDGET COPLEY AND PHILLIP WOLFF

2.1 Introduction

Linguistics has long recognized that causation plays an important role in meaning. Over the last few decades of the generative linguistic project, it has become clear that much of phrase structure is arranged around causal relationships between events (or event-like entities such as situations). Reference to causation in this tradition has most often taken the form of a relation CAUSE, with little further elucidation, in effect treating CAUSE as a primitive. This treatment of causation as a primitive relation has proved adequate to the task of developing grammatical structures that make reference to causation. But arguably, this hands-off approach to the meaning of causation has obscured potentially relevant details, impeding linguists' ability to consider hypotheses that might yield a more comprehensive analysis of the roles played by concepts of causation in language. Unpacking the notion of causation should, on this view, afford a deeper understanding of a range of linguistic phenomena, as well as their underpinnings in conceptual structure.*

In this chapter, we show how attention to the variety of existing theories of causation could advance the understanding of certain linguistic phenomena. In the first section, we review the two major categories of theories of causation, including some of the principal challenges that have been raised for and against each category. We identify in the second section a range of linguistic phenomena that we feel would benefit from a deeper investigation into causation—defeasibility, agentivity and related concepts, and causal chains—and also speculate on how theories of causation might inform our understanding of these phenomena. Since the linguistic theories make testable claims about cognition, they give rise to potential connections between syntactic structure and cognition. In the concluding section, we express our hope that further investigations along these lines

* Thanks to Kevin Kretsch and Jason Shepard for helpful discussion.

Causation in Grammatical Structures. First edition. Bridget Copley and Fabienne Martin (eds).

may pave the way for a theory of meaning grounded in both syntactic and cognitive realities, in a way that has not previously been possible.

2.2 Theories of causation

Given the importance of causation in linguistic theory, the question naturally arises as to whether some of the varied insights about causation from philosophy and cognitive psychology might have consequences for our understanding of linguistic theory. Up until now, they have not. The single theory of causation most often referred to in linguistic articles—by an extremely large margin—is Lewis's (1973) counterfactual theory of causation, which is discussed and adapted in Dowty's influential (1979) book. But even when Lewis or Dowty are cited, the causal relation is usually[1] treated by linguists essentially as a primitive. As a consequence, even Lewis's theory has not had a particularly meaningful impact on our understanding of the role of causal concepts in forming causal meanings.[2] Certainly, when causation has been treated as a causal primitive, it has just been a placeholder, a way of not having to deal with what causation is. Historically, this move was defensible since it was not clear that linguistic phenomena really depended on how causation was defined, or whether the grammar had access to anything more fine-grained than a primitive relation CAUSE. Arguably, it was even provisionally necessary to treat causation as an unanalyzed primitive at the outset of the development of the syntax–semantics interface, to avoid unnecessary complication.

As the generative enterprise has progressed, however, the need to address the lacunae still present in linguistic phenomena related to causation has become more and more pressing, both in familiar and in novel data. A number of linguistic phenomena, some of which we will present in this chapter, are not well addressed by appeal to a primitive CAUSE. It has therefore become increasingly apparent that

[1] Notable exceptions—i.e. authors who further investigate the Lewis–Dowty approach—include: Bittner (1998) (type lifting for cases where causal meaning is morphologically unmarked); Eckardt (2000) (focus sensitivity of the verb *cause*); Kratzer (2005) (a causal head in resultatives); Neeleman and van de Koot (2012) (questioning whether there is a causal event associated with causative predicates; see also Van de Velde 2001 for a similar point); Truswell (2011) (constraints that causal structure puts on extraction). A related theory, that of causal modeling (see section 2.2.1.4) is starting to be of interest to people working on modals and counterfactuals; e.g. Dehghani et al. (2012). Production theories of causation that rely on forces, or transmission of energy are very similar to a parallel development in cognitive linguistics that had its start with Talmy (1985a; 1985b; 1988). A few lines of inquiry in formal linguistics have explicit, or implicit links to production theories, most notably those of van Lambalgen and Hamm (e.g. 2003; 2005) and Zwarts (e.g. 2010).

[2] Lewis's theory has had an enormously meaningful and fruitful impact on semantics in the realm of conditionals (counterfactual and otherwise) and modals, stemming from initial work by Stalnaker (1968 and much later work), and Kratzer (1977; 1979 and much later work), as well as Dowty's work on the progressive (1977; 1979). The clear predictions and expressive power of the possible worlds approach have deservedly made it a jewel in the crown of modern semantic theory. However, this body of research has not generally been explicitly linked to the issue of causation. As we will discuss in section 2.2.2, causation and at least one kind of modal notion (that of volitionality) are related to each other, whether or not one agrees with Lewis on the best way to represent them; see also Ilić, Ch. 7, this volume, for discussion of linguistic data bearing on the relationship between causality and modality.

the story one tells about causal meanings will have to depend on one's theory of causal concepts. If, for instance, we were to take the details of Lewis' counterfactual theory of causation into account, it could have some interesting consequences.

Lewis's theory, though, is not the only theory in town. Other modern theorists have suggested that causal relations might be based on statistical dependencies (Suppes 1970; Eells 1991; Cheng and Novick 1991), manipulation (Pearl 2000; Woodward 2003), necessity and sufficiency (Mill 1973; Mackie 1974; Taylor 1966), transfer of conserved quantities (Dowe 2000), force relations (Fales 1990; White 2006), energy flow (Fair 1979), causal powers (Mumford and Anjum 2011a), and property transference (Kistler 2006a). While this list accurately reflects the considerable variation among philosophers as to the nature of causation, discussions of causation often categorize theories of causation according to several dimensions, such as whether the relata are single or generic, individual or population level; whether the causal relation is physical or mental; whether the causal relation is objective or subjective; or whether it is actual or potential (see Williamson 2009, e.g.). Many of these distinctions are not particularly relevant to current linguistic understanding of the causal relation as an element that occurs in a wide range of different environments. For example, linguistic consensus treats genericity as a separate operator from the causal relation CAUSE, so any viable proposal for the latter must be consistent with both generic and individual causation. In the following categorization of causation we emphasize two broad categories: *dependency* theories, in which A causes B if and only if B depends on A in some sense, and *production* theories (also commonly referred to as *process* theories), in which A causes B if and only if a certain physical transmission or configuration of influences holds among the participants in A and B.

2.2.1 Dependency theories

One major category of theories holds that causation is understood as a *dependency*. There are three main classes of dependency theory.

2.2.1.1 Logical dependency There is intuitive appeal in defining causation in terms of necessary and/or sufficient conditions. However, an analysis of such accounts raises a range of problems that are generally considered insurmountable (see Scriven 1971; also Hulswit 2002; Sosa and Tooley 1993). Consider, for example, a definition that identifies the concept of causation as a condition that, in the circumstances, is necessary. With such a definition, we might agree with Hume's comment that one event causes another "where, if the first object had not been, the second never had existed" (Hume 2007[1748])—i.e. a cause is a factor without which the effect would not have occurred.[3] The simplest version of such an account is contradicted by cases of late

[3] "Necessary" here is not necessarily to be thought of in the later modal logic sense of quantification over possible worlds; see e.g. Hume's "necessary connection" between cause and effect wherein "the determination of the mind, to pass from the idea of an object to that of its usual attendant" (Kistler 2006a).

pre-emption, i.e. cases where a potential alternative cause is interrupted by the occurrence of the effect.[4] For an example of late pre-emption, consider a scenario developed by Hall (2004).

There is a bottle on the wall. Billy and Suzy are standing close by with stones and each one throws a stone at the bottle. Their throws are perfectly on target. Suzy happens to throw first and hers reaches the bottle before Billy's. The bottle breaks. In this scenario, the effect of a particular candidate cause, Billy's throw, is "pre-empted" by another cause, Suzy's throw. As empirically verified by Walsh and Sloman (2005), Suzy's throw is understood to be the cause of the bottle's breaking, but Suzy's throw was not a *necessary* condition for the effect: if Suzy had missed, the bottle still would have broken because of Billy's throw.

An alternative account of causation in terms of logical dependency would be the proposal that causation is a sufficient condition for an effect. Under this view, a factor is the cause of an effect if the presence of that factor guarantees the occurrence of an effect. Of course, one problem with this view is that it is rare to find a case where single condition is sufficient in and of itself. An event is rarely, if ever brought about by a single factor; as Mill (1973[1872]) notes, every causal situation involves a set of conditions, which are sufficient for an effect when combined. Another problem for a sufficiency view is the case of late pre-emption described above. As noted, we would not say that Billy caused the breaking of the bottle. This is surprising from a sufficiency view, since Billy's throw is a sufficient condition for the breaking of the bottle.

Yet another possibility would be to define a cause as a necessary *and* sufficient condition (Taylor 1966). Such a definition fails because it entails that the cause would be a necessary condition and, as already discussed, there can be causes that are not necessary. A related view of causation is Mackie's (1965) INUS condition, that says a cause is an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result. The INUS condition, ultimately, defines causation in terms of sufficiency, but as discussed above, a factor (or set of factors) can be sufficient and yet not be a cause. A modern instantiation of an account of causation based on logical necessity and sufficiency can be found in Goldvarg and Johnson-Laird's (2001) model theory.

2.2.1.2 Counterfactual dependency Another type of dependency theory is based on the idea of counterfactual dependency: the counterfactual proposition that E would not have occurred without C. As we have seen, counterfactual dependency can be thought of as a paraphrase of the proposition that C is necessary for E. The modus operandi behind counterfactual theories of causation is thus to link two groups of

[4] Late pre-emption occurs when there are two potential causes but the occurrence of the effect prevents one of the causes from causing the effect. Early pre-emption (to be discussed in section 2.2.1.2) occurs when the initiation of one cause prevents the other potential cause from happening at all. See Menzies (2008) and Paul (2009) for more details.

intuitions: intuitions about whether certain counterfactual propositions are true and intuitions about whether certain events cause other events.

The simplest way to link these intuitions would be to identify causation with counterfactual dependency: i.e. to say that C is a cause of E if and only if E would not have occurred if C had not occurred. This looks as though we are equating causation with logical necessity, because it asserts that C must be present in order for E to occur. As we have seen, a definition of causation in terms of logical necessity erroneously predicts that C is not a cause of E if E could have been caused by something other than C. David Lewis, in the original version of his influential counterfactual theory of causation (1973 et seq.), proposed to avoid this problem by weakening the biconditional ("if and only if") to a mere conditional: counterfactual dependency entails causation, but causation does not entail counterfactual dependency. According to Lewis (1973), the reason that causation does not entail counterfactual dependency is because causal relations can sometimes emerge from transitive reasoning, but counterfactual relations, arguably, are not transitive (see Stalnaker 1968), and so causal relations may sometimes exist in the absence of a counterfactual dependency.

An example of such a scenario occurs in cases of so-called *early* pre-emption. Imagine, for example, a slightly different version of the Billy and Suzy scenario that was discussed above (which demonstrated *late* pre-emption). In this new scenario, Suzy throws a rock at a bottle (breaking it) and Billy acts as a backup thrower just in case Suzy fails to throw her rock. Here Suzy is the cause of the bottle's breaking, but just as in the case of late pre-emption there does not exist a counterfactual dependency between Suzy and the bottle's breaking; if Suzy had not thrown, the bottle would have still been broken because Billy would have thrown his rock.

To insulate his theory against such scenarios, Lewis (1973) proposes that C causes E if and only if *stepwise* counterfactual dependency holds between C and E, i.e. only if there are counterfactual dependencies holding between adjacent events in the chain, but not necessarily non-adjacent events in the chain. In the early pre-emption scenario, Lewis (1973) would argue that while the rock's breaking does not depend counterfactually on Suzy, there is a counterfactual dependency between Suzy and the intermediate event of the rock flying through the air, and a counterfactual dependency between the rock flying through the air and the bottle's breaking, and this chain of counterfactual dependencies licenses a judgment that Suzy's throw caused the bottle to break.[5] Lewis' approach to the problem raised by early pre-emption ultimately led to a definition of causation in terms of causal chains: specifically, C is a cause of E if and only if there exists a causal chain leading from C to E. Importantly, however, the links in the causal chain are defined in terms of counterfactual dependencies.

[5] Contra Lewis (1973), it is not entirely clear that there exists a counterfactual dependency between Suzy's throw and a rock flying through the air. Had Suzy not thrown her rock, there still would have been a rock flying through the air due to Billy.

There are a number of problems with Lewis's initial proposal, some of which continue to complicate counterfactual theories today. One kind of problem occurs in the case of *late* pre-emption. In both early and late pre-emption, a counterfactual dependency fails to hold between C and E, suggesting that counterfactual dependency is not *necessary* for causation. Lewis (1973) was able to address the lack-of-necessity problem in cases of early pre-emption by defining causation in terms of *stepwise* counterfactual dependency; but this fix only works for early pre-emption, not for late pre-emption, so the lack-of-necessity problem remains in the case of late pre-emption. Two other problems can be illustrated with a single type of scenario (see Hall 2000). Consider a case where an assassin places a bomb under your desk, causing you to find it, which causes you to remove it, which causes your continued survival. Without the assassin putting the bomb under your desk, you would not have removed it and thereby ensured your survival. Cases such as this demonstrate that counterfactual dependencies are not *sufficient* for causation. In this example, there exists a counterfactual dependency between the assassin and survival, but we would not want to say that the assassin caused your continued survival. Such cases also raise a problem for Lewis' (1973) definition of causation in terms of causal chains. As already noted, this definition was motivated by the assumption that causation is transitive; but, as shown in this example, there may be cases where transitivity in causation fails (see also McDermott 1995; Ehring 1987).

Lewis's 2000 theory attempts to address several of the problems facing his 1973 theory. In Lewis's new theory, counterfactual dependency exists when alterations in the cause lead to alterations in the effect. So, for example, if Suzy's throw is slightly altered—she throws the rock a bit faster, or sooner, or uses a lighter rock—the resulting breaking of the bottle will also be slightly altered. Lewis's new theory is able to explain why Suzy's throw, and not Billy's throw, is considered to be the cause: alterations to Suzy's throw result in changes in the effect, while alterations to Bill's throw do not. However, there is reason to believe that Lewis's new theory still does not escape the challenge raised by late pre-emption. As noted by Menzies (2008), in the case of Billy and Suzy, there is a degree to which alterations in Billy's throw could result in alterations of the final effect—if, for example, Billy had thrown his rock earlier than Suzy's. In order for Lewis's theory to work, only certain kinds of alteration may be considered. To foreshadow a point we will later make in the discussion of production theories (section 2.2.2), it may be that Lewis's theory can be made viable if the alterations are confined to those that are relevant to the creation of forces.

