CHAPTER THREE

# Hybrid Causal Representations

**Michael R. Waldmann**[1] **and Ralf Mayrhofer**
University of Göttingen, Göttingen, Germany
[1]Corresponding author: E-mail: michael.waldmann@bio.uni-goettingen.de

## Contents



## Abstract

The main goal of this chapter is to defend a new view on causal reasoning, a hybrid representation account. In both psychology and philosophy, different frameworks of causal reasoning compete, each endowed with its distinctive strengths and weaknesses and its preferred domains of application. Three frameworks are presented that either focus on dependencies, dispositions, or processes. Our main claim is that despite the beauty of a parsimonious unitary account, there is little reason to assume

*Psychology of Learning and Motivation*, Volume 65
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2016.04.001

that people are restricted to one type of representation of causal scenarios. In contrast to causal pluralism, which postulates the coexistence of different representations in causal reasoning, our aim is to show that competing representations do not only coexist, they can also actively influence each other. In three empirical case studies, we demonstrate how causal dependency, causal dispositional, and causal process representations mutually interact in generating complex representations driving causal inferences.

## 1. INTRODUCTION

Causal reasoning is one of our most central cognitive competencies, enabling us to adapt to our world. The ubiquity of causal reasoning has attracted researchers from various disciplines to this topic. Philosophers have studied causality for centuries, but more recently the topic has also motivated research in the fields of psychology, economics, biology, physics, anthropology, statistics, and artificial intelligence, to name just a few. Thus, causality is a genuinely interdisciplinary topic attracting both researchers interested in developing *normative methods* of causal discovery, and researchers pursuing the *descriptive* goal to capture how humans and non-human animals actually reason about causal relations (see Waldmann, in press; Waldmann & Hagmayer, 2013, for overviews).

Most theories of causal reasoning proposed in psychology have precursors in philosophy and other normative disciplines (see Waldmann & Hagmayer, 2013). Although the research goals of normative and descriptive theories differ, it is not an accident that the theories overlap. Both scientists and laypeople develop causal hypotheses that they intend to be correct. Thus, causal claims typically are associated with normative force (see Spohn, 2002; Waldmann, 2011). This commonality may be the reason why psychologists often turn to normative theories as an inspiration for psychological accounts. An examples of this long tradition are causal Bayes nets that have first been developed in philosophy and engineering (see Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 2000) but have also been adopted by psychologists as models of everyday causal reasoning (see Rottman & Hastie, 2014; Waldmann, 2016; Waldmann & Hagmayer, 2013, for reviews).

Despite the common goals of scientists and laypeople, however, it is implausible to expect that (descriptive) psychological accounts will exactly mimic normative theories that were developed for scientists to guide research in their specific domain. Causal domains substantially differ so that a method that has been developed for economics and sociology will differ from methods suitable for research in physics. By contrast, laypeople

use causal knowledge in various everyday domains including intuitive physics, biology, psychology, or sociology. Also, unlike scientists, they typically have little knowledge about the mechanisms governing these domains (see Rozenblit & Keil, 2002).

Another difference between normative and descriptive approaches is that philosophers and scientists interested in methodology generally try to develop a uniform coherent account that is grounded in few basic principles. Coherence, consistency, simplicity, and parsimony belong to the key qualities that researchers try to accomplish. By contrast, laypeople are often satisficers. They use methodological tools that work for a given problem but they often care little about overall coherence and consistency (see Arkes, Gigerenzer, & Hertwig, 2016).

A sign of the plurality of causal concepts in everyday thinking is that in psychology different frameworks and theories of causal reasoning compete. These frameworks and theories differ in terms of how they model causality and causal reasoning. We will use the term framework to describe classes of theories that use substantially different theoretical concepts to capture causality. They also often differ in the tasks they are trying to model. Within each framework there are various theories competing for the best explanation of the tasks addressed by the framework.

In the next section, we will briefly describe the main assumptions of different competing frameworks of causal reasoning. Then, we will elaborate our main claim that in everyday causal reasoning people simultaneously use multiple mutually interacting representations that can be grounded in the different frameworks of causality. These so-called "hybrid" causal representations may often lack overall consistency and parsimony but they may better capture reasoning in everyday contexts than approaches that strictly follow the regulations of axiomatized normative theories. In contrast to pluralistic views according to which different representations are independently used in different contexts, we argue that different causal representations constrain each other in a given reasoning context and that such hybrid representations are at least locally consistent. We present three empirical case studies that bolster our claims.

## 2. FRAMEWORKS OF CAUSAL REASONING

In this section, we describe different frameworks of causal reasoning that originally have been inspired by different philosophical accounts.

Each of the frameworks comes in numerous variants both in philosophy and psychology. We focus here on the prototypic features and only briefly point to variations. The main distinguishing features of these frameworks, which we discuss in the following sections, are the proposed *causal relata* (ie, the type of entities that enter causal relations) and the *causal relations* that are used to represent causal scenarios.

## 2.1 The Dependency Framework

The dependency view of causation is shared by several psychological theories that otherwise compete with each other, including associative theories (see Le Pelley, Griffiths, & Beesley, in press), covariation theories (eg, Cheng & Novick, 1992; Perales, Catena, Cándido, & Maldonado, in press), power PC theory (Cheng, 1997), causal model theories (eg, Gopnik et al., 2004; Rehder, in press; Rottman, in press; Sloman, 2005; Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne., 1995), and Bayesian inference theories (Griffiths & Tenenbaum, 2005, 2009; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008; Meder, Mayrhofer, & Waldmann, 2014; for overviews see Holyoak & Cheng, 2011; Waldmann, 2016; Waldmann & Hagmayer, 2013).

According to dependency theories, a variable $C$ is a cause of its effect $E$ if variable $E$ depends upon $C$. There is an extensive debate in philosophy about the proper *causal relata* in dependency theories (eg, events, propositions, facts, properties, or states of affairs; see Ehring, 2009; Spohn, 2012). For our purposes, however, it is sufficient to adopt the terminology of causal model theory (eg, structural equations and causal Bayes nets; see Halpern & Hitchcock, 2015), according to which the world can be properly represented in terms of random variables (and their values) and the dependencies between them. In causal model theories, *causal relations* are graphically depicted by causal arrows that are directed from cause to effect (see Fig. 1). For example, a causal model could be postulated that uses the binary variables representing the effect forest fire (present vs. absent) and the potential causes match (eg, dropped by an arsonist vs. not dropped) and lightning

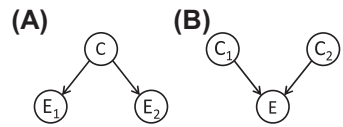


**Figure 1** An example of a common-cause structure (A) with a cause variable $C$ and two effect variables $E_1$ and $E_2$, and a common-effect structure (B) with two cause variables $C_1$ and $C_2$, and an effect variable $E$.

(present vs. absent) (see the common-effect model in Fig. 1B). The dependencies, then, encode a set of hypothetical situations consistent with the causal model. To describe an actual case of causation, the variables are instantiated (eg, match dropped, lightning absent, and fire present). All theories use some kind of statistical measure of covariation to describe the strength of these relations although they differ in terms of how these statistical measures are interpreted.

A useful distinction is to separate sample-based and model-based dependency theories (see Dwyer & Waldmann, in press; Griffiths & Tenenbaum, 2005; Meder et al., 2014). Sample-based theories assume that the observed covariation in a sample is a direct measure of causation. In the binary case, causes raise or lower the probability of their effects. Sometimes temporal order is added as a cue that helps distinguishing between cause and effect (ie, causes precede effects). Examples for these theories are associative theories or traditional probabilistic theories (see Le Pelley et al., in press; Perales et al., in press, for overviews). A more recent development separates the observed sample from the underlying causal structure that presumably generated the observed covariations. According to this view, observed data are used to make inferences about the hypothetical generating causal structure, for example about unobservable causal powers or whether or not there is a causal relation between two variables (eg, Cheng, 1997; Griffiths & Tenenbaum, 2005; Lu et al., 2008; Meder et al., 2014; Pearl, 2000). Causal directionality is a feature of the underlying hidden causal structure because the observed covariations are symmetric. Therefore, different proposals have been made about how to identify causal direction, including a recourse to temporal order (Johnson-Laird & Khemlani, in press; Spohn, 2012), counterfactuals (Lewis, 1973), hidden mechanisms (Pearl, 2000), or hypothetical interventions (Halpern & Hitchcock, 2015; Spirtes et al., 2000; Woodward, 2003).

Causal models also allow for a representation of mechanisms which within this framework are conceived as chains or networks of interconnected interdependent variables (see later sections for different views on mechanisms). For example, the covariation between smoking and lung cancer can be further elaborated by specifying intermediate variables, such as genetic alterations caused by the inhalation of carcinogenic substances.

Causal model theories are particularly good at explaining how people make statistical inferences from observed causes to effects (predictive reasoning) or from observed effects to probable causes (diagnostic reasoning; see Fernbach, Darlow, & Sloman, 2011; Meder et al., 2014; Meder & Mayrhofer, in press; Waldmann & Holyoak, 1992; Waldmann, 2000).

They can also capture teleological explanations (Lombrozo, 2010) and planning processes (Pearl, 2000). One particularly important feature that sets them apart from purely associative accounts is their capability to predict the outcomes of hypothetical interventions when only observational knowledge is available (Meder, Hagmayer, & Waldmann, 2008, 2009; Pearl, 2000; Spirtes et al., 2001; Waldmann & Hagmayer, 2005). Finally, an important feature of these theories is that they can be easily applied to the question how people learn and acquire causal representations through covariation learning.

Although various psychological studies have shown that causal model theories capture the key features of human causal reasoning well, there are also important deviations from the normative model, some of which are addressed later (see Rehder, 2014, in press; Rottman & Hastie, 2014; Rottman, in press; Waldmann & Hagmayer, 2013, for overviews).

An important distinguishing feature between frameworks are the tasks they use in experimental research. The fact that dependency theories focus on causal dependencies between variables is also manifest in the typical empirical research paradigms. In experiments, causal information is presented in terms of described (eg, Ali, Chater, & Oaksford, 2011; Fernbach et al., 2011; Rehder, 2014) or experienced (eg, Gopnik et al., 2004; Waldmann, 2000) covariations between causal variables that represent events. Typical examples of cover stories are scenarios that describe medicines causing headache (eg, Buehner, Cheng, & Clifford, 2003), foods causing allergies (eg, Shanks & Darby, 1998), chemicals or radiation causing the expression of genes or diseases (eg, Griffiths & Tenenbaum, 2005; Perales, Catena, & Maldonado, 2004), or fertilizers causing plants to bloom (eg, Lien & Cheng, 2000).

## 2.2 The Disposition Framework

A completely different view, which can be traced back to Aristotle's treatment of causation (see Kistler & Gnassounou, 2007), addresses the question why an observed lawful relation holds by focusing on the participants involved in a causal interaction; for example, the two colliding balls in Michotte's (1963) task or aspirin and a person with a headache in a medical scenario. A dispositional account of causation would say that the ingestion of aspirin relieves headache because aspirin has an intrinsic property, a disposition (or capacity, potentiality, power), to relieve headaches in suitable organisms, which interacts with the disposition of human bodies to be influenced by aspirin. According to this view, dependency relations are secondary; they arise as a product of the interplay of objects that are endowed with causal dispositions.

Thus, one important difference between dependency and dispositional theories concerns the *causal relata*. Whereas dependency theories focus on variables that, for instance, encode the presence or absence of events, dispositional theories use *objects* as primary entities. These objects can refer to both humans or nonhuman entities (eg, aspirin). The dispositions of objects can be static (eg, solubility of sugar) or they can be transient and dynamic such as the sudden exertion of force when pushing an object. Causal relations are not represented as dependency relations between variables or events but are situations that arise when objects are placed in specific situational contexts allowing them to express their powers. For example, neither aspirin nor the person suffering from headache are per se cause and effect. Only when placed in the right context (eg, aspirin being ingested by the body of a person), the observed causal relation between events arises (eg, relieving headache). Thus, the lawful relations between events that are the focus of dependency theories are actually secondary here; they arise because dispositional properties of objects generate them. In this way, dispositional views are looking for deeper explanations of observed dependencies underlying the observed covariation. One can see this as a focus on underlying mechanisms; however, the mechanisms have different properties from mechanisms modeled within the dependency framework (eg, as chains or networks of variables).

Different theories within the dispositional framework vary with respect to the abstractness of the postulated dispositions and object types. Some theories just distinguish between two classes of objects, for example causal *agents* and causal *patients*, others use more elaborate characterizations of dispositional properties.

A popular theory, especially in linguistics, is *force dynamics*. This theory has initially been developed and empirically tested in the context of verb semantics and uses fairly abstract characterizations (see Gärdenfors, 2014; Levin & Rappaport Hovav, 2005; Riemer, 2010; Talmy, 1988). Theories of force dynamics typically assume that in a specific causal interaction there are two types of entities, which have been labeled differently, but which we will call causal agent and causal patient (for short: *agent* and *patient*). This distinction between agents and patients can be traced back to Aristotle who explained efficient causation as a consequence of the interaction of these two entities. Talmy (1988), who invented the theory of force dynamics, uses the terms agonist and antagonist to describe the relevant objects. Talmy argues that intuitions about the interaction of causal forces are an important component of our general semantic intuitions.

Gärdenfors (2014) describes a patient as an animate or inanimate, concrete or abstract object that is acted on by causal agent. It can carry a counterforce resisting the action of the agent, which is the more active object that generates a force on the patient, either directly or indirectly via an instrument. The agent force represents the action of the agent. Forces are primarily physical but they can be extended metaphorically to social or mental forces (eg, threats, commands, and persuasions). Gärdenfors uses this framework to represent events and causation. In his two-vector model of a causal event, an agent exerts a force on a patient which leads to a result vector (eg, a movement of the patient). Like forces in general, the result vector need not be restricted to physical changes. Changes regarding other properties of the patients can also be represented.

Force dynamics has been used in linguistics to characterize verb semantics and argument structure. In these theories, verbs place constraints on the possible objects mentioned in the noun phrases. For example, in "Peter pushes Mary," "push" has two arguments, one describing an agent (Peter), the other the patient (Mary). Typically, agents are assigned the syntactic subject position.

The psychological sibling of linguistic versions of force dynamics, Wolff's (2007) force theory (later called dynamics model; Wolff, Barbey, & Hausknecht, 2010), initially aimed at elucidating our understanding of abstract causal concepts, such as *cause*, *prevent*, *enable*, and *despite* (see also Wolff, 2012; Wolff & Song, 2003). Later it has been extended to describe representations of specific visually or linguistically conveyed scenarios. Force theory states that people evaluate configurations of forces attached to affectors (ie, agents) and patients, which may vary in direction and degree, with respect to an end state. As in Gärdenfors' (2014) theory, forces are abstract representations and can describe physical, social, or psychological causal influences. Causal events are analyzed in terms of three components: (1) the prior tendency of a patient toward the end state, (2) the concordance between agent and patient force, and (3) whether the end state is reached or not. For example, a scenario in which the patient does not have a tendency toward the end state (eg, a boat standing still in the middle of a lake) and the affector force (eg, wind) is directed toward an end state that is eventually reached would be construed as a case of *cause* (ie, "The wind caused the boat to reach the harbor.").

While theories of force dynamics have primarily been developed in linguistics from where they were imported into psychology, philosophers have independently developed related kinds of dispositional theories. Unlike

psychologists, who are mainly interested in how people represent causality regardless of the correctness of their beliefs, philosophers endorsing *dispositionalism* try to develop a normative account. Therefore, philosophical theories use more elaborate characterizations of dispositional properties and do not restrict their theories to just two types of entities, agents and patients. For example, the philosopher Nancy Cartwright (1999) has proposed that observed lawful regularities (dependencies in our terminology) can only be understood if they are analyzed as arising from abstract or concrete "nomological machines" in which parts with attached causal powers are arranged in some spatiotemporal arrangement, which when put in the right constellation give rise to observed regularities. Cartwright discusses various examples of nomological machines, which range from abstract constellations (economics, planetary movements) to specific devices (pendulum, toilet cistern) (see also Cartwright & Pemberton, 2012).

One specific model of a dispositional theory of causation, which we have adopted in our second case study (see Section 4.2), is the *vector model* of Mumford and Anjum (2011). In their view, causation is a relation between properties of objects (see Fig. 6A, for an example). When a bag of apples on some weighing scale moves the pointer of the scale, it is the property of weight that does the causal work. Properties are, in the theory of Mumford and Anjum, clusters of powers that dispose objects in specific directions. For example, fire has the power or disposition to warm nearby things. Dispositions can be silent (eg, solubility) until put into the right circumstances; causation occurs when the dispositions manifest themselves.

Mumford and Anjum (2011) formalize this general idea using vector diagrams. The vector diagrams represent a specific moment of a causal situation. The various powers in operation at a particular moment can be represented as a bunch of vectors within a (multidimensional) quality space. In a simple case, the quality space is one-dimensional running from one extreme (eg, hot) to the opposite (cold). The quality space has a vertical line in the middle, which represents the momentary state of the situation with respect to some variable of interest, for example, the temperature of a room. Attached to this line are various vectors representing powers that dispose the situation in different directions. These vectors can vary in direction and length with length representing their intensity. For example, a fire strongly disposes a room toward warmth, whereas a simultaneously present open window may dispose it toward a colder state. Each situation therefore can be represented as a large set of vectors that represent powers in different directions and strengths. The authors suggest that the powers can be added

up leading to a resultant vector representing the overall causal disposition of the situation.

Another example for the vector model would be a situation with a one-dimensional quality space representing bodily health versus disease. Various factors, for example, lack of sleep, stress, and genetic dispositions, may represent powers pointing toward disease, whereas the ingestion of drugs and sunlight might represent countervailing powers. Thus, each situation needs to be characterized by a very large number of powers. This view contrasts with the typical analysis within dependency frameworks in which often only few causes are listed to explain an effect. The vector model captures causal changes well in which continuous properties change, such as heat. To be able to also explain causation with binary effects that can either be present or absent, Mumford and Anjum extended the vector model by adding a threshold that a resultant vector would have to pass to become visible as an effect (see Fig. 6A).

In psychology, the tasks studied to test dispositional theories differ from the ones used to test dependency theories. Psychological research on dispositional theories focuses on language or perceptual scenes as target domains. Tasks are presented that activate already present causal knowledge. Learning has not been formally addressed within this framework. Moreover, the causal scenarios that are typically studied are fairly simple. One reason for this limitation may be that verbs in most cases just involve one agent and one patient. There are studies on causal chains (Wolff et al., 2010) but inferences in other more complex causal models with multiple causal relations have not been studied yet. The main goal has been to study how people understand causal scenarios rather than modeling complex predictive or diagnostic inferences or learning.

## 2.3 The Process Framework

A third class of theories holds the assumption that causation cannot be understood as a relation between events or objects, but arises from continuous processes and interactions between processes. According to Salmon (1984) a process is anything with structure over time. A key issue is how causal processes can be distinguished from non-causal time lines. For example, atoms decaying or billiard balls moving across the table are examples of causal processes, whereas moving shadows or spots of lights are pseudo-processes according to this view. The core idea of process theories is that causation involves some kind of transfer of quantity from cause to effect. Most accounts are restricted to physical causation and turn to physics to identify

the right kind of quantity that is being propagated (see Paul & Hall, 2013). Fair (1979) suggests energy, while Salmon (1984) and Dowe (2000) propose that any kind of conserved quantity (eg, linear momentum, charge) is transmitted.

According to the process framework, causal processes are the primary basis of causation, whereas events are secondary abstractions of the underlying processes. Thus, whereas within the dependency view (eg, Bayes nets) mechanisms are represented as chains of events, process theories would view these chains as abstractions over causal processes that determine whether a chain of events is causal or spurious.

So far process theories are of limited value for psychology because most laypeople do not have deep knowledge about physics. Moreover, the theories seem to be restricted to physical domains, it is unclear how they would model causal reasoning in other domains, such as psychology, sociology, economics, or biology. However, these accounts do capture the intuition of people that some kind of hidden process seems to link causes and effects, for example when we observe billiard balls hitting each other (Michotte, 1963). Moreover, they could provide an account for why we often do not consider all dependency relations causal, for example the covariation between spuriously related events (eg, barometer and weather) or relations between omissions and outcomes. We do not, for example, consider it a cause of the drying of Putin's lawn that we did not water it. However, there are cases in which we do consider omissions to be causal (see Lombrozo, 2010) which led to extensions of process theories requiring the addition of counterfactual reasoning elements to account for these findings (see Dowe, 2000).

## 3. HYBRID CAUSAL REPRESENTATIONS

In Section 2, different frameworks of causal reasoning have been presented. They differ in terms of the *causal relata* they invoke and the way causal relations are construed. Moreover, these architectural differences are tied to specific kinds of tasks that each framework favors to support its theory. For example, dependency theories often are tested by using learning data presenting causal variables (eg, contingency tables), whereas dispositional theories are mostly studied presenting linguistic phrases or perceptual scenarios about interacting objects. Moreover, the frameworks differ in terms of the inferences that can easily be modeled. Dependency theories

are designed to explain learning and predictive and diagnostic inferences within causal models, whereas dispositional theories focus on causal understanding and the semantic parsing of causal scenes.

## 3.1 Unitary Versus Pluralistic Causal Theories

One clear evidence for the division of labor between frameworks is that they hardly ever are applied to the tasks of the competitor. For example, psycholinguists generally do not use Bayes nets, whereas dispositional theories are usually not applied to causal learning tasks.