2.2.1.3 Probabilistic dependency According to Hume (2007[1748]), if it is true that an event C causes an event E, it is true that events similar to C are invariably followed by events similar to E. This view is referred to as the "regularity theory" of causation. A well-known difficulty with the regularity theory is the simple observation that causes are *not* invariably followed by their effects. The observation has motivated

accounts of causation that ground the notion of causation in terms of probabilistic dependency.[6]

The simplest type of probabilistic dependency is one that relates causation to probability raising (Reichenbach 1956; Suppes 1970; Eells 1991). A variable C raises the probability of a variable E if the probability of E given C is greater than the probability of E in the absence of C (formally, $P(E \mid C) > P(E \mid \neg C)$). Thus on this theory, if smoking causes cancer, the probability of cancer given smoking is greater than the probability of getting cancer in the absence of smoking. An alternative way of describing the relationship between the conditional probabilities $P(E \mid C)$ and $P(E \mid \neg C)$ is to say that C is a cause of E when C makes a difference in the probability of E. Indeed, whenever $P(E \mid C) > P(E \mid \neg C)$ holds, E and C will be positively correlated and whenever E and C are positively correlated, $P(E \mid C) > P(E \mid \neg C)$. A relatively recent instantiation of probability raising is instantiated in Cheng and Novick's (1992) probabilistic contrast model.

While probabilistic approaches to causation address important limitations not addressed by other dependency accounts, they do not escape some other problems. Probability raising on its own seems to be not *sufficient* for causation: that is, C might raise the probability of E without C's being a cause of E. The reason it is not sufficient is because the presence of one event might (appear to) make a difference in the probability of another, but that appearance might in fact be due to a shared common cause, rather than from one causing the other (Hitchcock 2010). So, for instance, seeing a spoon raises the probability of seeing a fork; not because spoons cause forks, but rather because there is some overlap between the causes of seeing a spoon and the causes of seeing a fork. Reichenbach suggested that such cases could be flagged in the following manner: if two variables are probabilistically dependent and if one does not cause the other, they have a common cause that, if taken into account, renders the two variables probabilistically independent. Williamson (2009), however, points out that Reichenbach's characterization excludes cases of probabilistic dependency where C and E are related logically, mathematically, through semantic entailment, or accidentally.

Even with Reichenbach's common-cause cases excluded, however, sufficiency is still a problem. Returning to Suzy and Billy's case of late pre-emption, we can also see that probabilistic dependency is not sufficient for causation (C raises the possibility of E but C is not a cause for E). We can imagine that Billy's throw hits the bottle with a certain probability while Suzy's throw hits it with a certain, possibly different, probability. They both throw, and Suzy's stone hits the bottle, and breaks it. In that case we would say

[6] Unlike in other dependency theories discussed above, in probabilistic dependency theories there can be a causal relation (between kinds) even when the effect does not occur (at the individual level), since all that is needed to calculate a causal relation is the probability of the effect's occurring under certain conditions. As we will see in section 2.3.1, this property could be useful in understanding cases of defeasible causation in language, such as non-culmination of accomplishments.

Suzy's throw was the cause of the bottle breaking—and indeed her throw raised the probability of the bottle's breaking. However, Billy's throw also raised the probability of the bottle's breaking, although his throw was not the cause (Hitchcock 2010).

Additionally, probability raising is apparently not *necessary* for a factor to be considered a cause; cases exist where C is a cause of E but C does not raise the probability of E. Imagine that Suzy throws her rock with a 25% chance of shattering of the bottle. If Suzy had not thrown the rock, Billy would have done so, with a 70% chance of shattering. In this example, Suzy's throw would be the cause of the shattering, even though it lowered the chance of that effect (from 70% to 25%) (Hitchcock 2010).[7]

As usual, problems such as these are probably not insurmountable, but any viable solution would be expected to bring complications to the theory. Probabilistic dependency theorists have addressed such problems by getting more specific about the background contexts on which probabilities are calculated (Cartwright 1979; Skyrms 1980), as well as by recognizing differences between singular and general (kind) causation (Eells 1991; Hitchcock 2004), since probabilities can arguably only be calculated for kinds of events, not for individual events.

2.2.1.4 Causal modeling approaches to causation One particular formal implementation of the dependency view of causation has had a wide-ranging influence on a number of fields. As Williamson (2009) points out, the formalism of Bayesian networks developed in the 1980s (Pearl 1988; Neapolitan 1990) provided an efficient way to think about causal connections at a time when causal explanations were out of fashion in scientific fields, in part due to Russell's (1913) attack on the notion of causation as being unnecessary for scientific explanation.

A causal Bayesian network represents the causal structure of a domain and its underlying probability distribution. The causal structure of the domain is represented by a directed acyclic graph of nodes and arrows, whereas the probability distribution consists of the conditional and unconditional probabilities associated with each node. The alignment of these two kinds of information allows us to make predictions about causal relationships using probability theory. A simple causal Bayesian network is shown in Fig. 2.1. Each node in the network is associated with an unconditional, prior probability. For example, in the the network shown in

[7] Another example of how probability raising is not necessary for causation is seen in cases where the influence of one cause is overwhelmed by the influence of another (Cartwright 1979; Hitchcock 2010). For example, under the right circumstances, the probability of cancer might be less in the presence of smoking than the probability of cancer in the absence of smoking, that is P(cancer | smoking) < P(cancer | not smoking). Clearly, smoking causes cancer, but a positive correlation between cancer and smoking might be masked, or even reversed in the presence of another cause. Imagine a situation in which not smoking is correlated with living in a city, breathing highly carcinogenic air. In such a situation, not smoking could be more strongly associated with cancer than smoking, but the causal relationship between smoking and cancer could remain. Such reversals are widely known as examples of Simpson's paradox; see Kistler (Ch. 4, this volume) for additional discussion.

Fig. 2.1, exercise is associated with a 0.5 probability of being true and a 0.5 probability of being false, while debt is associated with a 0.2 probability of being true and a 0.8 probability of being false. The arrows in this graph represent causal relations (in the broad sense). In Fig. 2.1, the arrows from exercise and debt to happiness convey that these two variables affect happiness. The exact way in which they do so is described in the probability table associated with happiness, which specifies several conditional probabilities: for example, the probability of happiness being present when one exercises but also has debt, i.e. P(Happinesss | Exercise and Debt), is 0.6, and the probability of not being happy when one exercises and has debt, i.e. P(∼Happinesss | Exercise and Debt), is 0.4. The conditional probabilities specified in the probability table specify that exercise raises the probability of happiness, whereas debt lowers the probability of happiness. It is in this manner that a causal Bayesian network can represent both facilitative and inhibitory causal relations, and it is for this reason that the arrows are causal in a broader sense than is encoded in the meaning of the verb *cause*. Roughly, the arrows mean something like *influence* or *affect*.

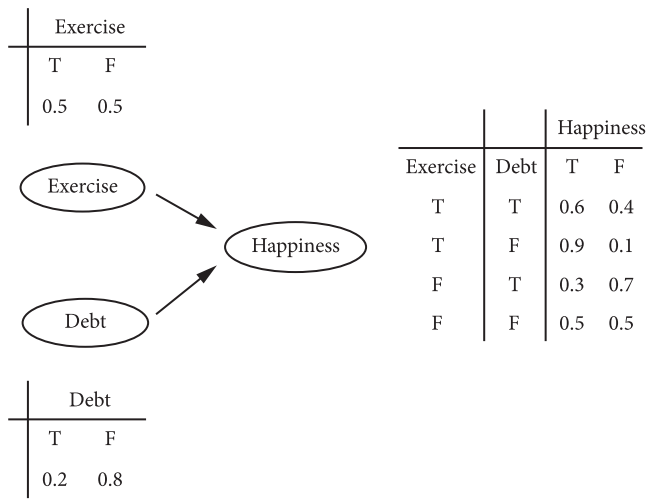

Fig. 2.1 Causal Bayesian network with associated probability tables

Causal Bayesian networks allow us to reason about causation in more than one dimension, i.e. in networks rather than mere chains. For example, they allow us to predict—using Bayes' rule—the probability of certain variables being true when other variables are either true or false, both in the direction of causation and diagnostically, i.e. working from effects to causes. However, in order to understand where the causal arrows themselves come from—i.e. when we are justified in asserting a causal relation between two variables—more must be said.

In order for a Bayesian network to qualify as a causal Bayesian network, it has been argued (Hausman and Woodward 1999) that its probabilistic dependencies and arrows must honor the *Causal Markov Condition*.[8] The Causal Markov Condition holds that a variable C will be independent of every variable in a network except its effects (i.e. descendants) (e.g. E), conditional on its parents. Hausman and Woodward (1999) use the Causal Markov Condition as part of a sufficiency condition on causation: if C and E are probabilistically dependent, conditional upon the set of all the direct parents of C in the given Causal Markov Condition-satisifying model, then C causes E. However, late pre-emption provides a counterexample to this sufficiency condition (i.e. a case where the condition holds but the intuition is that C does not cause E).

Another way to characterize causation in a causal Bayesian network is in terms of the notion of *intervention* (Pearl 2000; Woodward 2009). An intervention is a process by which a variable in a network is set to a particular value. The notion of intervention is closely related to our sense of causation. In effect, interventions allow us to conduct counterfactual reasoning. If C causes E, then intervening on, or "wiggling" the value of C should result in corresponding changes in the value of E. If we can intervene to counterfactually change the value of C to any possible value, and can still predict the probability of the value of E being true, we can be confident that C causes E. For example, suppose that you want to find out if a switch being in an up position causes a light to be on. The natural thing to do is to try the switch and see if the light is on when the switch is up and off when the switch is down. If the status of the light depends on the switch position in all positions (i.e. on and off), we feel justified in concluding that setting the switch to the up position causes the light to turn on. Note that wiggling the value of E should have no effect on the value of C. Thus, interventions allow us to determine the direction of the causal arrow. Interventions thus provide us with an alternative sufficient condition on causation: if interventions on C are associated with changes in E, C causes E (Hausman and Woodward 1999).[9]

[8] Strictly speaking, dependence means $P(E \mid C) \leftrightarrow P(E)$, so an arrow will be justified only if this is true. For example, if $P(E \mid C) = .5$ but $P(E) = .5$, there will be no arrow between C and E in a model that satisfies the Causal Markov Condition.

[9] The notion of intervention may seem to approach the notion of agency, and indeed an alternative approach to causation has pursued this idea. Woodward (2009) separates 'manipulation-based' accounts into interventionist theories such as we have described, which refer merely to intervention by whatever external cause, and agency theories (e.g. Menzies and Price 1993), which define causation in terms of explicitly animate or human agency. Menzies and Price propose that rooting the theory in our personal experience of agency keeps the theory from being circular, which is a desirable outcome (and foreshadows the production class of theories, section 2.1.2). On the other hand, there are counterexamples to their claim that agency is a sufficient condition for causation, including cases where there is no possibility for an agentive manipulation (Hausman and Woodward 1999); at any rate, an animate intervener is not necessary in order to define an intervention, just as inanimate entities can be causers (see section 2.2.2).

The light-switch example suggests an additional way to use causal models: it is possible to model deterministic causal structures as well as probabilistic causal structures. The special case where the values of each variable are limited to 1 and 0 (true and false, on and off) yields tables reminiscent of familiar Boolean truth tables and is therefore possibly of more interest to linguists (though of less interest to probability theorists; Bayes's rule is no longer relevant). One example of such an approach is Hitchcock (2010), which shows how a deterministic model accounts for the problem of late pre-emption that so bedeviled previous dependency theories of causation.

In addition to the node-and-arrow notation, Hitchcock presents his causal networks in terms of *structural equations* (see also Sloman et al. 2009).[10] Consider the causal network in Fig. 2.2.

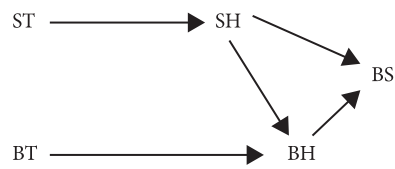

Fig. 2.2 An example of late pre-emption in terms of a direct graph (Hitchcock 2010)

The causal graph shown in Fig. 2.2 specifies the late pre-emption scenario discussed earlier, in which ST corresponds to Suzy's throw, SH to Suzy's ball hitting the bottle, BS to the ball's shattering, BT to Billy's throw, and BH to Billy's ball hits the bottle. The causal network shown in Figure 2.2 can be re-expressed in terms of structural equations as follows. The ":=" relation is an asymmetrical relation, read as "gets", in the opposite direction of the arrows.[11]

(1) SH := ST
BH := BT
BS := SH v BH
BH := BT and ∼SH

[10] These are equally available for the probabilistic case; here we examine the special case where values are either 1 or 0. One advantage to this notation over the node-and-arrow plus table notation is that it allows us to see at a glance whether a value of a parent variable has a positive or negative effect on the probability of a certain value of the child variable (since, as we have noted, both positive and negative effects are represented with the same kind of arrow).

[11] Note that there is an arrow from BT to BH and from BH to BS, even though we do not want to say that BT (or BH) causes BS. These arrows and the associated truth value distributions satisfy the Causal Markov Condition (see section 2.2.1.4), however. This failure in the face of late pre-emption shows that the Causal Markov Condition alone is not the correct sufficiency condition for causation.

As Hitchcock notes, an interventionist approach can offer a solution to the problem of late pre-emption. Given the equations in (1), we can simulate different scenarios by setting the variables to different values. For example, we could simulate the late pre-emption scenario by setting Suzy's hitting the bottle, SH, and Bill's hitting the bottle, BH, to "1". When this is done, the value of BS would be 1 as well; the bottle would shatter.

As we have said, a variable C is taken to cause E in a certain scenario if the values of E co-vary when C is wiggled and all external variables are held constant *at their actual values*[12] in that scenario. In the Suzy and Billy late pre-emption case, the actual values of Suzy's and Billy's throw, ST and BT, would be 1, the actual value of Suzy's hit, SH, would be 1, while the actual value of Billy's hit, BH, would be 0. It is interesting to see how such a graph is able to account for the intuition that, when both Suzy and Billy throw their rocks, with Suzy throwing first, we would describe Suzy and not Billy as the cause. To test whether Suzy is the cause, we need to hold BH fixed. In the actual scenario, BH is 0. Under these conditions, the value of BS would covary with the value of ST, implying that Suzy's throw is a cause of the shattering. To test whether Billy's throw is a cause, we need to hold SH to the value it has in the actual scenario, i.e., 1. With SH set to 1, the bottle would shatter regardless of the value of BT and the counterfactual test for BT would often be incorrect, offering evidence against BT being the cause of the shattering.

It is worth emphasizing the reason why the structural equation approach to encoding counterfactuals is able to account for late pre-emption. The reason why it succeeds is due to the asymmetry in the values of SH and BH. These two variables take on different values because of the requirement to freeze values at only their actual values; SH can be set to 1 while BH is set to 0, while the converse is not possible.

Causal Bayesian networks and structural equation modeling have several attractive properties: not only do they allow us to go beyond simple causal chains to specify causal networks in which some nodes have more than one parent (especially useful for counterfactuals; e.g. Dehghani et al. 2012), but they can be used to model both probabilistic and deterministic causation. Furthermore, they suggest straightforward accounts for late pre-emption. However, some concerns linger.

There are cases in which both of the sufficiency conditions mentioned hold between C and E, but C does not cause E: suppose that a villain gives the king poison (C), which causes the king's adviser to give the king an antidote, which on its own would kill the king but which neutralizes the poison harmlessly so the king survives (E) (Hitchcock 2007). In that case, it turns out that the intervention condition predicts C to be a cause of E, but we do not have the intuition that

[12] This requirement is an analogue to Lewis's similarity metric over possible worlds, relating them to the actual world: in both cases, certain other potentially interfering variables must be held constant at their actual value in order to determine if C causes E.

C causes E. Of course, one might propose a different sufficiency condition, and/or additional constraints on the model to explain these facts.

A more serious issue is the question of what these models are for. As Hausman and Woodward (1999) point out, it is curious to characterize causation in terms of intervention, which is itself arguably a causal notion. Such a characterization of causation is uninformative at best and circular at worst. This is not a problem if the models are used to analyze structures in which the direct causal relations are already known, and the question at hand is to find out how certain direct causal relations combine to yield causal relations in a complex structure. However, if these models are meant to be a theory of causation, and intervention is disqualified for circularity, it is only the Causal Markov Condition and other such conditions on the models that bear on the question of what causation actually is (and as demonstrated in the case of late pre-emption, the Causal Markov Condition is not enough to guarantee causation, though other conditions on the model can and have been added; see e.g. Woodward 2009). This is fine, but the complexity of the Causal Markov Condition and whichever additional conditions would be added to it raises the question of whether these are merely tests for whether certain structures can arise from causation, rather than accounts of our intuitive notion of causation itself (Mackie 1974).

2.2.2 Production theories

In the previous section, we touched on the major kinds of dependency theory of causation: logical dependency, counterfactual dependency, probabilistic accounts, and Bayesian and causal modeling accounts. What they have in common is the idea that causation can be explained by means of a dependency between the cause and the effect. The hope that motivates dependency theorists is that causation can be reduced to correlation or regularity if the conditions are pruned and the potentially confounding variables are fixed correctly. As we have seen, this hope is in large part justified by the success that such theories have had in providing appropriate sufficiency conditions for causal intuitions.

On the other hand, we seem also to have an intuition that something more than correlation or regularity is involved in causation (Pinker 2008; Saxe and Carey 2006). Hume recognized as much. He acknowledged that we often associate causation with a sense of force and energy. But for Hume, these were mental experiences that accompanied causation. He maintained that these notions could not be the basis for our understanding of causation, on the assumption that they could not be objectively observed. For Hume, these notions were imposed on experience by the mind, rather than experience imposing these notions on the mind. Ideas of force or energy are epiphenomena of our personal, subjective interactions with causation (Fales 1990; White 1999; 2006; 2009; Wolff and Shepard 2013).

It may be, however, that ideas of force and energy are more central to the notion of causation than was recognized by Hume, or for that matter by dependency

theories, which are largely descendants of the Humean perspective. One argument for why force and energy may be central to the notion of causation emerges when we consider the range of properties commonly associated with causal relationships.

One such property is temporal order: if C causes E, C must precede, or at least be simultaneous with E (Lagnado et al. 2007). This temporal relation between cause and effect is thus a necessary condition for causation. Like correlation, this relation is clearly not a sufficient condition for causation. Nonetheless, temporal precedence has been shown to be a stronger indication that C causes E than even correlation between C and E (Lagnado and Sloman 2006). The relationship between causation and temporality has been discussed by some linguistic researchers as well (e.g. Shibatani 1973a; Talmy 1976), though it has been ignored in much of the literature on the syntax–semantics interface.[13]

A second property is having a physical link between cause and effect (Salmon 1984; Walsh and Sloman 2011). This property requires some qualification. By physical link, we do not necessarily mean a direct physical contact; rather, that the cause and effect are linked in some way either directly or indirectly, through a chain of physical connections. This property appears not to be a necessary condition of causation, because of a large class of exceptions to this property that rely on "spooky action at a distance" (Einstein's famous description of quantum entanglement). This class includes not only gravity, electromagnetism, and quantum entanglement, but also magic and divine intervention.[14] Exactly when there is no plausible physical link, spooky influences such as these are called upon to justify impressions of causation.

These properties are problematic for dependency theories because these theories do not provide motivation for why these properties are relevant to causation. Temporal precedence or simultaneity, for example, is handled by stipulation, i.e. it needs to be explicitly stated in all these theories that the cause precedes the effect or occurs at the same times as the effect (Wolff, Ch. 5, this volume). The physical-link property is rarely if ever mentioned by dependency theorists. Why are these properties associated with causation? And is an answer to this question crucial to our notion of causation?

Our personal view is that the answer to that question is important, and since dependency approaches to causation give us no understanding of why these properties are relevant to causation, we must look elsewhere for an answer. In theories of causation based on concepts of force or energy, these properties of causation fall out naturally. A force is exerted or energy is transmitted, before or simultaneously with the effect that is provoked. Most forces also require a physical link, except, notably, for the class of spooky influences. These facts suggest that concepts such as

[13] See Copley and Harley (to appear) for a recent linguistic discussion of the difference between launching causation, in which the cause precedes the effect, and entrainment, where the cause and effect happen at roughly the same time (Michotte 1946/1963).

[14] Chains involving "social forces" might be thought to be part of this class, but as long as there is transmission of information from one person to another, there is still a physical link.

force and energy provide a necessary part of our notion of causation, and that Hume had it exactly backwards: that force and energy are in fact the basis of our notion of causation, while correlation and regularity are the epiphenomena.

Theories that characterize causation in terms of concepts such as force and energy view causation as a *production* or process. The production may involve a transmission of conserved quantities such as energy (Dowe 2000; Kistler 2006a; Ch. 4, this volume). It may also be viewed in terms of causal powers, namely the ability of entities to transmit or receive a conserved quantity (Mumford and Anjum 2011a). Yet another approach would be in terms of forces being imparted, for instance, by an agent to a patient (as in the parallel cognitive linguistic tradition, e.g. Talmy 1988; 2000; Gärdenfors 2000; Warglien et al. 2012; Croft 1991; 2012; Ilić, Ch. 7, this volume; also Wolff 2007; Ch. 5, this volume). See also Copley and Harley (Ch. 6, this volume) for a more abstract view of forces.

Theories of causation that characterize causation in terms of transmission include Salmon's (1984; 1998) mark transmission theory. In this theory, causation is understood primarily as a process rather than as a relation between events. A causal process is understood as a transmission of a causal mark, i.e. a propagation of a local modification in structure. A causal process would be instantiated if, for example, one put a red piece of glass in front of a light. In such a case, the red glass would impart a mark on a process that would transmit the mark to a different location, such as a wall.

Salmon's theory's greatest strength may be in its ability to distinguish, in certain circumstances, causation from pseudo-causation. However, because the theory emphasizes processes over events, it does not provide a direct definition of what counts as causation. It is not hard to imagine how Salmon's theory might be extended to provide such a definition. To say that A causes B might be to say, in effect, that a mark is propagated from A to B. In some cases, a procedure can be specified for determining whether a mark has been propagated. In the case of the light filter, one can check to see what happens when the filter is removed. However, in many other cases, procedures for determining whether a mark has been propagated are less clear. For example, in the ordinary billiard-ball scenario, what is the mark and how do we know it has been propagated? If the procedures cannot be specified, then the legitimacy of the causal relation should be ambiguous; but in the case of billiard-ball scenarios, at least, the the legitimacy of the causal relation is not in doubt. It might be possible, through further elaboration of the theory, to address this challenge. In particular, in order to make the criteria for causation easier to assess it would help to have a clearer idea of the notion of a mark.

A potential solution to this problem is offered by Kistler (2006a), who proposes a transmission theory of causation that brings back the idea of causation being a relation between a cause and an effect. According to this theory, "Two events c and e are related as cause and effect if and only if there is at least one conserved quantity P, subject to a conservation law and exemplified in c and e, a determinate amount of

which is transferred between c and e." Kistler (2006a) goes on to define "transference" as present if and only if an amount A is present in both events. In order for this to occur, events c and e must be located in space and time in such a manner that allows for the transference. In particular, the transference process requires spatial and temporal contiguity and implies that causation must take place over time (but does not, according to Kistler (2006a), necessarily imply that the cause precedes the effect).

Kistler's (2006a) proposal that causation involves a transference of a conserved quantity builds on a highly influential theory by Dowe (2000). According to Dowe's Conserved Quantity Theory, there are two main types of causation: persistence (e.g., inertia causing a spacecraft to move through space) and interactions (e.g., the collision of billiard balls causing each ball to change direction). Causal interactions are said to occur when the trajectories of two objects intersect and there is an exchange of conserved quantities (e.g. an exchange of momentum when two billiard balls collide). Unlike earlier theories, exchanges are not limited to a single direction (i.e., from cause to effect). One problem that has been raised for Dowe's theory—and that also applies to transference theories—is that such a theory is unable to explain the acceptability of a number of causal claims in which there is no physical connection between the cause and the effect. In particular, such a theory seems unable to handle claims about causing preventions or causation by omission (Schaffer 2000; 2004). Consider, for example, the preventative causal claim, "Bill prevented the car from hitting Rosy", assuming a situation in which Bill pulls Rosy out of the way of a speeding car. Such a causal claim is acceptable, even though there was no physical interaction between Bill and the car. Perhaps even more problematic are causal relations resulting from omissions, as when we say: "Lack of water caused the plant to die." The acceptability of such a statement cannot be explained by transmission or interaction theories since, plainly, there can be no transmission of conserved quantities from an absence.

Another type of production theory holds that causation is specified in terms of forces (Copley and Harley, Ch. 6, this volume; Talmy 1988; 2000; Gärdenfors 2000; Croft 1991; 2012; Ilić, Ch. 7, this volume; Warglien et al. 2012; Wolff 2007; Ch. 5, this volume). One such theory is Wolff and colleagues' force dynamic model (Wolff 2007; Wolff et al. 2010). According to this model, causation is specified in terms of configurations of forces that are evaluated with respect to an endstate vector. Different configurations of forces are defined with respect to the patient's tendency towards the end-state, the concordance of agent's and patient's vectors, and the resultant force acting on the patient. These different configurations of forces allow for different categories of causal relations, including the categories of cause-and-prevent relations. In a preventative relationship, there is a force acting on the patient that pushes it towards an end-state, but the patient is then pushed away from the end-state by the force exerted on it by the agent. Philosophers and cognitive scientists have

argued that transmission theories are unable to explain preventative relationships, as well as the notion of causation by omission. In order to capture these phenomena, it has been argued that theories of causation must go beyond a production view of causation to include, perhaps, counterfactual criteria for causation (e.g. Schaffer 2000; Dowe 2001; Woodward 2007; see also Walsh and Sloman 2011). Interestingly, Talmy's theory of force dynamics and, relatedly, Wolff's dynamics model are able to explain how the notion of prevention can be specified within a production view perspective without having to incorporate distinctions from dependency theories, such as counterfactual criteria. Wolff et al. (2010; see also Wolff, Ch. 5, this volume) also show how a production view of causation is able to handle the phenomenon of causation by omission—a type of causation which, according to several philosophers, is beyond the explanatory scope of production theories (Schaffer 2000; Woodward 2007).

Production theories have several attractive qualities. As already noted, they motivate why the concept of causation is associated with temporal and spatial properties. They also provide relatively simple accounts of people's intuitions about scenarios that are problematic for dependency theories, such as late pre-emption. In the case of late pre-emption, there are two possible causers and one effect. For example, in the Suzy and Billy scenario, Suzy and Billy both throw rocks at a bottle and the bottle breaks, but Suzy's rock hits the bottle first. Intuition says that Suzy's throw caused the bottle to break. This intuition falls out naturally from production theories: in transmission theories, in particular, Suzy is the cause of the breaking because it was from Suzy's rock that conserved quantities were transmitted to the bottle, while in force and power theories, Suzy's throw is the cause because it was from Suzy's rock, not Billy's, that force was imparted upon the bottle.[15]

Though production theories have several strengths, they also face several significant challenges. Without further qualification, production theories require that knowing that two objects are causally related entails being able to track the transmission of conserved quantities linking two objects. Such a requirement often does not hold for a wide range of causal relations. For example, common sense tells us that there is a causal relationship between the light switch and the lights in a room, but most of us could not say exactly how conserved quantities are transmitted through this system. As argued by Keil and his colleagues (Rozenblit and Keil 2002; Mills and Keil 2004), people often feel as if they understand how everyday objects operate, but when they are asked to specify these operations, it becomes clear that they have little knowledge of the underlying mechanisms. Keil and his colleagues refer to this phenomenon as the "illusion of explanatory depth". The

[15] As we have seen, Lewis in his later work (e.g. 2000) responds to criticism of his counterfactual approach by proposing that causation must be evaluated not just on whether C and E are true, but on finer properties of the events referred to by C and E. Late pre-emption is accounted for by noting that counterfactually changing the properties of Sally's throw changes the properties of the bottle's breaking, while changing the properties of Billy's throw does not. This interest in finer properties of events, rather than just truth values, is perhaps the closest approach of a dependency theory to the spirit of production theories, though it should be noted that Lewis's 2000 theory still relies on the notion of change rather than the notion of energy.

illusion of explanatory depth presents production theories with a challenge: how can causal relations be asserted of situations in which the underlying mechanism is not known? Production theory advocates might appeal to people's general knowledge of how things are likely physically connected, but such a move introduces uncertainty into their knowledge of causal relations, and if there is uncertainty, why not simply represent the causal relations in terms of probabilities and relationships between probabilities? Currently, there are no simple answers to such a challenge (but see Wolff and Shepard 2013).