Nevertheless, there are attempts to defend a unitary causal account against the threat of overly flexible multisystem accounts that often can be too easily adapted to whatever finding comes along. Unitary theories are attractive because they promise a maximum of coherence and consistency. Different attempts to explain tasks studied by competing frameworks have therefore been made. For example, Cheng (1993) has applied a dependency theory to the Michotte task, Wolff (2014) has argued for force dynamics as an overarching model, and Sloman, Barbey, and Hotaling (2009) have proposed that Bayes nets can account for different meanings of abstract causal verbs (*cause*, *prevent*, *enable*), thus directly competing with force dynamic theories. However interesting these attempts are, it seems fair to say that they did not convince the community of causal reasoning researchers to switch their respective theoretical framework and converge on a unitary one. Adaptations to different tasks are in some instances possible but tensions remain between the prime domain of application of the frameworks and their success in explaining phenomena in a different domain.

A tempting alternative that has been proposed by a number of philosophers and psychologists is to suggest *causal pluralism*. Since different domains seem to be best handled by different theories, why not accept all of them as possible accounts? An extreme version of pluralism has been proposed by Cartwright (2004) who argued in her article "Causation: One word, many things" that causal relations in the real world are too diverse to be captured by the abstract terms "cause" and "effect." For example, saying that pistons suck air in or that carburetors feed petrol to a car's engine provides specific information far beyond saying that some cause influences some effect.

More parsimonious accounts of causal pluralism have also been suggested (see Godfrey-Smith, 2009). A popular distinction has been proposed by Hall (2004) who differentiated between difference-making and production (roughly corresponding to our contrast between dependency theories and

process or mechanism theories). An example for a pluralistic account in psychology comes from Lombrozo (2010) who contrasted functional and mechanistic explanations which rely on different concepts of causation (dependency vs. process/mechanisms). Which concept is activated depends on the domain and the type of causal relation in question. For example, whereas, according to Lombrozo, omissions are rarely viewed as causes in physical domains, they may count as causes when intentional agents are involved.

In sum, the general idea of causal pluralism is that different concepts of causation may coexist, and be differentially activated by properties of the domain and task (see also Schlottmann & Shanks, 1992, for a different pluralistic account).

## 3.2 Hybrid Accounts

While pluralist theories suggest that different versions of causal representations coexist and are activated in a domain and task-specific fashion, hybrid theories assume some kind of active collaboration between different types of representations.

We have already encountered one example of such a view, Cartwright's (1999) dispositional power theory. Cartwright claims that causal dependencies (ie, lawful relations) are generated by *nomological machines* that consist of interrelated parts. The powers of these parts give rise to the dependencies we observe. Moreover, the empirically observed dependencies provide cues to dispositional properties of the parts of the underlying nomological machines.

Another example for an attempt to show that different views, in this case an interventionist dependency account and process theories, may collaborate is the proposal by Woodward (2011) who argued that dependency information and geometrical/mechanical information are not competing but may constrain each other. As stated above, causal mechanisms can be captured within dependency accounts as fine-grained chains or networks of interdependent causal variables. However, how these networks are configured also often depends on the components of a mechanism being in the right spatio-temporal configuration. For example, biochemical mechanisms only go forward when various reaction products are brought together in the right spatial position at the right time (Bechtel, 2006).

A related example of a hybrid account from psychology are hierarchical theories that combine top-down domain knowledge with causal Bayes nets. Waldmann (1996) has argued that the structure and parameterization of

causal Bayes nets often is influenced by abstract and domain-specific knowledge about properties of causal relations. For example, Waldmann (2007) has presented evidence showing that domain knowledge about different types of interactions of physical quantities influences the functional form of how multiple causes of a common effect are assumed to be combined (see also Griffiths & Tenenbaum, 2009). An example of the integration between intuitive Newtonian physics with probabilistic inference in causal models was offered by Gerstenberg and Tenenbaum (in press; see also Sanborn, Mansinghka, & Griffiths, 2013).

The present chapter also argues that people use hybrid rather than unitary or pluralistic representations in causal reasoning. We go beyond previous attempts of hierarchically combining domain knowledge with probabilistic dependency representations by including all three frameworks of causal reasoning: dependency, dispositional, and process theories. We demonstrate that people often use hybrid representations combining intuitions motivated by different frameworks that in the literature have often been described as contradicting each other and therefore as competing.

In some limiting cases, especially with tasks designed to test a specific theory, causal reasoning may be well explained by a unitary theory but our claim is that in more typical situations multiple representations interact and constrain each other. There is no reason to assume that people are restricted to one type of representation when trying to understand a causal situation. Outside the laboratory, causal information does not come in conveniently preprocessed modes as, for example, in trial-by-trial learning studies. When we observe causal scenarios there are multiple ways to categorize what we see. We can distill an event or a process representation from the scene or focus on the objects involved in causal interactions. All these possibilities do not exclude each other. It is more plausible that multiple sources of information are simultaneously processed and mutually constrain each other. Which of the different representations becomes the predominant driver of performance may also depend on the task at hand. Predictive inferences may rely more on information about event contingencies, whereas explanatory goals may lead to reflections about the powers of components. Language understanding will activate different processes than observational learning, but in many cases these two sources of knowledge will interact.

We believe that restricting the learning inputs to specific types of formats exaggerates the value of individual frameworks as models of causal reasoning. Both inside and outside the laboratory, we are confronted with

different formats. For example, most experiments studying human learning combine a phase in which verbal instructions inform about causal relations followed by a presentation of trial-by-trial information about individual cases. Theoretical accounts of learning then tend to focus on the trial-by-trial learning component instead of asking how the verbally conveyed instructions interact with the learning mechanisms.

A further advantage of a hybrid theory is that such a theory can close gaps that unitary accounts leave. Causal dependency information is often accompanied by information about the components and processes that constitute the mechanisms, with both types of information constraining our dependency intuitions. On the other hand, dispositional knowledge may be acquired on the basis of covariational learning input. For example, that wind has the power to move boats needs to be learned first based on observations of covariations. Thus, combining the different approaches should lead to a more complete theory of causal cognition.

Although the assumption that people use hybrid representations is certainly attractive, no formal theory has been developed that combines theories from all three frameworks into a unified overarching theory. Such a theory is certainly an important goal, but is beyond the scope of the present chapter. Our research goal is more modest. We are looking for experimental demonstrations of how representations from different competing frameworks interact. Using computational modeling, we will in two of these three case studies demonstrate how we envision the interaction between the frameworks in the particular cases.

## 4. CASE STUDIES

In the following sections, we will present three case studies, which demonstrate the usefulness of a hybrid account of causal representations. The first two case studies show how dispositional and dependency intuitions collaborate in the way causal inferences are made. The third case study addresses physical causation (Michotte task) and shows that both dispositional and process intuitions influence causal perception.

### 4.1 Study 1: The Interaction of Dispositional Intuitions and Dependency Representations—Markov Violations as a Test Case

Our first test case explores the interaction between dispositional knowledge and dependency representations (see Mayrhofer & Waldmann, 2015, for a

more detailed presentation). We have seen that both paradigms have their strengths and weaknesses. Dispositional theories dominate as explanations of linguistic intuitions about causality. By contrast, dependency theories, causal Bayes nets in particular, provide a compact theory of statistical inferences in complex causal models.

Experiments testing how people make such statistical causal inferences typically start with linguistic instructions about the presented causal scenarios. Experimenters use their semantic intuitions when proposing a causal model representation for the task, and frequently cover stories are modified when subjects do not seem to conform to the intuitions of the experimenter. The assumptions linking verbally conveyed cover stories to subjects' causal model representation are, however, mostly left implicit. Our proposal is that dispositional intuitions conveyed by the linguistic cover stories describing causal scenarios constrain the structuring and parameterization of subjects' causal models and therefore influence statistical inferences in a systematic fashion.

### 4.1.1 Agents and Causes

To demonstrate that dispositional intuitions can have an impact on reasoning with dependency representations, we aimed at presenting a situation in which causal dependencies are kept constant while dispositional intuitions about the participants taking part in the causal relations were varied. More specifically, we used the distinction between causal agents and causal patients that is fundamental in many dispositional theories of causation (see Section 2.2). To pit dispositional intuitions against dependency information, we compared situations in which the mapping between cause/effect and agent/patient roles was manipulated (see Fig. 2). While in one condition the cause involved an agent and the effect a patient, in the contrasted condition the cause involved the patient and the effect the agent. This way we could empirically dissociate the influence of the two distinctions.

Conditions in which an agent is involved in the cause event are ubiquitous and seem to be the standard case in causal scenarios. Most cover stories describe scenarios in which causes and agents are confounded, such as food causing allergies, radiation causing diseases, and medicine relieving headache (see also Section 2.1). In all these cases, the cause event involves an entity that is plausibly viewed as the more active part of the causal relation. However, there are situations in which the mapping is reversed. For example, consider a driver who stops in front of a red light. In this situation many people would see the driver as the causal agent who has control over the

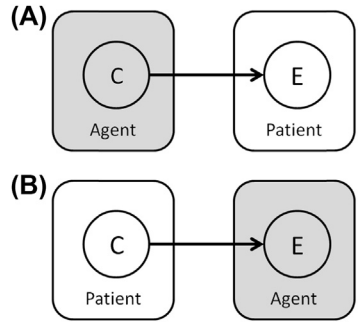


**Figure 2** A single cause–effect relation with (A) the agent role attached to the cause event and (B) the agent role attached to the effect event. *From Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: dispositional intuitions as a guide to causal structure.* Cognitive Science, 39, *72. Reprinted with permission from Wiley.*

situation. But the causal dependency actually runs from the light to the driver. The light is the cause of the driver's behavior which can be easily seen with an intervention test: If somebody intervenes and turns off the light, the behavior of the driver would change, whereas manipulating the behavior of the driver by other means does not affect the light.

In psycholinguistics these kinds of reversals have been discussed in the context of the semantic analysis of *psych(ological) verbs* (see Brown & Fish, 1983; Landau, 2010; Pinker, 1991; Semin & Fiedler, 1988, 1991; Rudolph & Försterling, 1997). Psych verbs reverse the usual mappings between causal roles and grammatical categories. In "Peter pushes Mary," Peter is the agent, the cause, and the subject of the sentence, while Mary is the object, the patient, and is involved in the effect event; this is the standard mapping. In "The show amused Bill," however, the agent (or more specifically the experiencer) is placed in the object role, while causal dependency runs from properties of the show to Bill's emotional reaction. Other examples of psych verbs, which do not necessarily refer to animate agents, are perceiving, receiving, detecting, or reading.