A second major challenge for production theories concerns the problem of how such an approach might be extended to represent abstract causal relations. Production theories are clearly well suited for causal relations in which the quantities being transmitted are grounded in the physical world. For example, production theories seem especially well suited for explaining the acceptability of statements such as *Flood waters caused the levees to break*, or *The sun caused the ice to melt*. Much less clear is how a production theory might represent statements such as *Tax cuts cause economic growth*, or *Emotional insecurity causes inattention*. Obviously, such abstract instances of causation cannot be specified in terms of physical quantities, so how might they be represented? According to some theorists, abstract causation might be represented in a fundamentally different manner than concrete causation. Such a view has been dubbed *causal pluralism* (Psillos 2008; see Kistler, Ch. 4, Wolff, Ch. 5, this volume). Another possibility is that abstract causation might be represented in a manner analogous to physical causation (Lakoff and Johnson 1999; Wolff 2007). While explaining abstract causation in terms of metaphor might be an easy move to make, questions soon arise about such an explanation's testability. As with the problem created by the illusion of explanatory depth, there is currently no simple answer to the challenge raised by abstract causation for production theories (but see Wolff, Ch. 5, this volume, for an attempt).

2.3 Linguistic phenomena to which causation is relevant

Recall that our main purpose in this chapter was to demonstrate that theories of causation are relevant to linguistic theory and vice versa. Having presented the state of research on causation as we see it, we now turn to examine three linguistic domains to which causation is relevant. We will argue that a more sophisticated understanding of different theories of causation has real potential to advance our knowledge of these phenomena, and conversely, that these linguistic analyses, especially those concerning data from less familiar languages, should be taken seriously by philosophers and cognitive scientists working on causation.

The phenomena we will examine are:

Defeasibility: A causal relation has been proposed between a cause and effect in the two sub-events of accomplishments, but in certain environments, such as progressives, non-culminating accomplishments, and frustratives, the effect does not occur.

Volition, intention, and agency: Numerous linguistic phenomena seem to distinguish animate agents from inanimate causers. Language thus appears to be sensitive to whether the causing entity is volitional (or intentional) or not, though volitionality seems to not always be quite the right notion; rather, a broader causal notion that subsumes but is not limited to volitionality is called for. We argue that disposition fits the bill.

Representations of causal chains: How conceptual representations of events and participants in causal chains are mapped onto language, both to syntactic event chains and within certain lexical items.

Apart from being of interest in themselves, discussion of these phenomena will allow us to demonstrate several ways in which the choice of causal theory can bear on linguistic theory. For example, one instance of defeasible causation—non-culminating accomplishments—illustrates the heuristic that complex semantics should be adduced only when there is visible morphology in some language. Since different theories of causation predict different parts of causal semantics to be complex, choosing a causal theory whose distribution of complexity matches what is seen in the morphosyntax has the opportunity to greatly simplify the (morpho)syntax–semantics interface. Another example of how causal theory can be of use to linguists is when grammatical evidence suggests that certain concepts are linked, for example volitionality and the ability of inanimate objects to be the external arguments of activities (Folli and Harley 2008). In such cases, using a causal theory that explains the link between these concepts is preferable to one that does not, since it has a better chance of informing the semantic theory. Finally, it is fairly easy to see how causal theory can bear on the question of how conceptual causal chains correspond to the causal chains that are represented in language. In the mapping from conceptual causal structure to semantics, certain phenomena corresponding to components of causal theories are observed at the syntax–semantics interface, and/or the lexicon, while others are not. Ideally, for whichever causal theory is chosen, the components that are observed linguistically should be those that are important to the theory; conversely, nothing important to the causal theory should be completely invisible to language.

Two caveats must be mentioned. First, we are prepared for the possibility that different causal theories may be useful for different linguistic phenomena. This possibility is related to the notion of causal pluralism discussed earlier. However (and this is the second caveat), it should be clear that the study of language's relation to a mental representation of causation by definition has only to do with how causation is represented in the mind, not with the actual nature or metaphysics of causation. To the extent that certain philosophers are concerned with the metaphysics of causation rather than the mental representation of causation, a causal pluralism in language would not bear directly on their findings, although the use

of linguistic data by even metaphysically-oriented philosophers suggests that they might do well to pay attention to what linguists find.

2.3.1 Defeasibility

The first phenomenon we will examine is that of defeasibility, especially in the case of accomplishment predicates. These have been argued to involve a causal relationship between two eventualities (Pustejovsky 1991; 1995; Giorgi and Pianesi 2001; Kratzer 2005; Ramchand 2008; Higginbotham 2009; see also Thomason, Ch. 3, Ramchand, Ch. 10, and Lyutikova and Tatevosov, Ch. 11, this volume). So, for instance, *Mary build a house* is true, roughly speaking, if Mary is the agent of an event that causes the state of a house existing.

Not everyone treats accomplishments as involving a causal relation. Parsons (1990), building on Bach (1986), uses a part–whole relation between a non-culminated event (e.g. an event of Mary's building, up to but not including the part where there is a completed house) and a culminated event (an event that includes e.g. the part where there is a completed house). While a *Mary build a house* event holds if the event in question is at least a partial event of Mary's building a house, it culminates just in case it is a complete event of Mary's building a house. As Portner (1998) has noted, Parsons has no particular definition of the relation between events of a certain sort that hold and events of the same sort that culminate. That said, recent non-causal treatments of accomplishments (Piñon 2009, e.g.) characterizes such a relation in terms of gradability,[16] relying on the assumption that we intuitively have a sense of the degree to which an event is complete.

Despite the existence of this alternative perspective, however, the idea that accomplishment predicates involve a causal relation has widespread support. This idea is generally represented by causal relation that relates two existentially bound events.

(2) $\exists e_1\ \exists e_2 : e_1$ CAUSE e_2

However, accomplishments can occur in environments in which the caused event does not occur; these are the cases of "defeasible" causation. The particular theory of causation one chooses will have consequences for one's linguistic theory of these phenomena. The reason why is because some theories of causation are "result-entailing"—i.e. they assume that the caused event occurs—whereas other, "non-result-entailing" theories lack this assumption. As we will see, defeasible environments suggest that it might be best that one's theory of causation be non-result-entailing.

Cross-linguistically, there are a number of expressions which include accomplishment predicates but which do not entail the final result.[17] If there is indeed a causal

[16] Koenig and Chief (2008) also use gradability to account for non-culminating accomplishments. However, they retain the notion of causation.

[17] Sometimes the "accomplishments" are predicates that are normally considered achievements; we follow existing practice in calling all cases "non-culminating accomplishments".

relation in accomplishments, this may be surprising, since there is a causal connection and yet the final caused event is not actualized. The simplest case of these is non-culminating accomplishments, as in the examples below.[18] The assertion in such expressions is that something has been done that would normally be expected to cause the effect. Thus culmination is implicated but not entailed. Most typical are cases where the non-culminating case is expressed with an ordinary verb form, where the culminating case expressed by a morpheme meaning something like 'finish', with or without the (non-culminating) verb, as is shown in (3) and (4):

(3) a. Watashi-wa keeki-o tabeta dakedo keeki-wa mada nokotteiru.
I-TOP cake-ACC ate-PERF but cake-TOP still remains
'I ate the cake but some of it still remains.'

b. *Watashi-wa keeki-o tabeteshimatta dakedo keeki-wa mada nokotteiru.
I-TOP cake-ACC eat-finish-PERF but cake-TOP still remains
'I ate the cake but some of it still remains.'

(Japanese; Singh 1998: 173–4)

(4) Wo gai le xin fangzi, fangzi hai mei gai-wan.
I build PERF new house house still not build-finish
'I built a new house, but it is still not finished.'

(Mandarin; Koenig and Chief 2008: 242)

The 'finish' verb can often be used as a main verb to express culmination:[19]

(5) a. K'ul'-ún'-lhkan ti ts'lá7-a,
make-DIR-1SG.SU DET basket-DET
t'u7 aoy t'u7 kw tsukw-s.
but NEG just DET finish-3POSS
'I made the basket, but it didn't get finished.' (St'át'imcets)

b. Kw John na kw'el-nt-as ta skawts
DET John RL cook-DIR-3ERG DET potato
welh haw k-as 7i huy-nexw-as.
CONJ NEG IRR-3CONJ PART finish-LC-3ERG
'John cooked a potato but never finished.' (Skwxwú7mesh)

(Bar-el et al. 2005: 90)

[18] Note that the non-culminating accomplishments are not progressives, or imperfectives; see Bar-el et al. (2005).

[19] See also Zucchi et al. (2010) for an account of FATTO/FINISH in Lingua dei Segni Italiana and American Sign Language as completion markers.

(6) Kerim ešik-ni ac-xan-di, alaj boša-ma-ʁan-di.
Kerim door-ACC open-PERF-3SG but finish-NEG-PERF-3SG
(Context: The lock is broken and Kerim tries to open the door.)
Lit. 'Kerim opened the door, but he did not succeed.'
(Karachay–Balkar; Tatevosov 2008: 396)

Verbs other than 'finish' are sometimes also used to signal culmination, such as a verb meaning "take" in Hindi:

(7) a. Mãĕ ne aaj apnaa kek khaayaa aur baakii kal khaaũũgaa.
I ERG today mine cake eat-PERF and remaining tomorrow eat-FUT
'I ate my cake today and I will eat the remaining part tomorrow.'

b. *Mãĕ ne kek khaa liyaa, jo bacaa hae wo raam khaayegaa.
I ERG cake eat take-PERF what remain is that RAM eat-FUT
'I ate the cake and Ram will eat the rest.'
(Hindi; Singh 1998: 173–4)

In all these examples, the non-culminating form may not be as morphologically marked as the culminating form (as in (3) and (7)). Something is added to the non-culminating form to get the culminating form. Mandarin (as in (4)), and certain Austronesian cases (Tagalog and Malagasy) discussed by Dell (1987) and Travis (2000), seem to be a counterexample to this generalization. Tagalog, for instance, distinguishes between a neutral (N) form, which does not entail culmination, and an abilitative (A) form, which does:

(8) Inalis ko ang mantas, pero naubusan
N-PERF-remove GEN-I NOM stain, but run-out-of
ako kaagad ng sabon, kaya hindi ko naalis.
NOM-I rapidly GEN soap hence NEG GEN-I A-PERF-remove
'I tried to remove (lit. 'I removed') the stain, but I ran out of soap, and couldn't.'
(Tagalog; Dell 1987: 186)

In these Austronesian cases, the culminating and non-culminating forms are both (differently) morphologically marked. The appropriate generalization thus seems to be that the culminating cases are at least as morphologically complex as the non-culminating cases, and very often more so.

A non-morphologically-marked alternation between culminating and non-culminating readings of accomplishments can also show up with certain lexical items which can have either culminated readings (as for *warn* in (9a), where an unsuccessful warning is described via the addition of *try*) or non-culminated readings (as for *warn* in (9b), where *try* is not used to describe an unsuccessful warning):

(9) a. I tried to warn him, but he didn't listen.
b. I warned him, but he didn't listen.

Defeasibility also arises for accomplishments in combination with certain morphemes that seem to carry additional meaning, notably frustrative morphemes, as in (10) and progressives (as has long been noted; see e.g. Dowty 1979).

(10) Huan 'o cem kukpi'ok g pualt.
Juan aux-IMPF FRUS open DET door
'Juan *cem* opened the door.'
= 'Juan pulled on the door but failed to open it.'
(Tohono O'odham; Copley 2005a; Copley and Harley, Ch. 6, this volume)

(11) Mary was building a house, but then she had to leave the country, so the house was never finished.

If we still want to maintain that cross-linguistically there is a causal relation in accomplishments and a causal relation requires the caused event to occur (i.e. the caused event argument is existentially bound), such cases of defeasibility are problematic, because in these cases a causal relation is supposed to hold, despite the non-occurrence of the final event.

Dowty's (1979) solution to this problem with respect to the progressive was to say that in these cases, the caused event occurs only on certain possible "inertia" worlds —the worlds where events proceed normally or stereotypically. This kind of approach has also been adopted for non-culminating accomplishments by Matthewson (2004), Tatevosov (2008), and Martin and Schäfer (2012). Using this solution, both the causation relation in accomplishments and the insistence that the caused event occurs (the existential quantification on the caused event) can be maintained. This solution comes at the cost, however, of additional semantic machinery for which there is cross-linguistically little or no morphological support; as we have seen, languages with both culminating and non-culminating forms of predicates do not mark the culminating forms with more morphology than the non-culminating forms, so it is odd that the non-culminating forms would involve so much extra meaning. Furthermore, as Martin and Schäfer (2012) point out for their data, there is no evidence that such modals take scope over other modals, which would also tend to indicate that they are not visible to the logical form per se.[20]

Dowty (1979) uses a result-entailing theory, a version of Lewis's (1973) counterfactual theory of causation. Though we are used to expressing Lewis's counterfactual condition as a mere conditional—if C had not occurred, E would not have occurred—in fact the inverse is also asserted: if C had occurred, E would have occurred. Causal dependency of E on C thus requires that E occur, since there is no way to simultaneously have E not occur on the actual world, and at the same time for the causal dependency biconditional to be true, i.e. for some C-world where E occurs to

[20] Note however that if the inertia world possibilities were not represented in the semantics, but somewhere else, e.g. in a conceptual model inaccessible to (morpho)syntax, neither of these problems would arise.

be closer to the actual world than any C-world where E does not occur ("Our actual world should be closest to actuality, resembling itself more than any other world resembles it": Lewis 1973: 560). Stepwise dependency does not improve matters. So Dowty was right to see that Lewis's theory did not sit easily with defeasibility.

Other result-entailing theories are equally problematic. For instance, logical dependency theories that define causation in terms of necessity and sufficiency (e.g. Sosa and Tooley 1993) also require the caused event to occur. Under the view that a cause is a sufficient condition for an effect, the occurrence of the effect is guaranteed in the presence of a cause. However, if actuality is dropped, assessing sufficiency becomes difficult. For example, if a cause occurs and the effect does not, this presumably should count as evidence against the sufficiency of the cause, and hence as evidence against there being a causal relation, assuming causation is based on sufficiency. Structural equation theories can be instantiated with probabilities, in which case (as we will see just below) they inherit the benefits of probability relationships, and do not entail actuality of the caused event. However, if the values are limited to truth values, as in the model we discussed earlier, then the dependencies used to check for the causal relation are logical necessity and sufficiency. But it is difficult to model defeasible causation in terms of logical necessity and sufficiency, as we have seen. Certain production theories, such as Kistler's (2006a) transference theory, are also result-entailing. In this theory, causation is defined as a transference of a conserved quality from the cause to the effect. If the effect does not occur, there can be no transference on conserved quantities.[21]

We may provisionally conclude that, if the idea that there is a causal relation in accomplishments is to be maintained, while still being faithful to the morphosyntactic facts, it is better to choose a theory of causation in which the effect is not required to be actual, i.e. a non-result-entailing theory. Two kinds of causal theory fit the bill: probability theories and certain production theories. To use either one of these in a semantic theory of non-culminating accomplishments, a technical innovation (different in either case) is required to ensure that the effect does not occur.

One approach to specifying defeasible causation would be in terms of probability raising, which allows for exceptions to sufficiency and necessity. As we have seen, probability-raising theories would say that C (the occurrence of e_1) causes E (the occurrence of e_2) iff $P(E \mid C) > P(E \mid \sim C)$: the probability of E occurring is greater if C occurs than if C doesn't occur. In such probability statements there is no existence or occurrence asserted of the caused event E, but just the assertion that the frequency of a certain kind of event is higher in the presence of another kind of event. This seems to indicate that a probability-based theory should make use of event kind arguments (Gehrke 2012), as individual events do not have frequencies; only kinds of events do. The use of event kind arguments is equally helpful with the technical issue

[21] However, as we will see next, Kistler's theory may offer a solution to this problem.

of ensuring that e_2 does not necessarily occur, as an event kind argument does not refer to an actually occurring event, even when it is existentially bound.[22]

We might then say that if the probability statement, using event kinds, is true, one can assert the non-culminating predicate with an individual event argument as the causing event and an event kind argument as the caused event. Using 'O' in the spirit of Lewis's "occur" typeshifting predicate to turn events into propositions, so that we can use negation, we might write the following. On the left-hand side, c is an individual event, because neither culminating nor non-culminating accomplishments predicate something of kinds. The caused event is either an event kind, as in (12a), reflecting a non-culminating causal relation or an individual event as in (12b), reflecting a culminating causal relation. The definition given on the right-hand side is the same in either case.

(12) $O(e_k) = 1$ iff there is an eventuality e such that e realizes e_k

a. for any eventuality c and eventuality kinds c_k and e_k where c realizes c_k:
$c\ cause_{non\text{-}culminating}\ e_k$ iff $P(O(e_k) \mid O(c_k)) > P(O(e_k) \mid \sim O(c_k))$

b. for any eventualities c, e, and eventuality kinds c_k and e_k where c realizes c_k and e realizes e_k:
$c\ cause_{culminating}\ e$ iff $P(O(e_k) \mid O(c_k)) > P(O(e_k) \mid \sim O(c_k))$

This solution does not seem to predict, as we might wish, that the culminating case is cross-linguistically the more morphologically marked case, but in any case it is not incompatible with the morphological facts.

Probability theories of causation are useful here because they are non-result-entailing. Another way out of the actuality requirement on the caused event is to adopt a production theory of causation that is non-result-entailing.[23] If we define causation in terms of forces, such an account offers us a possible explanation for what accomplishment predicates might denote, whether they culminate or not. That is, they might refer to force interactions. Forces can be imparted with the occurrence of an effect, such as when a person kicks a tire or slaps a table; or when another stronger force opposes it, as when you hold an object, opposing a gravitational force. Whether there is a result or not is an entirely different matter. The non-culminating

[22] The idea of treating defeasible causation in terms of event kinds is due to Berit Gehrke (p.c.); the particular implementation in terms of probability raising is ours.

[23] Yet another way out of the actuality requirement on the result event is to change the relata—a tactic we have already used in our appeal to event kind arguments. Instead of using events, or event kinds, one could also take accomplishments to involve inferential relations between propositions, as Asher (1992) and Glasbey (1996) have proposed for the progressive. The inferential relation is in effect a defeasible causal relation between propositions rather than events. The inferred proposition, i.e. the proposition that the result event occurs, which is caused to be defeasibly inferred when the speaker learns that the proposition that the causing event occurs, does not require existential binding, as it is a proposition and not an event. Such a solution is also offered by Thomason in the last part of Ch. 3, this volume. But note the similarity to production theories, with information playing the role of force; both are naturally understood as having effects that are defeasible in the presence of perturbing influences (i.e. information to the contrary).

cases are thus seen as semantically more basic than the culminating cases, which is also morphologically desirable, as we have seen.

Another advantage to production theories is that they make more sense of accomplishment predicates in which the causal relation is less plausible. It is arguably odd to say that *Mary walked to the store* is analyzed as one in which Mary's walking event *causes* a state of her being at the store. However, it is unobjectionable to say that the force, or effort, expended by Mary in walking *resulted in* her being at the store. This is possible with production theories because they are not only theories of intuitions about the word *cause*, but about our intuitions about forces (or energy, etc.) and their results. Note that this point represents a retreat from the idea that accomplishments involve a causal relation in the narrow sense. This retreat provides no aid to result-entailing theories, however. The reason is because in these theories, causation is defined in terms of the occurrence and non-occurrence of the result. If you peel away the occurrence of the outcome, very little or nothing is left in the case of result-entailing dependency theories such as Lewis (1973). In contrast, in the case of production theories, if the outcome does not occur, there can still be a very specific kind of relationship referred to by the accomplishment predicate. According to force theories, that relationship would be a specific pattern of forces or force interaction. In terms of transmission theories, if the outcome is removed, there can still be the occurrence of a transmission because there can be the transmission of a conserved quantity (e.g. momentum, energy) without the occurrence of a (noticeable) result. According to Kistler, causation entails the occurrence of a result. Kistler would therefore not view non-culminating accomplishments as instances of causation; however, his account would be able to represent non-culminating accomplishments using some other kind of transmission relation besides causation.

As far as semantic theory is concerned, the use of forces or energy transmission is prima facie problematic if Davidson's (1967) event arguments are to be maintained. However, Copley and Harley (to appear; Ch. 6, this volume) have integrated force-dynamic theory into the syntax–semantic interface by treating the Davidsonian argument as a force, where such arguments formally have the meaning of functions from situations to situations. This move also provides a solution to the technical problem of the non-actuality of the caused eventuality, as the output of a function does not need to be existentially bound.[24]

We have seen that if we want to maintain a causal relation as a component of accomplishments, we must ask whether and how our conception of causation accounts for cases of defeasibility. Although many theories of causation have no

[24] And thus its referent need not (be asserted to) occur. On the other hand, why some cases are culminating and others are non-culminating must still be accounted for. A situation kind argument could be employed, or the assumption that the situation is all the speaker need consider (equivalent to the assumption that the modal base is totally realistic, the "closed world" assumption, or Copley and Harley's (to appear; Ch. 6) efficacy presupposition; such assumptions would correspond to the "finish" morphology).

easy way to address this question, the use of one of the theories that does—probabilities with event kinds or force dynamic theories—should be of use to linguists studying this phenomenon. Likewise, philosophers would be well advised to pay attention to the existence of non-culminating accomplishments and their treatment in linguistic theories.

2.3.2 Understanding agency through dispositions

We have addressed how theories of causation may apply to apparent cases of defeasibility. Researchers should also consider how different theories of causation might apply to agency and related notions. Languages are sensitive to a collection of concepts related to agency that seem particular to—or at least prototypical of—humans, including animacy or sentience (Dowty 1991; Ritter and Rosen 2010), and volitionality, which for our purposes we will consider equivalent to intentionality, though that may not be precisely the case.

The notion of volitionality has played a significant role in linguists' understanding of the syntax–semantics interface, especially through the related notion of agency (see, e.g. Duffield, Ch. 12, this volume). Intention is another similar concept. The precise way in which grammar interacts with these concepts is, however, still the subject of very active discussion. Recent theories have proposed that, although there are numerous cases where volitionality seems to be a distinguishing factor, volitionality is not the concept that the grammar is sensitive to in these cases; rather, it is sensitive to properties of the causal structure being referred to. This idea suggests that the notion of disposition—a notion that has links to causation—is relevant. Our point here is that a wider appreciation of existing theories of causation could be very helpful to understanding these issues.

2.3.2.1 Volitionality is not always the right notion While certain kinds of linguistic data are sensitive to whether the subject is volitional (or intentional, but we will largely not distinguish these notions in this chapter), it is not always clear whether this sensitivity reflects a real grammatical distinction or an epiphenomenon. Some of the clearer cases are those in which there is a strict selection of either animate, volitional agents (as in lexical cases such as *murder*) or (possibly) inanimate, non-volitional causers, as in *from* phrases. The latter seem to denote the cause of an event, as in (13). Only non-volitional entities are acceptable as complements of *from*; volitionally acting entities, as in (14), are unacceptable (see Piñon 2001a; 2001b; Kallulli 2006; Alexiadou et al. 2006; Copley and Harley, to appear):

(13) a. The window broke from the pressure/the rock.
b. The door opened from the wind.
c. The floor collapsed from the elephant's weight.
d. He got sick from pneumonia.
e. The sidewalk is warm from the sun.

(14) a. *The window broke from Mary.
b. *The door opened from the man.
c. *The floor collapsed from the elephant.

However, it is not always clear whether grammars always "see" volitionality per se or whether something that looks like a volitionality requirement is actually a case where language refers to a certain kind of causal structure, of which volitionality is just one example (albeit the most prototypical one). This seems to be the case in many phenomena which appear at first to distinguish animate, volitional agents from possibly inanimate, non-volitional causers, but which upon closer inspection do not make such a clear distinction.

Thomason (Ch. 3, this volume) reaches such a conclusion on the basis of considerations of how conceptual causal structures map to our intuitions about subjects. Likewise, a similar consensus has been reached in the literature on argument structure. For instance, Folli and Harley (2008) argue that while agents are prototypically animate and volitional, in certain cases inanimate entities behave like agents as well. The subjects of unergatives such as verbs of sound emission like *whistle* are without question agents rather than themes, as shown, for example, by the fact that they pass the *V x-self Adj* test (Chierchia 1989; 2004; Levin and Rappaport Hovav 1995), as in (15). The agent of *whistle* can be an inanimate, non-volitional entity, as in (16a). Notably, however, certain inanimate entities require an extra phrase such as *into the room* in order to serve as the agent for a verb like *whistle*, as demonstrated in (16b) (Levin and Rappaport Hovav 1995).

(15) a. John whistled himself silly.
b. *John fell himself silly.

(16) a. The teakettle whistled.
b. The bullet whistled *(into the room).

Folli and Harley attribute the contrast in (16) to the difference between the teakettle's causal properties and the bullet's causal properties. "Agents, then, are entities which can produce particular events by themselves: they are sufficient on their own to initiate and carry out the entire event denoted by the predicate" (Folli and Harley 2008: 192). The teakettle can initiate and carry out the whistling event on its own, while the bullet can only do so if it is in motion (hence the need for a path-denoting prepositional phrase). They conclude that agency is sensitive not to animacy or volitionality, but to a more fundamental property termed by Higginbotham (1997) *teleological capability:* "the inherent qualities and abilities of the entity to participate in the eventuality denoted by the predicate".

Latrouite (Ch. 14, this volume) proposes a similar but more permissive causal condition for voice selection in Tagalog. Subjects in Tagalog, as in other Austronesian languages, are not prototypically agents; the question of which argument is realized as the subject is a complex one and as yet no consensus exists. Latrouite argues that subjects in Tagalog are *event-structurally prominent*: that is, a subject

must be an argument that is crucial for the event because it delimits the run time, either because it is an agent and can volitionally decide the course of the event up to the endpoint, or because it is not an agent, but otherwise causally determines the endpoint of the event.

In a similar vein, Sichel (2010) observes that certain English nominalizations of non-alternating lexical causatives, unlike their corresponding verbal forms, apparently require *direct participation* of their subject in the event. This is demonstrated by the fact that although the verbal form in (17a) is felicitous with *the results* as the subject, the nominalized version in (17b) is not. *The expert* is felicitous, as in (17c), but this is apparently not because of its animacy or volitionality, because *the sun* can appear as the subject of the nominalization *illumination*. Sichel argues that the difference between (17b) and (17d) is that the results do not directly by themselves verify the diagnosis (though an expert does) while the sun does directly illuminate the room.

(17) a. The results/the expert verified the initial diagnosis.
b. #The results' verification of the initial diagnosis.
c. The expert's verification of the initial diagnosis.
d. The sun's illumination of the room.

Alexiadou et al. (2013) who investigate this phenomenon in English, German, Greek, Romanian, Spanish, French, Hebrew, and Jacaltec, provide further support for the idea that volitionality is not responsible for the effect, even though there is a great deal of variation among constructions and across languages.

Another causal property that is more general than volition is the ability to have *direct causation of a temporally distant effect*. Copley (2014) argues that (18a), which lacks future marking, is acceptable in English ultimately because a volition today for the Red Sox to play the Yankees is asserted to directly cause them to do so tomorrow, and volitions can directly cause temporally distant effects. No volition is plausible that would make (18b) similarly felicitous. Copley argues that cases such as (18c) and (18d) also involve current properties of the subject that have temporally distant direct effects, though no volition is involved.