To implement different mappings between cause/effect vs. agent/patient, we used a cover story about aliens adapted from Steyvers, Tenenbaum, Wagenmakers, and Blum (2003). The cover story generally mentioned four aliens whose thoughts are being transferred to each other. In one experiment, we used a common-cause model with one alien's thoughts being transmitted to the three other aliens (see Fig. 1A, for a common-cause model with two effects). In general, we kept the roles of cause and effects constant across different conditions. One alien (eg, top alien in Fig. 3), the cause, was

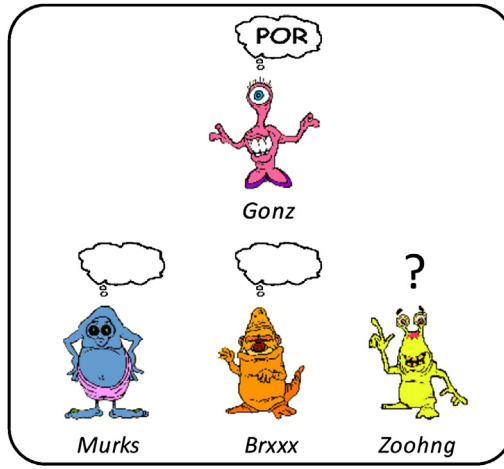


**Figure 3** A common-cause model of aliens whose thoughts were transferred (sending vs. reading) from the top alien, the cause, to the bottom aliens, the effects. *From Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: dispositional intuitions as a guide to causal structure.* Cognitive Science, 39, *85. Reprinted with permission from Wiley.*

described as having a specific thought first which then causes the same thoughts in the effect aliens (eg, bottom aliens in Fig. 3). Thus, the thoughts of the effect aliens were temporally preceded by the thought of the cause alien and depended on it. Represented as a causal Bayes net, the arrows need to be directed from cause alien to effect aliens as in Fig. 1A.

To manipulate the dispositional roles of agents and patients, we used different causal verbs. In one condition, the cause alien was described as being capable of *sending* its thoughts. This verb should establish the cause alien as the agent and the effect aliens as patients (see Fig. 2A). In a contrasting condition, the effect aliens were described as being capable of *reading* the thoughts of the cause alien. In this condition, the effect aliens should be viewed as the agents and the cause alien as the patient (see Fig. 2B).

Discussions with colleagues often led to the question whether the dependency model was really kept constant across conditions; some suggested that in the reading condition the causal arrows need to be reversed (as in Fig. 1B). We believe that the reason for this confusion is that causal agents are typically associated with the cause role. To make sure that subjects' representations of the task conform to the intended causal dependency, we tested their intuitions in an experiment (Experiment 1a in Mayrhofer & Waldmann, 2015). We told subjects about two aliens, Gonz and Murks, who occasionally think of the artificial word "POR." As described above, we contrasted two

conditions in which either the cause alien was capable of sending its thoughts to the effect alien (sending condition) or the effect alien was capable of reading the thoughts of the cause alien (reading condition). We additionally instructed in both conditions that causal strength was high but not perfect and that the effect alien occasionally also spontaneously thinks of POR on its own. To test intuitions about causal dependency, we asked about the outcomes of hypothetical interventions implanting thoughts in the cause or the effect alien. If causal dependency runs from cause to effect in both conditions, as assumed, implanting a POR-thought in the cause alien should increase the probability of POR-thoughts in the effect alien independent of condition (sending vs. reading). Implanting a POR-thought in the effect alien by external means should not change the probability of the cause alien thinking of POR beyond the base rate. The results clearly confirmed these predictions showing that subjects' representations of causal dependency were not influenced by the manipulation of dispositional properties of the alien mind readers.

#### 4.1.2 Agency, Accountability, and Error Attribution

Our general goal was to link dispositional intuitions with dependency representations. As we will see, this is particularly interesting in cases in which causal transmissions fail; that is, the cause is present but unexpectedly its effect fails to materialize. For example, the cause alien may think of POR, but the effect alien does not. In this case, the question of accountability naturally arises. Who is to blame? Our general assumption was that intuitions about responsibility would be moderated by dispositional role. In general, in agent—patient relations a failure will always be a joint result of the strength of the agent and the resistance of the patient. However, since patients are considered passive, being subject to acts by the agent, we suspected that without further knowledge, agents should be primarily blamed for failures. We tested this prediction in our domain by presenting subjects with two aliens as in the intervention study (Experiment 1a, Mayrhofer & Waldmann, 2015) but now we asked them about which of the two aliens was more responsible when the thoughts of the cause alien failed to be transferred to the effect alien (Experiment 1b, Mayrhofer & Waldmann, 2015). The results showed that errors were attributed differently in the two contrasted conditions with stronger attributions to the cause in the sending condition (cause as agent) compared to the reading condition (effect as agent). While in the reading condition the cause alien was only picked by 17.5% of the subjects as being more responsible for causal failure, this number went up to 50% in the sending condition (see also Mayrhofer & Waldmann, 2015, for further discussions).

### 4.1.3 A Bayes Net Account of Error Attribution

The predictions about how accountability is distributed between agents and patients can be translated into causal Bayes net representations. In causal Bayes nets (see Fig. 1), failures of causes to generate their effect are typically coded by the strength parameters attached to the causal arrows. Following Cheng (1997), causal power (or strength) can be interpreted as the probability of a cause producing its effect when alternative causes are hypothesized to be absent. Since causal failure is uniformly expressed in the strength parameter regardless of its source, standard Bayes nets are ill equipped to express differential error attribution to agents and patients. To remedy this deficit, we proposed to split up the error in two components, one attached to the cause ($F_C$) and one to the effect event ($F_E$) (see Fig. 4A). With each cause C, an independent hidden preventive node $F_C$ is associated that is connected to each of its effects with equal strength. Moreover, each effect event has its own error term $F_E$. Thus, in the common-cause model shown in Fig. 4B a single hidden error term $F_C$ is attached to the cause event with equally strong links leading to each of the effects of the common cause. Moreover, there are three effect-related errors $F_E$ attached to each of the effects individually.

Manipulating the strength of the links emanating from $F_C$ allows the network to express how failures of sufficiency are distributed between cause and effects. If the weights coming from $F_C$ are relatively high, errors are mainly attributed to the object involved in the cause event (eg, the agent). If these weights are low, failures are attributed more to the effect side.

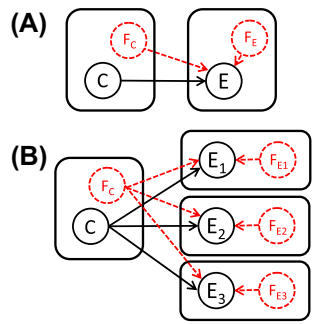


**Figure 4** Panel A shows an elemental cause—effect relation with two sources of failure, a cause-related error node $F_C$ and an effect-related error node $F_E$ *(From Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: dispositional intuitions as a guide to causal structure. Cognitive Science, 39, 75. Reprinted with permission from Wiley)*. Panel B shows the augmented representation of a common-cause model with a single cause-related preventive error node $F_C$ and three effect-related error nodes $F_E$.

Depending on where in the network agents and patients are located, setting the weight parameters for $F_C$ relative to $F_E$ therefore allows us to express differential attributions of errors.

### 4.1.4 Markov Violations as a Test Case

One of the central features of Bayes nets is the Markov property according to which each variable conditioned upon its direct causes is independent of all other events, except for its direct and indirect effects. This property is one of the defining features of Bayes nets. It is the key reason for their parsimony because it allows for making local inferences without having to consider all variables in the network. It suffices to focus on the causes of an effect event to make a prediction about its status.

Despite its computational advantages, some philosophers (eg, Cartwright, 2007) have cast doubt on the adequacy of the Markov condition as a property of causal relations in the world. Moreover, a number of empirical findings have shown that subjects' causal reasoning routinely tends to violate the Markov condition. Initial evidence for this phenomenon came from experiments by Rehder and Burnett (2005) in which subjects were presented with a common-cause model in some domain (see Fig. 1A) and were asked to judge how likely one of the effects is when they know for sure that the cause is present. According to the Markov condition, the inference should flow from cause to the target effect while being unaffected by the status of the other effect variables. However, the results showed that subjects did not ignore the collateral effects. When the other effects were present, the estimates for the target effect were higher than when they were absent (see also Walsh & Sloman, 2007).

A first reaction to these results was that subjects might not restrict themselves to the instructed common-cause model but might have augmented it with hidden variables that expressed their domain knowledge about the complex relations underlying the shown variables. Such an augmented network may honor the Markov condition, while explaining the apparent Markov violations in the restricted model presented in the instructions. This explanation certainly is reasonable for real-world domains, for example disease scenarios, to which subjects bring to bear prior knowledge. However, this interpretation is weakened by a further experiment of Rehder and Burnett (2005) that demonstrated Markov violations of equivalent size with tasks in which just abstract lettered variables (A, B, C, and D) had been presented without any reference to specific domains and mechanisms (see also Rehder, 2014).

In light of Markov violations even in abstract domains for which no domain-specific knowledge is available, we suspected that abstract dispositional assumptions may provide a more general explanation of this phenomenon. Given that causal agents are typically associated with cause events, an abstract verbal instruction such as "A causes B" does not only describe a dependency relation between A (the cause) and B (the effect) but also implicitly assigns the agent role to the subject (A) and the patient role to the object of the sentence (B).

To demonstrate that Markov violations were mediated by dispositional intuitions about agency, we presented subjects with our instruction about four aliens and manipulated agency as described above (sending vs. reading; see Experiment 2 in Mayrhofer & Waldmann, 2015). To test for the existence and strength of Markov violations, we gave subjects in the test phase several hypothetical test cases in which the cause alien either thinks of POR or not, and in which the number of collateral effect aliens also thinking of POR was manipulated (none, one, or two). For all these cases, we asked subjects to estimate the number of cases out of 10 in which target effect alien probably thinks of POR.

Our central hypothesis was that Markov violations should be observed to be stronger when the cause alien was associated with the agent role (ie, sending condition) than when the effect aliens were the agents (ie, reading condition). In the sending condition, responsibility for errors should be more strongly attributed to the cause alien. When, for example, the two collateral effect aliens did not receive the thought of the cause alien, a plausible interpretation is that something must have gone wrong with the sending capacity of the cause alien, something which should affect all effects at once; thus, lower ratings for the target effect are to be expected relative to cases in which everything seems fine (eg, all collateral effect aliens had received the thought of the cause alien).

By contrast, in the reading condition errors should more strongly be attributed to the effect aliens. The fact that two collateral readers fail to achieve their goal should not be predictive of the capacity of the target alien. Its reading capacity may still be intact despite the problems of the collateral effect aliens. Thus, no (or at least much smaller) Markov violations were predicted for the reading scenario.