(18) a. The Red Sox play the Yankees tomorrow.
b. #Mary gets sick tomorrow.
c. The sun rises at 5 tomorrow.
d. The hurricane arrives tomorrow.

Like the ability to be the agent of an activity, then, event-structural prominence, direct participation, and direct causation of a temporally distant effect are conditions that make no reference to animacy or volition, but rather to causal properties of the entity.

Many of the above examples make reference to lexical causatives; analytical causative facts also suggest that volitionality is often not quite the right notion. For example, Ramchand (Ch. 10, this volume) relates an apparent volitionality requirement in certain Hindi analytic causatives to the lexicalization of both the

"initiation" and "process" subevents—again, a causal notion rather than volitionality. We can also add English *have* causatives as in (19) (Dowty 1979; Ritter and Rosen 1993; Copley and Harley 2009; 2010). *Have* causatives have been claimed to permit only subjects with volitional control (Ritter and Rosen 1993), but inanimate causers are possible too, as in (19b)—as long as the caused eventuality is a stage-level stative, e.g. *laughing* instead of *laugh*, in this case.[25]

(19) a. Mary had Sue laugh/laughing.
b. The book had Sue *laugh/laughing.

Again, such examples suggest that the grammar cares about a causal property which can be exemplified by volition, but is not limited to it.

2.3.2.2 Implications for the kind of volitionality involved in agency The above discussion suggests that linguists would benefit from making use of theories of causation in which these notions such as teleological capability and direct causation or participation can be easily explained. In other words, what is needed is a theory of causation that ensures that volitional causing entities and certain non-volitional causing entities are treated similarly. In part this will require us to rethink what volition is in the context of agency. For instance, Folli and Harley propose that the requirement for agents is not volition, but is rather teleological capability, the "inherent qualities" that are among the causing conditions causing an entity to participate in an event. Volition might be thought of as one of these qualities, and its analysis should therefore be similar to the analysis of other such qualities in relevant ways.

Volition has been studied in the context of the main verb *want*, which probably does not have the same meaning as volition in the context of agency—indeed, unlike agency, (standard) English *want* really does require animacy—but which is a good place to start. Heim's influential 1992 account presents a denotation which is squarely in the counterfactual tradition of Lewis (1973), via Stalnaker's (1968) extension of Lewis's theory to non-counterfactual conditionals:

(20) 'α wants φ' is true in w iff: for every w' $\in$ Dox_{α} (w):
Every φ-world maximally similar to w' is more desirable to α in w than any non-φ- world maximally similar to w'.

[25] Another reason why *have* causatives make us think that a better notion of causation would be useful is the fact, originally pointed out by Ritter and Rosen (1993) that a sentence such as *Mary had Sue laughing* has a special "director's" reading. The sense in which Mary is a "director" is in that Mary must have near-omnipotent power over Sue: she is either the director of a play that Sue is in, or an author writing about Sue and choosing what she will do in a scene, etc. The particular oddness of the director's reading ought equally to stem from some peculiarity of the causal relationship between the subject and the caused eventuality; Copley and Harley (2009; 2010) suggest that the director's reading occurs when the director's volition directly causes the eventuality.

The idea is that a wanter considers only the worlds they believe to be possible; "α wants φ" holds if and only if, among the worlds that are maximally similar to the belief worlds, the φ-worlds are all preferable to the non-φ-worlds. Wanting is thus explained in terms of belief and desirability, i.e. preference (see also Egré, Ch. 8, this volume).

Such a denotation cannot obviously be extended to describe teleological capability. The teakettle does not whistle due to any beliefs or preferences. Now, it is not likely that we will be able to entirely elide the reference to the animate agent's beliefs, as the ability to hold beliefs is a significant difference between animate and inanimate entities.[26] But regardless, there are apparently also similarities between animate and (certain) inanimate entities in terms of their causal abilities. It would therefore be desirable to replace preference with a property that is more conducive to understanding how the grammar can sometimes see physical causing properties in the same way it sees volition.

An alternative way of defining *want* has been explored not only by Lewis (1986b) but also by Stalnaker (1978), Portner (1997), and Condoravdi and Lauer (2009). This alternative way of defining *want* treats desire as a disposition instead of a preference: i.e. roughly, as a state of readiness, or tendency to act in a specific way. The general idea is that wanting p is a property (a state, since dispositions are states) that holds if, should the circumstances be right, the entity would act to bring p about. Consider one dispositional definition of volition in (21), taken from Portner's (1997) discussion of exactly this issue; note that belief is still a part of this definition, through the wanter α's doxastic worlds $\text{Dox}_{\alpha}(b)$.

(21) Portner (1997)
For any wanting situation s of α and belief state b of α, $\text{want}_{\alpha,\, b}(s)$ = the set of plans which would satisfy α's desire in s, relative to his or her beliefs in b = {s′:
a. for some $w \in \text{Dox}_{\alpha}(b)$, $s' \leq w$, and
b. s′ begins with a dispositional counterpart s″ of s,
c. α acts in s′ in ways which tend, given Dox_{α} (b), to bring it about that s″ develops into s′, and
d. α is disposed in s to act in those ways}

Both (20) and (21) are, as far as we know, perfectly reasonable characterizations of volition. The fundamental difference between them is that the preference denotation in (20) explains wanting in terms of preference, whereas the dispositional denotation in (21) explains wanting in terms of disposition to act. Inanimate entities, as well as animate ones, have dispositions. We would not generally call their dispositions "dispositions to

[26] Although the beliefs might be understood as another causal property.

act", but it is not a stretch to call them "dispositions to cause", and this perspective could equally apply to the animate entities. Thus, unlike the preference definition, the dispositional denotation immediately suggests a commonality between a volitional entity and a non-volitional but teleologically capable entity.[27]

We can say, then, that there are two kinds of disposition: psychological (of which we are concerned with one, namely volition) and physical (non-volitional teleological capabilities). Apparent volitionality requirements in language often are dispositionality requirements instead. But here a problem seems to arise regarding intensionality.[28]

A widely-accepted thesis attributed to Franz Brentano (1874) holds that intensionality distinguishes psychological from physical phenomena: only psychological phenomena are intensional. Volition, for instance, is clearly intensional, according to several non-controversial criteria for intensionality (Place, 1996; Molnar 2003), including: (i) directedness towards something, e.g. the directedness of a desire toward the propositional content of the desire; (ii) the fact that an intensional object may be either existent or nonexistent, e.g. John may want a unicorn although no unicorns exist; and (iii) referential opacity, the fact that co-referring expressions are not substitutable, e.g. just because Mary wants to see the morning star does not mean she wants to see the evening star. If physical dispositions are analogous to volitions, they should be intensional too, contra Brentano's thesis. And in fact this is what Place argues, that "inten[s]ionality is the mark, not of the mental, but of the dispositional" (Place 1996: 91). Namely, (i) dispositions are directed toward their proper manifestation; (ii) dispositions exist whether or not their manifestations exist; and (iii) they provoke referential opacity, as shown by the fact that *Acid has the power to turn this piece of litmus paper red* does not entail that *Acid has the power to turn this piece of litmus paper the colour of Post Office pillar boxes* (Molnar 2003: 64).

The idea of physical intensionality is controversial, and previous arguments for it have been challenged on various grounds, the strongest perhaps being that the criteria for them such as (iii) are themselves linguistic, i.e. in the mind (Crane 1999). In Mumford's introduction to Molnar (2003), he points out that criteria (i) and (ii) are not in the mind, though from a linguistic perspective it would not matter if they were; if dispositions are in the mind we do not care, as long as they are intensional. This point suggests that we are on firm ground: the intensionality of volition is no barrier to understanding non-volitional teleological capabilities as (intensional) dispositions. A dispositional definition of volition can and (given the

[27] It is intriguing to speculate about whether psychological research methods might bear on the question, by distinguishing preference vs. disposition to act a certain way.

[28] *Intensionality*, often (though not always) spelled with an "s", refers to a kind of meaning that e.g. distinguishes *the morning star* from *the evening star* even though both expressions refer to the same object. Intentions (with a "t") and beliefs are examples of intensions. Some characterizing properties of intentions are given just below in the text.

linguistic data) should be used to characterize the volition that is relevant to agency, and apparent volitionality requirements can and should be analyzed in terms of dispositionality.

2.3.2.3 Intentions, like volitions, can be seen in two ways Shortly we will ask how the theories of causation can be used to address dispositions, so that we can determine how they account for volitionality. First, however, we will make a small digression. We said at the beginning of this section that we were going to treat volitions and intentions as essentially the same. And in fact, experimental evidence relating to intentions mirrors the two forms of volition that have been proposed. Recent experimental work has demonstrated an intriguing connection between an action's intentionality and the goodness or badness of an outcome brought about by that action. The connection between intentionality and good and badness can be illustrated by scenarios in which an actor moves forward with a plan that brings about either a bad or good side-effect. A scenario in which the side-effect is bad is shown below (see Knobe 2003b):

(22) The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment."

The chairman of the board answered, "I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was harmed.

After reading the above scenario, Knobe (2003b) asked participants: "Did the chairman intentionally harm the environment?" Knobe (2003a) found that 82% of the participants respond that the chairman intentionally harmed the environment. A very different result was found when the side-effect was described as good. When the scenario was changed so that the business plan not only made a profit but also helped the environment, only 23% of the participants felt that the chairman intentionally helped the environment (Knobe 2003a). The basic finding has been replicated by a number of researchers across a wide range of scenarios (Shepard and Wolff 2013; Sloman et al. 2012; Knobe 2010). The only outward difference between the bad and good side-effect scenarios was the badness of the side-effect. According to Knobe (2003b; 2006; 2010), difference in intentionality are driven by differences in badness, but a number of other explanations have been offered. One general class of alternative explanation holds that the asymmetry observed in Knobe (2003b) is due to differences in causal structure (Shepard and Wolff 2013; Sloman et al. 2012; Nanay 2010).

Shepard and Wolff (2013) have found that when Knobe's scenarios are changed to include unjust laws in which there is greater pressure (a production theory notion)

against doing a good than bad thing, the alignment between badness and intentionality flips: participants are more willing to say that doing a good thing is more intentional than doing a bad thing, but the relationship between intentionality and causal structure remains the same (Shepard and Wolff 2013). Relatedly, Nanay (2010) notes that the good and bad scenarios used in Knobe's research are associated with a difference in counterfactual dependency: if the chairman had not ignored the environmental ramifications of harming the environment, he might not have made the same decision, whereas if the chairman had not ignored the environmental ramifications of helping the environment, his decisions would probably have been the same.

Note how these three explanations mirror what we have already seen for volitionality in section 2.3.2.2. Volitionality can be thought of in terms of preference (i.e. goodness/badness), like Knobe's theory; cf. Heim's (1992) denotation for *want*. Alternatively, volitionality can be thought of in terms of dispositionality to act, as per Portner (1998). Note too that with a dispositional view, as with volitions, intentions can either be seen in terms of counterfactuals as in Nanay's account, or in terms of tendencies, as in Shepard and Wolff's account.[29]

To summarize: we have seen evidence so far that both volitionality and intentionality can be viewed either in terms of preferences or dispositions. If they are viewed in terms of dispositions, this makes it easier to understand why neither volitionality nor intentionality is exactly the right notion for the data discussed in section 2.3.2.1. These data presented a number of cases where an entity is very nearly, but not quite, required to be volitional; a constrained set of non-volitional entities is possible. Since non-volitional entities can have dispositions but not preferences, it makes sense to think of these data as requiring a certain kind of disposition, of which volition is but one example. The last goal of this section is therefore to address how different causal theories treat dispositions.

2.3.2.4 How different causal theories fare on dispositional explanations for volition

This section is interested in how causal theories can help us to understand volition in language, especially in the case of agency. We have argued that, at least in the case of agency, a dispositional view of volition is more successful than a preference view of

[29] Just as Knobe's work is useful for linguists, linguistic theory about intentionality can be useful to philosophers asking Knobe-type questions. Egré (Ch. 8), for instance, takes up a linguistic property (gradability) and applies it to the notion of intention. He argues that Knobe contrasts can be explained by the proposal that agents have degrees of intentionality, where intentionality is further broken down into properties of foreknowledge and desire (cf. again Heim's 1992 denotation of *want*). This is partly based on Tannenbaum et al.'s (2007) finding that the fact that an action was done *intentionally* does not necessarily entail that the agent *had an intention* for the outcome.

volition. We now therefore consider how causal theories bear on the representation of dispositions.[30]

As we have seen, dispositions are states, but—as is evident for the state of wanting—they are states for which certain future possibilities are somehow relevant. The idea is to provide an essentially causal analysis of the relationship between the disposition state and the occurrence that is sometimes—but not always—caused by the disposition.[31]

Note that this means that dispositions are cases of defeasible causation. Since we are analyzing volition in terms of dispositions, this would suggest that all volitional cases should be defeasible.[32] On the other hand, since not all dispositions are volitions, we would expect to see cases of defeasible causation occurring with certain non-volitional cases involving physical dispositions. This is exactly what we see. The presence of a volitional agent seems, on the one hand, to sometimes be linked to the defeasibility of the occurrence of the caused event (Demirdache and Martin 2013). For instance, non-culminating accomplishment readings, for instance, seem at first glance to occur only with animate agents, in the French examples below (Martin and Schäfer 2012; 2013):

(23) a. Pierre l'a provoquée, mais elle n'a même pas réalisé.
Pierre her-has provoked, but she NEG-has even not realized
'Pierre provoked her, but she didn't even realize.'

b. La remarque l'a provoquée, #mais elle n'a même pas réalisé.
The remark her-has provoked, but she NEG-has even not realized
'The remark provoked her, but she didn't even realize.'

Similar generalizations have been noticed for Tagalog (Dell 1987) and Skwxwú7-mesh (Jacobs 2011). However, Martin and Schäfer point out that certain cases do allow inanimate (i.e. necessarily non-volitional) causers to have defeasibility:

(24) a. Cette situation leur a montré le problème, #mais ils ne l'ont pas vu.
This situation them has showed the problem but they NEG it-have not seen
'This situation showed them the problem, but they didn't see it.'