Fig. 5 shows the results. Generally the ratings were higher when the cause alien thought of POR than when it did not think of POR, which is consistent with the instructions. Moreover, there were no statistically significant Markov violations when the cause was absent (ie, the cause alien did not think of POR). This is also to be expected because in the absence

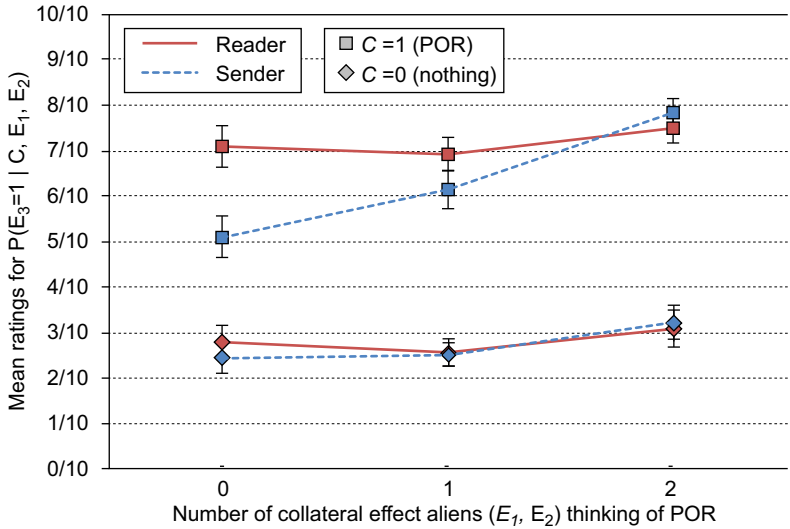


**Figure 5** Mean ratings (and standard errors) representing the estimates of the relative number of times the target alien thinks of POR in 10 hypothetical situations. *From Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: dispositional intuitions as a guide to causal structure.* Cognitive Science*, 39, 86. Reprinted with permission from Wiley.*

of the cause no transmission of thoughts and hence no failure is possible. Most importantly, we found the hypothesized interaction when the cause was present (ie, the cause alien thought of POR). The positive slope in the sender condition reveals a clear Markov violation. The estimates of the likelihood of the target alien went up the more collateral aliens thought of POR. By contrast, there was only a slight, but non-significant increase in the reading condition, which is in line with our prediction. Thus, we were able to manipulate the size of the Markov violations by manipulating the dispositional properties of the involved causal participants.

Our proposed Bayes net representation can account for these findings (Fig. 4B). In a common-cause model, a single independent hidden preventive variable node representing $F_C$ is added to the model (and three error terms for the effect nodes, $F_E$). The state of the preventer $F_C$ is inferred based on the status of the effects. Absent effects signal a higher likelihood of the presence of $F_C$ than present effects. Since $F_C$ is linked to all three effects with equal weights, its activation also dampens the prediction for the target effect. The strength of the links of this preventive node $F_C$ relative to the strengths of $F_E$ represent how strongly subjects attribute errors to the cause node.

Thus, when the cause event involves the agent, the strength of the weights is set to a relatively high level, which entails a strong Markov violation. When the weights are relatively low, a weaker Markov violation is predicted, as in the case when the agents are located on the effect side. In this case, errors are more strongly attributed to each effect individually. The model is similar to others suggested in the literature (Hausman & Woodward, 1999; Rehder & Burnett, 2005; Walsh & Sloman, 2007). The new feature is our proposed separation of sources of errors that are attached to the cause and effect sides and whose settings are motivated by the dispositional distinction between agents and patients involved in the causal relations.

### 4.1.5 Alternative Theories

While many early studies on Markov violations had focused on the demonstration and explanation of its existence (eg, Rehder & Burnett, 2005), subsequently the question whether and how the size of Markov violations can be manipulated came to the forefront. Since we have proposed a hybrid account to explain such effects, it is interesting to compare our view with the theories of others who argued with principles coming from within their chosen unitary framework.

An important theory was proposed by Park and Sloman (2013). They have presented several accounts, but we focus on the one which can be viewed as a direct competitor to ours. Adopting a causal Bayes net representation, they argue that the size of Markov violations is influenced by assumptions about the causal mechanisms. Since causal mechanisms can be easily represented in causal Bayes nets as chains of variables, this approach does not require assumptions coming from other causal frameworks. Their main hypothesis is that Markov violations in common-cause models will be observed when the cause generates its different effects using the same type of mechanism. Whether a mechanism is of the same or a different type is determined by looking at the intervening variables mediating between cause and effects. For example, when the causal model links smoking as the cause with the two effects impairment to lung function and damage to blood vessels, it is assumed that smoking leads to the two effects via the same mechanism (ie, the same intermediate variable). The intermediate variable plays a similar computational role as our hidden preventive node $F_C$, which both predict Markov violations. By contrast, when in a different causal model smoking is linked to both an impairment of the lung function

and a financial burden on the family budget, different mechanisms with different intervening variables are involved. Hence, no Markov violation is expected. These predictions were largely confirmed in the experiments of Park and Sloman.

We did not directly test our theory against Park and Sloman's (2013) because our studies were conducted prior to the publication of their results. However, it is useful to compare their account with the one we would propose. One advantage of our theory is that it is framed at a more abstract level than the mechanism account; hence, it can also be applied to more abstract tasks for which no mechanism knowledge is available, such as our alien scenario or Rehder and Burnett's (2005) experiment in which letters were used to describe causal variables. Moreover, we would analyze the tasks of Park and Sloman differently. In our view, it is not the different intervening variables that lead to the effects but the fact that in the two situations different dispositional properties of the cause are relevant. In the homogeneous disease context, it is plausible that a single agent, smoke, is responsible for both effects. However, in the separate mechanism condition, different dispositional properties of cigarettes (smoke vs. cost), and, therefore, different causal agents generate the two effects. Thus, in this case we would also not expect a Markov violation.

This analysis also applies to Park and Sloman's (2013) Experiment 3 in which an abstract task with sliders being in different positions were presented as causal variables. In the same mechanism condition all sliders have the same color whereas in the different mechanism condition the two effects had different colors and the cause was split in the middle with one of the colors on one side, the other on the other side. Looking at the materials from our perspective, we doubt that subjects have intuitions about same or different intervening variables linking cause and effects here. What is salient, however, are the differences of the cause display, which either has a single or two color features. Again, from a dispositional perspective one might argue that subjects may have viewed the different colors as indicators of two independently operating causal agents.

We do not want to argue that assumptions about mechanisms do not play a role in explaining Markov violations. When mechanism knowledge is available, it will certainly be used. However, in many cases we doubt that people have the required knowledge (see also Rozenblit & Keil, 2002). Dispositional theories may be plausible candidates for explaining different intuitions without requiring elaborate mechanism knowledge about the nature of intermediate variables.

## 4.2 Study 2: Mutual Constraints Between Dispositional Intuitions and Dependency Knowledge

When we access knowledge about causal relations, we often have intuitions about the strengths of the relations (see Fernbach et al., 2011; Fernbach & Erb, 2013, for Bayes net models of real-world knowledge). However, it is far from clear where this knowledge comes from and how it is applied in different contexts. One source of knowledge about probabilistic causal relations may certainly be experience. This factor has often been studied in learning experiments (eg, Cheng, 1997; Griffiths & Tenenbaum, 2005; Waldmann, 2000; see Rottman, in press, for an overview). Outside the laboratory, examples may include physicians who see lots of patients or traders who watch changes of the value of stocks. However, in many cases our knowledge is based on verbal communication. We may read in text books or newspapers about causal relations (eg, medicine), the nature of which is often not quantitatively specified. Nevertheless, when we learn a new fact about a causal system, our intuitions about the dependency relations may be altered.

In the present case study, our aim was to investigate the interplay between verbally conveyed dispositional knowledge and dependency intuitions in a more systematic fashion (cf. Mayrhofer, Quack, & Waldmann, 2016). While our focus in Case Study 1 was on how dispositional assumptions affect the structuring and parameterization of causal models, here we were interested in the two-way interaction between dispositional intuitions and dependency knowledge.

### 4.2.1 Probabilistic Force Model

Fig. 6 displays the key features of our *probabilistic force model* which integrates statistical dependencies and force intuitions. It is inspired by Wolff's (2007) force theory and Mumford and Anjum's (2011) dispositional vector model (see Fig. 6A and Section 2.2), here applied to agents and patients instead of elemental properties of causal objects (Fig. 6B). In the following examples, we will focus on standard cases in which agents are involved in cause events.

Fig. 6B displays a simplified case of forces associated with an agent and patient starting at a neutral point directed toward an effect (right side) or away from the effect (left side). A threshold on the right side determines whether the effect will be observed (shaded area). The key assumption is that the resultant force of the interaction between causal agent and patient ($F_{\text{Result}}$) is an additive function of the force of the agent, $F_{\text{Agent}}$, and the

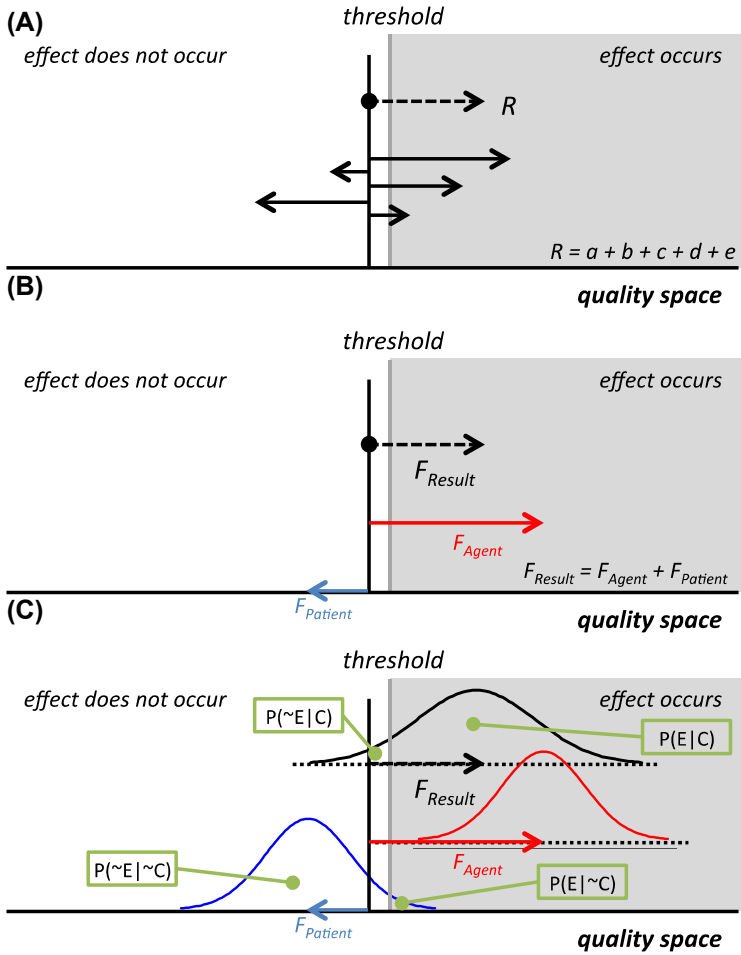


**Figure 6** (A) An illustration of the dispositional vector model of causation (Mumford & Anjum, 2011), (B) a simplified version which is the basis for the probabilistic force model, (C) a demonstration of how vectors with uncertain length are linked with probability distributions to predict dependency intuitions *(From Mayrhofer, R., Quack, B., & Waldmann, M. R. (2016). Causes and forces: A probabilistic force dynamics model of causal reasoning (in preparation))*.

counterforce of the patient, $F_{\text{Patient}}$ (see also Wolff, 2007). In Fig. 6B the resulting force ends well beyond the threshold so that the effect is expected to occur.

So far the vector representation is deterministic. To add uncertainty and link force vectors with probabilistic dependency representations, we added the assumption that some degree of uncertainty is attached to force

representations. This uncertainty is represented by the Gaussian distributions in Fig. 6C. In the present demonstration, we assume the standard deviations of these distributions to be 1 and an arbitrary threshold. With these assumptions, conditional probabilities (and hence contingencies) can be translated into force vectors and vice versa. For example, if in a data set a specific probability of the effect conditional upon the presence of the cause, $P(E|C)$, is observed, then this probability should correspond to the result vector, $F_{Result}$, which represents the outcome of the interaction between agent and patient. $P(E|C)$ then is represented by the area of the distribution of $F_{Result}$ that lies beyond the threshold; the remainder represents $P(\text{non-E}|C)$ (see Fig. 6C).

$F_{Patient}$ in turn represents the prior tendency of the patient that, in the example, disposes away from the effect. Assuming uncertainty again, the effect may occasionally occur even when the agent (ie, cause event) is absent. The likelihood of this happening depends on the length and the distribution of the patient vector (the standard deviation is again assumed to be 1). Given these assumptions, the expectation of the length of the patient vector is the value for which $P(E|\text{non-C})$ corresponds to the area to the right of the threshold of the patient vector distribution. The remaining area represents $P(\text{non-E}|\text{non-C})$. Assuming additivity between agent and patient vectors when determining the result vector, the expectation of the agent vector can be calculated (see Fig. 6C).

### 4.2.2 Experiment

As an initial test of this idea, we ran an experiment in which 32 subjects participated. Initially subjects were instructed that cows were bitten by snakes that normally inject a certain amount of poison (eg, 400 mg). Some of the cows receive a specific amount of an antidote (eg, 200 mg). Then, in Phase 1 of the experiment, subjects were presented with contingency data showing the number of dead or surviving cows when no antidote was given versus when antidote was given (see, eg, Fig. 7). Four different contingencies were presented to each subject in which we varied the base rate of the effect, $P(E|\text{non-C})$, in two levels (0.4 vs. 0.8) and causal power also in two levels (0.5 vs. 0.8) using Cheng's (1997) power equations.

In this scenario, the survival of the cow is the target effect. The cover story describes being poisoned as the default situation which was therefore modeled as a property of the patient (along with other properties of cows that were assumed to be invariant). The antidote represented the agent, which disposed the cow toward the target effect. (Note, however, that these assignments are

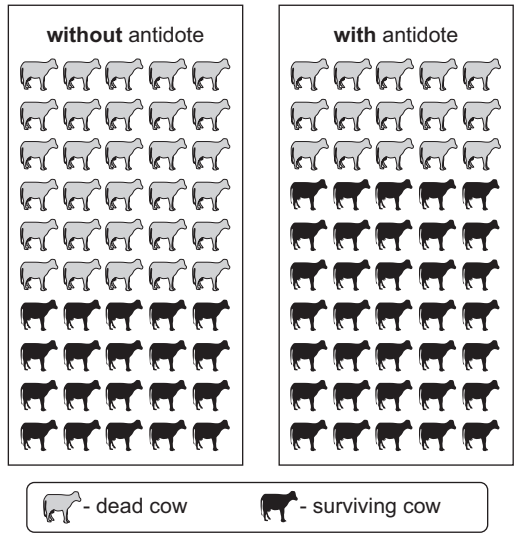


**Figure 7** Example of contingency information presented to subjects in the experiment. *From Mayrhofer, R., Quack, B., & Waldmann, M. R. (2016).* Causes and forces: a probabilistic force dynamics model of causal reasoning *(in preparation).*

relative to the given situation. In other contexts, the poison may be viewed as the agent which interacts with physiological properties of the cows.)

In Phase 2 of the experiment, a new situation was verbally presented that had never been observed before. Our key question was whether subjects were able to translate these verbally conveyed changes into new base rate and causal strength estimates without having seen new contingency data. To study this competency, we varied agent force (amount of antidote: 50%, 100%, and 150% of previously observed amount) and the prior tendency of the patient (amount of snake poison: 50%, 100%, and 150%), yielding a $3 \times 3 \times 4$ (ie, 36 conditions) within-subject design. For example, in the test phase subjects in one condition were asked to imagine a new geographical area in which cows bitten by snakes are injected with 100 mg of poison (that is, 50% of the previous amount). In this area, whenever an antidote was delivered, 300 mg were given (ie, 150% of the previous amount). Subjects were then asked to estimate for this scenario how many out of 10 cows who had been bitten and who would otherwise die would survive had they been given the antidote. This question measures subjects' intuitions about causal strength. To measure intuitions about base rates, $P(\mathrm{E}|\text{non-C})$, subjects were asked to imagine 10 bitten cows and were then requested to judge how many of these cows would survive without the antidote.

The key question was whether subjects could translate the verbally presented numerical context changes into sensible probability estimates. The probabilistic force model allows for such a translation between force representations and dependencies by multiplying the length of the vectors by the instructed change proportions (0.5, 1, and 1.5). These new vectors can then be translated by the probabilistic force model into new base rate and strength estimates.

The model makes four qualitative predictions, which were tested in the experiment: First, base rate judgments, $P(\text{E}|\text{non-C})$, should be lowered when the patient force (that disposes away from the effect in this case) is increased by means of the instructions. In the experiment, this effect was observed, $F(2,62) = 99.95$, $p < .001$, $\eta_p^2 = .76$ (see Fig. 8A). Second, causal strength estimates should decrease when the counterforce attached to the patient is strengthened, which we also found, $F(2,62) = 100.34$, $p < .001$, $\eta_p^2 = .76$ (see Fig. 8A). Third, causal strength estimates are expected to increase when the agent force becomes stronger, $F(2,62) = 153.03$, $p < .001$, $\eta_p^2 = .83$ (see Fig. 8B). Finally, the model predicts no effect of agent force on base rate judgments. However, this effect unexpectedly turned out to be significant, $F(2,62) = 6.95$, $p < .01$, $\eta_p^2 = .18$, although it is barely visible (see Fig. 8B) and the smallest of the observed effects in the highly sensitive within-subject design.

In sum, Case Study 2 presented initial evidence for the newly developed probabilistic force model that is capable of translating force changes conveyed by verbal instructions into probabilistic inferences. Thus, the model formalizes a possible interaction between two types of representations, dispositional intuitions about forces and causal dependencies.

### 4.3 Study 3: Dependencies, Processes, and Dispositions: The Michotte Task

The third case study looks at a different phenomenon: causal perception. A classic task demonstrating phenomenal causality is the Michotte task in which subjects are presented with moving colliding objects (Michotte, 1963). In a *launching* scenario, for example, Object X, a ball, moves toward a resting Object Y, another ball, and touches it. At this moment, Object X stops and Object Y is set into motion eliciting a causal impression (see Fig. 9, Condition A, for an illustration). The strength of the causal impression depends on a number of parameters, including the time lag between X stopping and Y starting its movement, the spatial gap between X

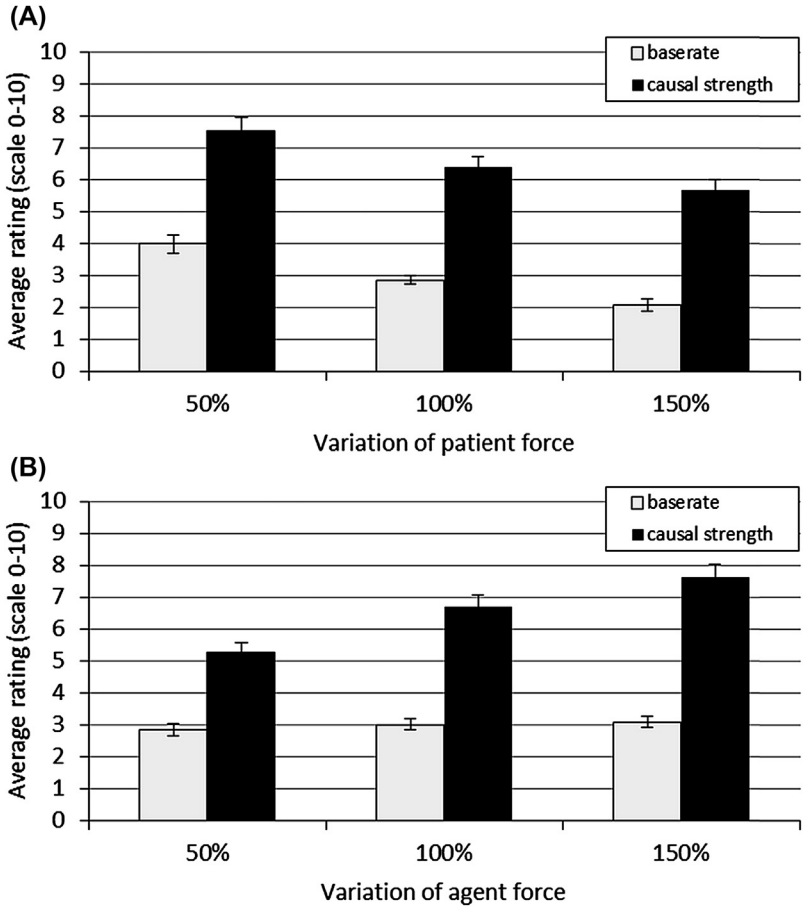


**Figure 8** Results of the probabilistic force experiment for (A) variation of agent force (marginalized over patient force variation), and (B) variation of patient force (marginalized over agent force variation). *From Mayrhofer, R., Quack, B., & Waldmann, M. R. (2016).* Causes and forces: a probabilistic force dynamics model of causal reasoning *(in preparation).*

and Y, or the ratio of pre- and postmovement velocities of the objects (see, eg, Hubbard, 2013; Scholl & Tremoulet, 2000).

In physics, movements and collisions of macroscopic objects can be predicted by Newtonian mechanics. Recently, Sanborn et al. (2013) proposed the *noisy Newton* model that implements probabilistic Bayesian inference over a Newtonian representation of the world. The key feature distinguishing this psychological model from Newtonian physics is that it is assumed that observations (eg, of object velocities) are noisy and therefore lead to some degree

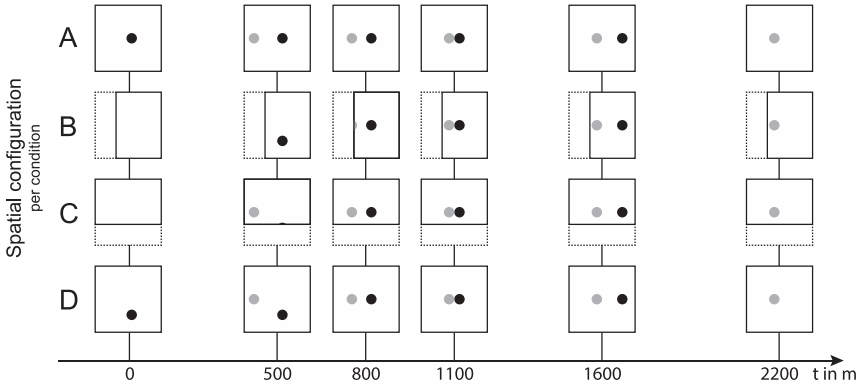


**Figure 9** Illustration of the experimental setup showing the spatial configuration of the balls at significant time points in Conditions A to D. *From Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: the role of the prior context.* Cognition, 132, *486. Reprinted with permission from Elsevier.*

of uncertainty. The noisy Newton model has proven successful in predicting judgments about masses of colliding objects. Moreover, Sanborn et al. applied their noisy Newton model to launching scenarios in which the spatiotemporal gaps between the two balls were manipulated. The model correctly predicts that increasing gaps lead to a lowering of judgments of causality.