[30] We are not going to do justice to the large literature on disposition in philosophy, although, like causation, this is a topic where linguists and philosophers could learn from each other.

[31] There is not even consensus on whether causation is relevant to dispositions (Choi and Fara 2012). For our linguistic purposes we assume that they are, based on two premises: the fact, discussed above, that linguistic data requires volition in agents to have a similar account as causal properties of certain inanimate causers; and the plausibility of a dispositional analysis of volition in this context.

[32] We should qualify this statement: actually (departing from our assumption that intention and volition are the same), there is a linguistic approach whereby we might view intentions as "total", or "net" volitions, i.e. taking all volitions and all circumstances into account, which means that intentions *do* entail the effect. See Condoravdi and Lauer (2009), Lauer and Condoravdi (2011), Copley (2012). Though, even with intention, some cases are defeasible (Martin and Schäfer 2012). If one *intentionally* performs an action, that action must occur, while if one *intends* to perform the action, that action does not have to occur. See also the previous footnote.

b. Clairement, cette situation leur a bel et bien montré le problème! C'est fou
clearly this situation them has well and truly shown the problem! It's crazy
qu'ils ne l'aient pas vu!
that-they NEG it-have not seen
'Clearly, this situation well and truly showed them the problem! It is crazy that they didn't see it!'

Similarly, non-volitional subjects are consistent with a defeasible result in Finnish morphological causatives (Ilić, Ch. 7, this volume).

(25) Vitsi naura-tt-i minu-a (mutta en nauranut).
joke.NOM laugh-CAUS-PAST I-PART but not.1SG laugh.SG.PAST
'The joke made me feel like laughing (but I did not laugh).'

While all of these facts must be examined in closer detail, they are squarely in keeping with the limited non-volitional exceptions we saw in section 2.3.2.1, supporting the idea that the grammar, in these cases of defeasible causation, cares about dispositions rather than volitions. We may indeed provisionally suppose that defeasible causation might in some or all cases be identified with disposition.

The question now is how causal theories help us understand dispositions, so that we can ultimately understand the reasons for the distributions of agents and causers. Since we are understanding dispositions to be cases of defeasible causation, we might expect the same answer as in section 2.3.1 on defeasible causation: namely, that result-entailing theories of causation (counterfactual, transmission) fare poorly, while non-result-entailing theories (probability, force) fare well.

In fact, however, a popular starting point for philosophers concerned with dispositions is a counterfactual theory (Choi and Fara 2012). A basic counterfactual analysis of dispositions is given in (26a). We can see how dispositions on this analysis might relate to causation by means of Lewis's 1973 counterfactual theory of causation, which gives us the proposition in (26b). Together (26a) and (26b) allow us to conclude (26c).

(26) a. [O is disposed to M when S] iff [if it were that S, then O would M]
b. if [if it were that S, then O would M] then [S causes O to M]
c. [if O is disposed to M when S], then [S causes O to M]

So, for example, supposing a glass is disposed to break when struck, we may conclude that if it were struck, the glass would break, and indeed that striking causes the glass to break.[33] This seems largely correct. Note the similarity between this approach and inertia world solutions to non-culminating accomplishments (section 2.3.1), as well as Martin and Schäfer's stereotypical modal base they use to account

[33] For some reason, *allow* sounds better than *cause* with a volitional subject: if Mary is disposed to eat doughnuts in the presence of doughnuts, then the presence of doughnuts seems to *allow* her to eat doughnuts, rather than to *cause* her to eat doughnuts. This relates to a familiar issue about whether a contributing factor is "the" cause; see Dowty's (1979) revision of Lewis's (1973) theory.

for the facts in (23) and (24). In all these cases, defeasibility is attained by restricting the worlds in which the causal relation holds to a certain subset of possible worlds—the normal, stereotypical, or optimal[34] worlds. These are all thus essentially the same solution to the problem of trying to use a result-entailing theory of causation to account for defeasible causation.[35]

Despite the popularity of this solution among philosophers for dispositions, recall that there were linguistic reasons for dispreferring this kind of semantic solution for non-culminating accomplishments: the lack of scope-taking (Martin and Schäfer 2012) and the cross-linguistic lack of morphology. Now that we are understanding defeasible causation in terms of disposition (both volitional and non-volitional), these reasons should be re-examined.

There are cases of morphology that are likely to denote dispositions: derivational morphology such as *-able* in English, for instance, and even possibly generic/habitual markers (as in Hindi). Thus there may not be such a strong case to make from morphosyntax that a simpler semantics is to be preferred.

On the other hand, in the agent/causer cases we have been discussing, volitional and other dispositional meanings indeed do not seem to take scope over other modals (Martin and Schäfer 2012). We think the lack of scope-taking in agency/causerhood indicates that we should still (just) disfavor complex semantic solutions. If this complexity were taken out of the logical form and put elsewhere, for example in the cognitive system (which would interface with the semantics via axioms of the model), the scope problem would not arise, as desired.

The problems of scope-taking and semantic complexity did not arise for defeasible causation with non-result-entailing theories of causation such as probability and force-based theories. Further consideration of dispositions, however, shows that if these theories are used, they run up against a couple of different problems that would need to be resolved.

We saw that a probability theory of causation, augmented with a distinction between events and event kinds, can account for defeasible causation. However, as far as dispositions are concerned, it may be objected that a disposition should be more than something that raises the probability of the kind of event occurring. One difficult point for dependency theories is that the notion of a disposition typically is thought of as a property of an entity. Dependency approaches such as counterfactual or probability theories, however, do not easily specify whether the likelihood of an outcome is due to internal or external factors, so they are not well suited for

[34] There is a question that arises here (as in discussions of inertia worlds) about how one decides what the normal, stereotypical, or optimal worlds are. It is not trivial.

[35] The remaining result-entailing theories of causation (logical, transmission, some causal models) run into the same issue as the counterfactual theories.

capturing the idea that the disposition is due to properties that hold of an entity. And in fact, a prominent line of criticism of basic counterfactual treatments of dispositions raises exactly this point (see e.g. Martin 1994).

We might expect production theories to fare better in constructing a notion of an internally-generated tendency (or force or energy) because, as we have seen, spatial information is relevant to production theories of causation, while it is irrelevant to dependency theories of causation. Indeed, one of the strengths of transmission theories of causation is that they can specify both the origin and destination of the quantity that is transmitted. As a consequence, such theories can specify whether a conserved quantity emerged from forces internal (e.g. a car's engine) or external (e.g. a leaf falling and being blown about by the wind) to an entity. However, while transmission theories are certainly able to distinguish internal from external influences, they do not provide a motivation for why making such a distinction might be of value. Such motivation is provided in force theories of causation, which, according to White (1999; 2006; 2009) and Wolff and Shepard (2013), are based on our personal experiences of acting on objects and having objects act on our bodies.

But how important is the intrinsic/extrinsic distinction to language? In terms of objective reality, being non-committal about whether the tendency is due to internal or external factors might be a positive feature. Consider, for example, a tendency of physical objects to fall. Objectively, this tendency emerges in part from factors that are external to the object—specifically, the force field created by the earth. Psychologically, however, people typically will attribute this tendency to factors that are internal to the object, a point captured in Aristotle's theory of impetus (Aristotle 1999). To the extent that we want our theories to accurately capture people's representations of entities and their properties, we should prefer theories of causation that allow the internal/external distinction to be made. There is already an argument, for instance, that the internal/external distinction is relevant to the event and argument structure of verbs (Levin and Rappaport Hovav 1995), and it is evident at least that a volitional disposition is conceived of as being proper to the entity whose volition it is. On the other hand, certain dispositions are seemingly not entirely intrinsic: the disposition of a key to open a particular lock, visibility, etc. However, whatever extrinsic relations are involved in such dispositions, they are still based on intrinsic properties (e.g. the shape of the key, the material of which the visible object is constructed).[36]

This makes it seem as though force theories are the winners of this particular round. Yet at first glance it is hard to see how a theory of force dynamics can provide a basis for the intensional nature of dispositions. Talmy (2000), for instance, treats

[36] Molnar (2003) argues that there is no way to remove the requirement for intrinsicality from all dispositional properties.

desires and physical forces as the same.[37] However, desires and physical dispositions are intensional while physical forces are not (cf. e.g. substitutability of co-referring phrases: *This acid changed the color of the litmus paper to red = This acid changed the color of the litmus paper to the color of Post Office boxes*). This is a devastating problem for a Talmy-like system, and it arises because while propositions are crucial for dealing with intensionality, propositions are not particularly well dealt with in systems that address physical forces.[38]

On the other hand, it is possible to borrow the notion of proposition from dependency theories into the spatio-temporal anchoring of force theory. For example, physical forces can be understood as abstract functions from situations to situations (Copley and Harley, to appear),[39] while dispositions can be understood as "second-order forces", i.e. functions from situations to propositions (sets of situations; Copley 2010). The result of a "second-order" force is thus a proposition or set of states of affairs rather than a single state of affairs.[40] In such a way it is possible to integrate forces into a system that does justice to propositions, and hence to the intensionality of dispositions. Likewise, a primitive "dispositional modality" could be understood to be at work (Mumford and Anjum 2011); if modality is taken to involve propositions, this is a very similar solution to the problem.

To summarize: the fact that some inanimate causers can appear in contexts that by and large seem to have only volitional agents indicates that what the grammar cares about in these cases is not volitionality but a causal structure of which volitionality is just one example. That being the case, in studying agents and causers, it behooves us to adopt a theory of causation that can make reference to a notion of volition that has similarities with non-volitional causation. The dispositional approach to volition (e.g. Portner 1997) does better than the preference approach to volition (Heim 1992) on these terms, which (after a detour through research on intention) led us to question which theories of causation can deal with dispositions. Several theories can represent dispositions, but if dispositions are to be understood

[37] On the other hand, philosophical production theories of causation often have a different focus when discussing psychological states in the context of causation: the questions of whether intentions have effects (i.e. whether there is 'mental causation') and whether intentions have causes (i.e. whether there is free will).

[38] See Kistler, Ch. 4, this volume, for related discussion of this point.

[39] As we have seen, this functional treatment of forces also allows a result situation to be the output of the force function; it is thus not existentially bound, and therefore its referent is not asserted to occur. This also yields intensionality, in that various properties can be attributed to that situation without asserting that the situation occurs. However, this move is not really available for *want*, as *want* has a propositional complement. This point underlines the idea that there are currently several different technologies available that can account for defeasibility: not just first- and second-order force functions, but also event kinds, the lack of an efficacy presupposition, scales and less than totally realistic modal bases. All of them apparently model intensionality appropriately. It remains to be seen how many ways to achieve defeasibility and/or intensionality are suggested by the cross-linguistic data.

[40] See also Kamp (1999–2007) for related discussion on this very Davidsonian question of how to link intentions with mental representations of the physical world.

as *intrinsic* tendencies, force-based theories have the upper hand. The intensional nature of dispositions still poses a major problem for force-based theories. It can be overcome, however, by appealing to the idea that a proposition can formally be the result of a more abstract kind of "force".

2.3.3 Representations of causal chains

So far we have investigated how different causal theories account for two related linguistic issues: defeasibility, and the role of volitionality in agency. Our third linguistic issue is the question of how conceptual representations of participants and events in causal chains are mapped onto linguistic representations of these chains. This issue arises both in syntax and in the lexicon, as well as in determinations of the boundary between the syntax and the lexicon (e.g. Folli, Ch. 13, this volume). Since several of the chapters in this volume deal with this issue in detail (Thomason, Wolff, Ilić, Martin and Schäfer, Ramchand, Lyutikova and Tatevosov), we will not do so here. We still want to ask, however, whether the linguistic facts related to the representation of causal chains favor the choice of one causal theory over another, either for use of the theory as a tool in further linguistic analysis or to provide evidence of the theory's cognitive reality.

2.3.3.1 Causal chains in grammatical structures There are several linguistic phenomena that suggest that causal chains—i.e. sequences where more than one causal relation is linked together—are visible to grammar. In other words, more than just the beginning cause and the ending effect of the causal chain are represented in the grammar; intermediate steps are also represented. One case would be the distinction between lexical (e.g. (27a)) and periphrastic (e.g. (27b)) causatives, which have been claimed to differ in terms of whether the causation is "direct" or "indirect" (Fodor 1970; Cruse 1972; Shibatani 1976; Smith 1970).

(27) a. John turned the baby.
b. John made the baby turn.

So, for example, for (27a) to be true, John must physically turn the baby himself (direct causation), whereas for (27b) to most felicitously be true, John does something that causes the baby to turn herself (indirect causation). If the baby turns herself, (27a) cannot be truthfully asserted. The question that this contrast raises is how to best represent the difference between direct and indirect causation, noting that "direct" and "indirect" are intuitive judgments which must be treated as non-linguistic intuitions about the world. One general approach is to define direct causation in terms of how causal chains in the world are conceptualized, and to put a restriction on lexical causatives such that they can only be used to describe causal relations that lack intermediaries. Causal relations in which the cause and effect are mediated through some intermediate causal agent would thus need to be described using periphrastic causatives.

Futurates provide a similar example of indirect causation (Copley 2014). On this view, a futurate sentence such as *Mary is building a house next year* makes reference not only to the subeventuality of which Mary is the agent and the result eventuality of the existence of the house, but also to a subeventuality corresponding to the volition of the entity causing Mary to build the house (the "director"; possibly but not necessarily Mary). This extra eventuality can be modified by a temporal adverbial, as shown in (28a) (Prince 1973) and (26b).

(28) a. Yesterday morning I was leaving tomorrow on the Midnight Special.
b. For a moment Mary was building a house next year.

Copley and Harley (2009; 2010) point out that the semantics of futurates is very similar to the semantics of *have* causatives.

These phenomena might make us wonder whether the intermediate step strictly needs to involve an agent if more than one causal relation is to be licensed in the syntactic structure (Cruse 1972). This seems to be the case, but "agent" must, as usual, be understood in terms of teleological capability. Even though a non-teleological entity might *conceptually* be seen as an intermediate causer, as in a situation where John pushes a blue marble and the blue marble's motion causes the green marble to move (29),[41] the blue marble is not felicitous as an intermediate agent. Rather, a teleologically capable entity is needed, such as a machine.

(29) a. #John made the blue marble make the green marble move.
b. John made/had the machine move the green marble.

Going just by (29), then, a teleologically capable intermediate agent might seem to be necessary in order to represent causal chains. However, what (29) shows is merely that an agent seems to be necessary for periphrastic or indirect causation. There is another way in which multiple causal relations seem to be visible to the grammar, namely by the representation of an instrument in a prepositional phrase. Croft (1988), for instance, presents instruments as intermediate entities in the causal chain between an agent and a patient. There are indeed differences between cases with intermediate agents and cases with instruments: unlike intermediate agents, instruments don't give the causation an "indirect" feel and is an enabler rather than a causer. Nevertheless, the fact that they are represented seems to indicate that the grammar has access to the intermediary causal participation of the instrument.[42]

Given that there is evidence that the grammar appears to be sensitive to the presence of multiple causal relations, the question arises as to how different accounts of causation might handle the representation of causal chains. In the

[41] Note, however, that it is arguable whether the blue marble can be seen as an intermediate causer even conceptually. In the paraphrase it is the blue marble's *motion* that makes the green marble move, rather than the blue marble itself.

[42] See also Thomason, Ch. 3, this volume for discussion of this point.

philosophical literature, the need for chains has been motivated by the belief that causal relations can be derived from transitive reasoning (Hall 2004): if A causes B and B causes C, we can infer that A causes C. As recognized by Lewis (1973), causal relations derived from transitive reasoning can sometimes be problematic for a counterfactual analysis of causation. Such problems arise in cases of late pre-emption. Consider, again, the case of Suzy and Billy; each throws a rock at a bottle, but Suzy's rock hits the bottle first and shatters it. As noted by Lewis (1973) (see section 2.2.1.1), we would say that Suzy caused the shattering of the bottle. The problem for counterfactual theories is that the the bottle's shattering does not depend counterfactually on Suzy; in Lewis's (1973) words, the causal dependence here is "intransitive". The problem for a counterfactual analysis is that there can exist a causal relation without causal dependence.

To get around this problem, Lewis proposes that the link between Suzy and the bottle's shattering can be decomposed into a series of steps or links in a causal chain. He argues that while there may not be causal dependence between the first and last event in the chain, causal dependence can hold between individual steps in the chain, and the presence of these causal dependencies can license the inference that there is a causal relation between the first and last events in the chain, i.e. they can license the transitive inference that the first and last events of the chain are connected by a causal relation. While this possible solution looks at first like a viable fix to the theory, even Lewis (1973) recognized that it is vulnerable to the criticism that in the cases of late pre-emption—like the Suzy and Billy case—counterfactual dependence will not hold for at least one of the intermediate steps. This problem led to a revised version of his theory in which counterfactual dependency is said to exist when an alteration in the cause leads to an alteration in the effect (Lewis 2000). However, as we have seen, this new version of the theory brings with it its own set of problems. As noted by Menzies (2008), if Suzy's throw is altered such that she throws after Billy, Suzy might no longer be considered the cause of the shattering, but it would still be the case that an alteration to her actions led to an alteration in the final effect, which according to Lewis's (2000) new criterion, would make Suzy a cause. In sum, a counterfactual approach to causal chains, in the style of Lewis's theories of causation, is fraught with problems and is not likely to be able to serve as a theory of how people represent causal chains.

While counterfactual theories of causation likely cannot serve as theories of how people represent causal chains, other kinds of dependency theories hold greater promise.[43] As discussed earlier, a Bayesian network approach to causation is by design specifically formulated for the representation of causal chains and causal networks. As we saw, such an approach is able to handle problematic situations, such as late pre-emption, by representing these situations in terms of a

[43] Aside from logical dependency theories, which do not do so well even at representing single causal relations, as noted in section 2.2.1.1 with respect to late pre-emption.

network rather than in terms of a chain. As described in Hitchcock (2010), in a network representing late pre-emption, Suzy's throw leads to the hitting of the bottle, which not only leads to the bottle's shattering but also prevents Billy's throw from hitting the bottle, and hence blocks Billy's throw from being the cause of the shattering.

Causal networks raise an interesting issue for the relationship between theories of causation and linguistic theory. The solution to late pre-emption offered by Hitchcock involves a branching causal chain, but according to some linguists, language only encodes non-branching causal chains (Croft 1988; Talmy 1983). If certain kinds of causal relationship require representations specifying causal networks, complications arise with regard to the relationship between linguistic and conceptual representations. It may be that for the sake of language, conceptual representations specifying causal networks are, in some sense, reduced, or simplified into causal chains. Another possibility is that, contrary to Croft and Talmy, representations in language might directly specify causal networks. Yet another possibility is that the nature of the representations specifying causation do not specify causal networks, even in the case of late pre-emption.

According to Hall (2004), production theories of causation are well suited for the representation of causal chains. Unlike dependency theories, production theories do not need to use a network to handle late pre-emption (see Walsh and Sloman 2011). As we saw in section 2.2.2, Suzy's throw causes the bottle to break because she imparts its force or energy to her rock, which imparts its force or energy to the bottle, breaking it. There is no need to consider Billy's throw. This means that Croft's (1988) principle that linguistically represented causal chains are non-branching can be maintained.

One problem, however, with using production theories of causation to represent syntactic causal chains lies in the causal relata they assume. Production theories typically treat causation as relating entities to one another, in terms of how entities exert force on or transfer energy to one another (see Wolff, Ch. 5, this volume; also very much like cognitive linguistic literature on the topic, esp. Croft 1988). Using entities as the causal relata is inconvenient, given that causal chains in syntax are generally assumed to have events as the causal relata. Bayesian theories are inconvenient too, in that they treat the causal relata as variables which have true or false outcomes (with a given level of certainty), which seems to treat them as propositions.[44] For this reason, counterfactual theories might be preferable, since they take the causal relata to be events. However, even counterfactual theories ultimately define causation in terms of propositions—namely, the proposition that the causing event occurs and the proposition that the result event occurs—and it might be questioned whether propositions as such are really relevant to meaning that is

[44] See Thomason (Ch. 3) for a critical discussion of Dowty's (1979) use of propositions as causal relata.

located so low in the verb phrase. One alternative would be to rescue the production perspective by using a more abstract notion of force that acts on situations rather than entities, as do Copley and Harley (Ch. 6, this volume; to appear). The similarities between events and situations (Kratzer 2011) mean that this move addresses the basic problem. Another solution could come from Lyutikova and Tatevosov (Ch. 11, this volume), in which the causal relata are essentially properties of events. Properties could arguably stand in for propositions within the verb phrase in such a way as to make the dependency theories more plausible.

2.3.3.2 Sometimes causal chains are not important to linguistic representations We have argued that conceptual causal chains are sometimes visible to grammar, and that this point suggests that some causal theories—production theories and possibly causal networking theories—would be more useful than others in accounting for the representation of causation in language. However, quite to the contrary, there are also cases where a conceptual causal chain has intermediate links that are not represented in language; the grammar is apparently indifferent to anything but the beginning and end of the conceptual causal chain.

One such case, as is noted in Thomason's and Ramchand's chapters in this volume, is what we might call "conceptual-linguistic causal mismatch": the fact that virtually any causal relation can be conceptualized as involving an intermediary (Pinker 1989). For instance, lexical causatives can have unexpressed instruments or body parts; if John broke the table, he did so either with an instrument, or with his hand, or another part of his body (Guéron 2005; 2006). Volition itself can be thought of as an extra initial step in the conceptual causal chain (Davidson 1963; Talmy 2000; Ramchand 2008; Ch. 10, this volume; Copley and Harley, Ch. 6, this volume); theories differ as to whether the volition of an agent is represented in the syntax, but if it is not, then it provides another example of a complex conceptual causal chain but a simple morphosyntax. To take another example, it is clear from Sichel's (2010) nominalization data presented in (17) that the nominalization *justification* places a requirement of more direct participation on the subject than does the verbal form *justify*; the possibility of having *the hurricane* as a subject of *justify*, but not of *justification*, indicates that the former, but not the latter, must allow a conceptually complex causal chain. So even though we perceive *justify* to be more "direct" than, e.g. *cause to justify*, this intuition of directness does not necessarily correspond to a simple (one-link) conceptual causal chain, because the nominalization has a requirement for an even simpler causal chain.

So what is "direct causation" as we perceive it in lexical causatives? In Ramchand's chapter, she argues that this direct causation has to do with lexicalization of two heads with one morpheme. On the other hand, for Lyutikova and Tatevosov (Ch. 11, this volume), the directness/indirectness distinction stems from different kinds of causal/temporal structure, as represented in the semantics. If the latter, production

theories should be preferred again as they necessarily represent spatio-temporal information, while dependency theories need it to be added on.

A better case for dependency theories can be made if we consider lexical connectives such as *because (of)*, *since*, *as a result of*. *Because of*, for instance, is apparently not sensitive to how far away in the causal chain an agent is, or if there are any intermediate agents or causers in the chain. The sentence in (30a) is felicitous and true even if Mary opens the window and the wind blows the door open, or if she tells John to open the door and he complies; in these cases (30b) is not. And when Mary opens the door herself, (30a) is still felicitous and true.

(30) a. The door opened because of (what) Mary (did).
b. Mary opened the door.

Because, unlike the verb *cause*, is also indifferent to the type of causal relationship. As noted by Talmy, the notion of cause seems to be a family of notions that includes CAUSE (in the narrow sense), ENABLE/ALLOW, PREVENT, and DESPITE. These different notions of causation are often differentiated in verb meaning. In causal connectives, in contrast, they are generally not differentiated. For example, it has been observed that *because* can refer to enabling conditions instead of a single most important cause, unlike the main verb *cause*. This is demonstrated by the fact that the sentence in (31a) with *because* is true, because drugs were an enabling condition of Armstrong's seven victories (though not "the" cause), while a similar sentence with *cause* in (31b) is false.

(31) a. Lance Armstrong won seven Tours de France because he took drugs.
b. Drugs caused Lance Armstrong to win seven Tours de France.

The existence of a connective like *because* might be slightly unexpected under a theory of causation where causing and enabling are necessarily represented differently. The surprise increases once we consider that it is likely that all causal connectives are apparently indifferent to the distinction between CAUSE and ENABLE (Wolff et al. 2005).[45]

Some production theories necessarily distinguish CAUSE and ENABLE. For example, in Talmy's (1988) theory of force dynamics, CAUSE is reflected by an agent vector opposing the patient vector, while ENABLE is represented by the removal of a preventive force (see also Wolff, Ch. 5, and Ilić, Ch. 7, this volume). So the nonexistence of causal connectives that differentiate these notions is surprising in such a theory. In dependency theories, however, causes and enabling conditions are represented similarly (but see Sloman et al. 2009). Dowty (1979) and Eckardt (2000) make this point for Lewis's (1973) theory, for instance, and in Bayesian causal models, the arrows can represent either causing or enabling

[45] Some verbs also are vague in this way, e.g. *influence*, *affect*, *lead to*, *link to*.

relations. So on a dependency theory of causation, the fact that causal connectives do not make the distinction is expected.

2.4 Conclusion

We have argued that certain theories of causation are more suited to certain sets of linguistic data and should therefore be utilized in developing theories of these data. This story we have been telling is thus a familiar Occam's Razor story, but the appeal to Occam's Razor is meant to be understood in a quite practical and preliminary way: we have not argued that in every case the most plausible theory of causation is the correct one. Rather, we suggest simply that one needs to be aware of the range of theories of causation while collecting, organizing, and analyzing data. Questions to be asked are: how does a particular theory of causation help in formulating linguistic hypotheses and the linguistic theories that arise from them? How does a particular theory of causation tell us which things language might care about? Does the causal theory allow us to make the semantics simple by putting any meaning in the conceptual system instead of having to represent it in (syntax-visible) semantics, in places where it seems as though the syntax cannot see that meaning? If we can investigate the possibilities afforded by different theories of causation for the data we are interested in, the choice will likely lead to different and (dare we say) better explanations, as well as interesting avenues for further research. Of course it must be emphasized that even a theory of causation that is not particularly well suited to a particular set of data still can be adequate, and could even be right. We have surely not thought of every way in which every theory could be extended. Still, if theories are to be compared, the burden of proof is, as always, on whichever theory is less obviously suitable for the data.

Much of this chapter has focused on how theories of causation might inform linguistic theory. However, the potential benefits of this conversation also extend in the other direction. To the extent that philosophers and cognitive scientists use language-related data to support claims about cognition, such research will benefit from insights from linguistic theory. Linguistic theory reveals syntactic and semantic distinctions that are likely to reflect conceptually significant categories. Consider, for example, the Minimalist Program (Chomsky 1995) principle that the properties of language depend in part on the need for language to interface with phonology on one hand and the conceptual system on the other. (Cognitive linguistics, true to its name, has always viewed representations as having cognitive reality.) Any philosopher or psychologist who last looked at linguistics as recently as the early 1990s and decided it was not germane to their research should know that the development of Minimalism, among other advances, has made syntactic theory both more cognitively plausible and more amenable to questions of meaning.

We have made no claims about the particulars of the interface between language and mind in this chapter, but predictions should emerge quite naturally from the juxtaposition of theories of causation, linguistic theory, and data. For example, we have observed an interesting relationship between structural height and the specificity of the causal expression. We have seen that volition, temporal-spatial considerations, and complex causal chains are relevant to the semantics of the verb phrase, which tends to point toward a production theory of causation. We have also seen, however, that it is less clear that production theories are relevant to higher structural domains, as in lexical connectives like *because* which occur outside the verb phrase and take propositions as their arguments. In fact, the lack of specificity of these expressions seems to indicate that a dependency theory would be preferable higher in the structure. Could this structural disparity reflect a causally pluralistic mental representation, where production theories are relevant to the lower part of the tree, and dependency theories are relevant to the higher part of the tree? Might linguistic structure even suggest a way to relate the two kinds of theory to each other or define one in terms of the other?

We do not yet know the answers to these speculative questions. But we hope that linguists, philosophers, and cognitive scientists will see fit to ask and answer such questions together. Although interdisciplinary ventures are never automatically fruitful merely by virtue of being interdisciplinary, we think that a cognitive and linguistic conversation on causation is now possible, and that this conversation is likely to advance the long-term goal of integrating linguistic theory with the science of the mind.