However, one phenomenon resists explanation for Newtonian theories. White (2006a) has pointed out that subjects when confronted with a launching event tend to view Object X as the agent and Object Y as the patient (or cause and effect object in his terminology). This so-called *causal asymmetry* effect manifests itself in the preferred descriptions of subjects. They tend to describe the launching scene as an event in which "X launched Y" instead of the equally valid description that "Y stopped X." Moreover, force estimates for X tend to be higher than force estimates for Y. Both findings are indicators of the underlying dispositional distinction between causal agents and causal patients, according to White (who uses the terms cause object and effect object). Causal asymmetry contradicts Newtonian physics because the physical force on Object Y exerted by Object X is equal in magnitude (but opposite in direction) to that on Object X exerted by Object Y. From a Newtonian perspective, the collision is perfectly symmetric, and both descriptions (ie, "X launched Y" and "Y stopped X") should be equally appropriate.

White (2009) has proposed a dispositional theory of causal asymmetry that links perceived scenes to stored representations of sensomotoric experiences of our actions on objects (see Wolff & Shepard, 2013, for an alternative theory). According to White, we experience our own agency and the force

we impose upon objects we manipulate during the course of our ontogenetic development. When perceiving a scene, we compare the movements of the objects with these stored representations. We tend to overestimate the force of the causal agent (cause object in his terminology) relative to the counterforce of the manipulated patient (ie, effect object) because the (counter-)force exerted by the patient is perceptually attenuated in cases in which we manipulate objects (ie, the source of our stored representations). White's theory is a unitary dispositional theory: Both the description of causal scenes and the attribution of forces are driven by the asymmetry of the agent–patient relation, which is primary in White's theory.

In our view, causal perception of collision events is better captured by a hybrid account that combines a dispositional component and a process component. First, we will show that linguistic descriptions of perceptual causal scenarios are influenced by dispositional properties of objects in the scene. This speaks against a pure process account of causal perception and is consistent with White's view. However, we also show that measures of properties of the observed causal process (ie, perceived forces) do not necessarily covary with dispositional assignments, which contradicts White's unitary account according to which both linguistic descriptions and force assignments are influenced by the dispositional properties of the observed interacting causal objects (see White, 2014, for a different view).

To disentangle agency and force judgments from the observed collision event, we conducted two sets of experiments. In our first set, our goal was to dissociate agency assignment from the collision event, which are typically confounded in the Michotte task (see Mayrhofer & Waldmann, 2014, for a more elaborate description). Our experimental goal was to keep the collision event constant but manipulate agency through indicators that are perceptually available prior to the collision event. Thus, in all conditions of our experiments, the events at and after the collision of Balls X and Y were identical. Therefore, all factors leading to the distinction between agent and patient that were associated with the moment of collision and subsequent events were kept constant.

To manipulate agency using features available prior to collision, we turned to Dowty's (1991) theory of agency, which he had developed to explain how people distinguish between agents and patients in language. Dowty suggests that agent and patient roles are *prototype* concepts. For example, a prototypic agent is, among other properties, volitional, sentient, and causes or changes an outcome. None of these features is necessary for agency, but the more features a causal participant shares with the

agent or patient prototype, the more likely it is that it plays the respective semantic role.

Dowty's (1991) features were developed with language in mind; therefore, we adapted his list to the Michotte task so that it applies to movements in perceptual scenes. Thereby we focused on features that can be seen in the precollision phase. It is notable that most of Dowty's properties constituting the prototype of an agent can be viewed as properties of active human interventions.

One important property of agents is that they tend to move prior to contact while the patient remains stationary until launched. Thus, following White (2006b), we expected that moving objects should be attributed more agency than stationary ones. Relative movement was our first feature distinguishing agents from patients.

As a second feature, we manipulated the sequence of appearance of causal participants. Since a prototypical agent intervenes into an existing scenario that is either stationary or changing in a predictable way, the object that enters the observed scene last should tend to be attributed relatively more agency than objects that are already part of the observed scene. To test this hypothesis, we kept the movements constant, but manipulated the sequence of visibility by hiding parts of the scene.

Finally, we manipulated cues indicating volitional action which is a key property of human intervention: When a spontaneously moving object behaves in a manner not obviously explainable by physical knowledge (eg, self-propelled motion), the object's behavior tends to be interpreted as a volitional act by an animate agent (see Csibra, Gergely, Bíró, Koos, & Brockbank, 1999; Muentener & Carey, 2010; Saxe, Tenenbaum, & Carey, 2005).

We manipulated these precollision cues to disentangle factors affecting agency from the properties at and after the collision. White's (2009) extensive set of studies shows that the launching event at the point of contact constitutes a strong cue suggesting Object X to be the agent and Object Y the patient. Since the launching event was kept constant, our goal therefore was to test how much the precollision cues we tested can override the cue that is inherent in the causal interaction.

We tested the influence of the three precollision cues on agency assignments in four within-subject conditions of an experiment (see Mayrhofer & Waldmann, 2014; Experiment 2). In all conditions, the movement properties of the two objects at and following the collision event were kept constant. Thus, when Ball X, coming from the left side, hit Ball Y, Ball X

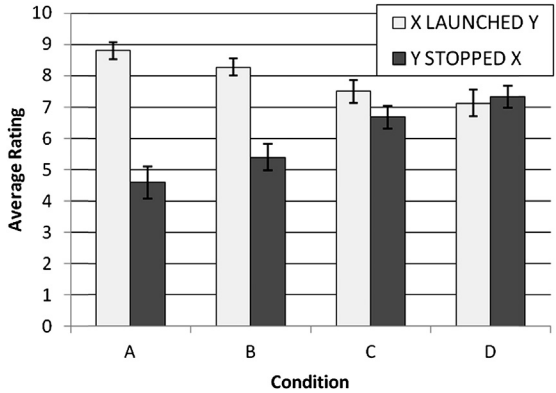


**Figure 10** Results of experiment (error bars indicate standard error of means). *From Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: the role of the prior context.* Cognition, 132*, 489. Adapted with permission from Elsevier.*

stopped and Ball Y moved toward the right hand edge of the screen (see Fig. 9). In all conditions, Ball Y stood still in the middle of the screen immediately prior to contact. Condition A (Fig. 9, top row) represented a standard Michotte launching case in which Ball Y stands still in the middle of the screen until launched by Ball X. Here Ball X should be clearly viewed as the agent. In all other conditions, Ball Y moves from the bottom of the screen toward its collision point where it stops 300 ms prior to the collision. By hiding either the left margin (Condition B; second row in Fig. 9) or the bottom margin (Condition C; third row in Fig. 9) or by letting Ball Y start a self-propelled movement (Condition D; bottom row in Fig. 9), we added in each condition one additional agency indicator suggesting Ball Y as the agent (for details see Mayrhofer & Waldmann, 2014).

Fig. 10 shows the results of the experiment (Experiment 2, Mayrhofer & Waldmann, 2014). As an indicator of agency attributions, we asked subjects to rate on a scale from 1 to 10 how much they agree with the statements "X launched Y" or "Y stopped X," respectively (which in a specific condition was, for example, instantiated as "The red ball stopped the blue ball."). As can be seen in Fig. 10, adding features of agency to Ball Y (successively from Conditions A to D) clearly had an impact on the ratings. The more agency cues for Ball Y were present, the higher was the agreement with the statement that "Y stopped X." However, in no condition agency attributions for Y turned out to be higher than for X. This is to be expected because the collision event, which was kept constant across conditions,

generally leads to the strong impression that Object X is the agent. In sum, the experiment showed that dispositional assumptions triggered by perceptual cues influence causal perception, contradicting a Newtonian (Sanborn et al., 2013) or a pure causal process account.

In an additional set of experiments, we were interested in the second indicator of causal asymmetry, force asymmetries. Asymmetric force ratings for agents and patients have been viewed as another hallmark evidence for dispositional theories (see White, 2009). Although dispositional theories explain force asymmetry as due to the asymmetric roles of agent and patient in causal interactions, in most studies agency has not been manipulated independently of properties of the collision. This confounding opens up the possibility that properties of the collision (eg, pre- and postcollision velocities) may independently influence agency and force perception, thus creating a spurious correlation.

In order to test whether causal agency influences force judgments, we again manipulated agency indicators independent of an otherwise invariant launching event (Mayrhofer & Waldmann, 2016, for a more detailed presentation). We focused on two conditions, the standard Michotte launching event (Condition A in the previous set of studies, Fig. 9) and the condition in which we presented all three additional agency cues for Ball Y (intervention condition; corresponding to Condition D, Fig. 9). We chose these two conditions because they led to the strongest effect on agency attributions in the previous study. In the present study, we measured agency attributions using a forced choice measure by presenting subjects with the two alternative statements "X launched Y" and "Y stopped X" (see also Mayrhofer & Waldmann, 2014; Experiment 1). Additionally, we requested subjects to rate the forces attached to X and Y using a rating scale ranging from 0 to 100. This time we ran the study online as a between-subjects design.

The results were clear (see Fig. 11). The force ratings (right panel) exhibited a clear causal asymmetry effect. Object X was uniformly assigned more force than Object Y. However, whereas we found significantly more attributions of agency for Y in the intervention than in the standard condition (left panel), the different agency attributions in the two conditions had no effect on the force ratings. This pattern was replicated in a second study (Mayrhofer & Waldmann, 2016; Experiment 1).

The results of the experiment cast doubt on the adequacy of theories based on Newtonian physics (Sanborn et al., 2013) and on purely dispositional theories (White, 2009). Noisy Newton theory has problems with explaining the stable findings of agency and force asymmetries, whereas

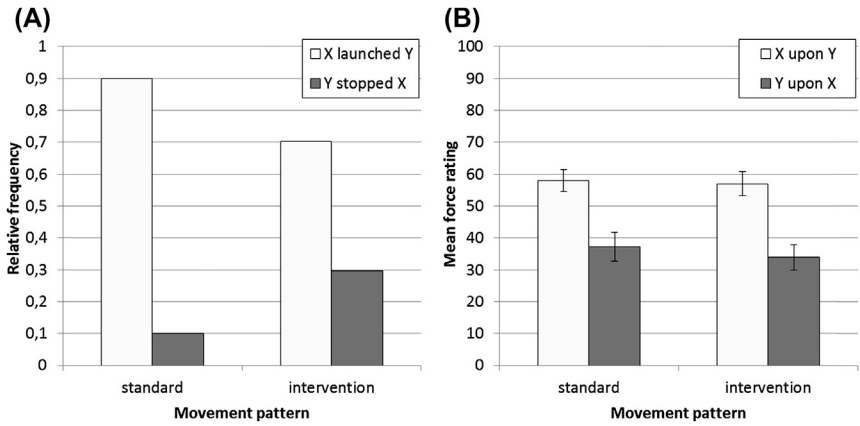


**Figure 11** Relative frequency of causal-agency assignments (A) and force ratings (B) for Ball X and Ball Y in the two movement conditions (standard launching vs. intervention). Error bars indicate 95% confidence intervals. *Mayrhofer, R., & Waldmann, M. R. (2016). Causal agency and the perception of force.* Psychonomic Bulletin and Review *(in press).*

dispositional theories arguing for a direct link between agency and force assessments cannot explain the dissociation between the two measures in our experiments.

A psychological version of causal process theories may be an alternative candidate for explaining force asymmetry. According to Dowe (2000), causal processes carry a quantity, such as linear momentum, mass-energy, or charge, which is conserved within the process. Of course, only experts know these physical quantities whereas most subjects do not have deep knowledge about physics (see Rozenblit & Keil, 2002). However, despite the lack of elaborate physical domain knowledge it seems plausible to assume that even laypeople represent the Michotte task as a causal process in which some sort of hidden placeholder property is transmitted when Ball X moves toward Ball Y and makes contact. A psychologically plausible candidate for such a property might be the (pre-Newtonian) concept of impetus, which is usually represented as an internal force that keeps an object moving and which can be assumed to be transferred from one object to another in a collision event (see, eg, Kozhevnikov & Hegarty, 2001; McCloskey, 1983). If force intuitions traced the transference of such impetus (ie, internal force), one would expect an asymmetrical assignment of forces that expresses the directionality of the causal interaction. Force asymmetry then can be used as one of several cues of an agent prototype rather than an effect of it as in White's (2009) theory. This finding again suggests a hybrid account; in this case a combination between dispositional and causal-process theories.

## 5. CONCLUSION

The main goal of this chapter was to defend a new view on causal reasoning, a hybrid representation account. In our review of theoretical frameworks, we have shown that different types of theories of causal reasoning compete with each other, each endowed with its distinctive strengths and weaknesses and its preferred domains of application. We have argued that despite the beauty of a parsimonious unitary account, there is little reason to assume that people are restricted to one type of representation of causal scenarios. When trying to make sense of the world, we receive information in different input formats which we then have to translate into some plausible representation of the causal texture of the world. Unlike in the psychological laboratory, we are rarely confronted with conveniently precategorized representations that invite us to only use a specific framework of causal reasoning.

In contrast to causal pluralism, which postulates the coexistence of independent modes of causal reasoning, our aim was to show that competing representations not only coexist, they can also actively influence each other. In three empirical case studies, we have demonstrated how dependency, dispositional, and process representations mutually interact in generating complex representations driving causal inferences. Using computational modeling, we have in two of these three case studies demonstrated how we envision the interaction between the frameworks in the particular cases.

Our three case studies just represent a first step in the direction of developing a hybrid account of causal reasoning. Future empirical studies will have to systematically explore when and how different competing representations influence each other. So far we have mainly focused on the standard experimental tasks that present causal situations in ways optimized for the modeling goals of the favored framework. To overcome these limitations, it would be desirable to study more realistic scenarios that are closer to what we encounter in our everyday experience. When learning about causal relations, we often combine different sources of knowledge that influence each other. In particular, studying the interaction between causal intuitions conveyed by linguistic cover stories and statistical input seems to be particularly important to understand the results of experiments in which typically both sources of causal knowledge are provided. It is unfortunate that theoretical models usually only focus on the experiential input while the role of the linguistic cover stories is not addressed. Our Case Study 1 is just a first step in the investigation of the interaction of these two sources of knowledge.

So far the general strategy in research on hybrid representations has been to show how knowledge from different sources constrain causal representations. Of course, the most ambitious goal for future research is to develop a more general hybrid theory of causal reasoning that combines concepts from the all three frameworks within a unified theory.

## ACKNOWLEDGMENT

We thank J. Nagel for helpful comments.

## REFERENCES

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: mental models or causal models. *Cognition, 119*, 403–418.

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision, 3*, 20–39.

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

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14*, 237–273.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1119–1140.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge, UK: Cambridge University Press.

Cartwright, N. (2004). Causation: one word, many things. *Philosophy of Science, 71*, 805–819.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge, UK: Cambridge University Press.

Cartwright, N., & Pemberton, J. M. (2012). Aristotelian powers: without them, what would modern science do? In J. Greco, & R. Groff (Eds.), *Powers and capacities in philosophy: The new Aristotelianism*. New York: Routledge.

Cheng, P. W. (1993). Separating causal laws from casual facts: pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 215–264). New York: Academic Press.

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365–382.

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

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

Dowty, D. (1991). Thematic proto roles and argument selection. *Language, 67*, 547–619.

Dwyer, D. M., & Waldmann, M. R. Beyond the information (not) given. Representation of stimulus absence in rats (*Rattus norvegicus*). *Journal of Comparative Psychology*, in press.

Ehring, D. (2009). Causal relata. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 387–413). Oxford, UK: Oxford University Press.

Fair, D. (1979). Causation and the flow of energy. *Erkenntnis, 14*, 219–250.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140*, 168–185.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1327–1343.

Gärdenfors, P. (2014). *The geometry of meaning. Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gerstenberg, T., & Tenenbaum, J. Intuitive theories. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.

Godfrey-Smith, P. (2009). Causal pluralism. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 326–337). Oxford, UK: Oxford University Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review, 111*, 1–30.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716.

Hall, N. (2004). Two concepts of causation. In L. A. Paul, E. J. Hall, & J. Collins (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge, MA: MIT Press.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science, 66*, 413–457.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science, 50*, 521–583.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: the new synthesis. *Annual Review of Psychology, 62*, 135–163.

Hubbard, T. L. (2013). Phenomenal causality I: varieties and variables. *Axiomathes, 23*, 1–42.

Johnson-Laird, P., & Khemlani, S. (2016). Mental models and causation. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*. New York: Oxford University Press (in press).

Kistler, M., & Gnassounou, B. (Eds.). (2007). *Dispositions and causal powers*. Aldershot, UK: Ashgate.

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin and Review, 8*, 439–453.

Landau, I. (2010). *The locative syntax of experiencers*. Cambridge, MA: MIT Press.

Le Pelley, M., Griffiths, O., & Beesley, T. Associative accounts of causal cognition. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge, MA: Cambridge University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: a coherence hypothesis. *Cognitive Psychology, 40*, 87–137.

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

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955–982.

Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: the role of the prior context. *Cognition, 132*, 485–490.

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: dispositional intuitions as a guide to causal structure. *Cognitive Science, 39*, 65–95.

Mayrhofer, R., & Waldmann, M. R. (2016). Causal agency and the perception of force. *Psychonomic Bulletin and Review* (in press).

Mayrhofer, R., Quack, B., & Waldmann, M. R. (2016). *Causes and forces: A probabilistic force dynamics model of causal reasoning* (in preparation).
McCloskey, M. (1983). Naive theories of motion. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 229–324). Hillsdale, NJ: Erlbaum.
Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin and Review, 15*, 75–80.
Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory and Cognition, 37*, 249–264.
Meder, B., & Mayrhofer, R. Diagnostic reasoning. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.
Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review, 121*, 277–301.
Michotte, A. E. (1963). *The perception of causality*. New York: Basic Books.
Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology, 61*, 63–86.
Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. New York: Oxford University Press.
Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. *Cognitive Psychology, 67*, 186–216.
Paul, L. A., & Hall, P. (2013). *Causation: A user's guide*. New York: Oxford University Press.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan-Kaufmann.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
Perales, J. C., Catena, A., & Maldonado, A. (2004). Inferring non-observed correlations from causal scenarios: the role of causal knowledge. *Learning and Motivation, 35*, 115–135.
Perales, J., Catena, A., Cándido, A., & Maldonado, A. Rules of causal judgment: mapping statistical information onto causal beliefs. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.
Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530–535.
Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology, 72*, 54–107.
Rehder, B. Categories as causal models: Categorization. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.
Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264–314.
Riemer, N. (2010). *Introducing semantics*. Cambridge, UK: Cambridge University Press.
Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: inferences on causal networks. *Psychological Bulletin, 140*, 109–139.
Rottman, B. The acquisition and use of causal structure knowledge. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.
Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, 26*, 521–562.
Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs. *Psychological Bulletin, 121*, 192–218.
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120*, 411–437.
Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: inferences about hidden causes by 10-and 12-month-old infants. *Psychological Science, 16*, 995–1001.

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44*(A), 321–342.
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299–309.
Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: social cognition and language. *Journal of Personality and Social Psychology, 54*, 558–568.
Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications, and range. *European Review of Social Psychology, 2*, 1–30.
Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 405–415.
Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.
Sloman, S. A., Barbey, A. K., & Hotaling, J. (2009). A causal model theory of the meaning of "cause," "enable," and "prevent." *Cognitive Science, 33*, 21–50.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction and search*. New York: Springer.
Spohn, W. (2002). The many facets of the theory of rationality. *Croatian Journal of Philosophy, 2*, 247–262.
Spohn, W. (2012). *The laws of belief. Ranking theory and its philosophical applications*. Oxford, UK: Oxford University Press.
Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489.
Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100.
Waldmann, M. R., & Hagmayer, Y. (2005). Seeing vs. doing: two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 216–227.
Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.
Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222–236.
Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124*, 181–206.
Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *Causal learning: Vol. 34. The psychology of learning and motivation* (pp. 47–88). San Diego, CA: Academic Press.
Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 53–76.
Waldmann, M. R. (2007). Combining versus analyzing multiple causes: how domain assumptions and task context affect integration rules. *Cognitive Science, 31*, 233–256.
Waldmann, M. R. (2011). Neurath's ship: the constitutive relation between normative and descriptive theories of rationality. *Behavioral and Brain Sciences, 34*, 273–274.
Waldmann, M. R. (Ed.). *Oxford handbook of causal reasoning*, in press. New York: Oxford University Press.
Walsh, C. R., & Sloman, S. A. (2007). Updating beliefs with causal models: violations of screening off. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *A Festschrift for Gordon H. Bower* (pp. 345–358). New York: Erlbaum.
White, P. A. (2006a). The causal asymmetry. *Psychological Review, 113*, 132–147.

White, P. A. (2006b). The role of activity in visual impressions of causality. *Acta Psychologica, 123*, 166–185.
White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review, 116*, 580–601.
White, P. A. (2014). Perceived causality and perceived force: Same or different? *Visual Cognition, 22*, 672–703.
Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111.
Wolff, P. (2012). Representing verbs with force vectors. *Theoretical Linguistics, 38*, 237–248.
Wolff, P. (2014). Causal pluralism and force dynamics. In B. Copley, & F. Martin (Eds.), *Causation in grammatical structures* (pp. 100–119). Oxford, UK: Oxford University Press.
Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: how absences cause events. *Journal of Experimental Psychology: General, 139*, 191–221.
Wolff, P., & Shepard, J. (2013). Causation, touch, and the perception of force. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 58, pp. 167–202). New York: Academic Press.
Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276–332.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.
Woodward, J. (2011). Mechanisms revisited. *Synthese, 183*, 409–427.