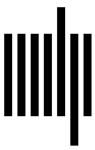
The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

7.2 The Rationality of Everyday Causal Cognition

Michael R. Waldmann

Summary

Normative and descriptive theories of causal cognition are tightly connected. Theories of causal cognition proposed in psychology have in most cases precursors in philosophy and other normative disciplines. For example, associative theories, causal Bayes net theories, and dispositional theories of causal cognition all are inspired by positions developed in philosophy. The competition between normative accounts is therefore mirrored in psychology. Findings that seem rational within one normative account may appear deficient within a competing theory. Consistency with a specific normative account is, however, only one criterion to assess rationality; to assess rationality, also correspondence of causal representations to causal relations in the world has been used. The present chapter will selectively review research, focusing on causal reasoning and learning, to demonstrate the tight coupling between normative and descriptive theories of everyday causal cognition.

1. Normative and Descriptive Theories

Causal cognition belongs to our most central cognitive competencies. Without causal knowledge, we would not be able to make predictions and diagnoses, generate explanations, design artifacts, solve problems, or plan and act. The ubiquity of causal cognition has attracted scientists 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, artificial intelligence, anthropology, biology, economics, physics, and statistics (among others). Causality is a genuinely interdisciplinary topic attracting both researchers interested in developing *normative methods* of causal analysis and researchers pursuing the *descriptive* goal of understanding how humans (in different age groups) and nonhuman animals reason about causal relations (for overviews, see Waldmann, 2017; Waldmann & Hagmayer, 2013).

Yet until recently, causal reasoning has been curiously absent from mainstream cognitive psychology. Although there has been some limited work in more specific areas, such as developmental or social psychology, it did not play a significant role in fundamental theories of human cognition. To date, textbooks on cognitive psychology do not contain chapters on causal reasoning. One reason for this neglect is that psychology was for many decades dominated by the view that cognitive mechanisms, such as associative learning, are general and can be specified independently from the domains to which they are applied. In many areas that addressed knowledge acquisition, domain-general similarity-based theories dominated, which neglected the role of causal knowledge (e.g., categorization, induction, conditional reasoning, learning). The situation has slowly changed in the past two decades, with more and more research devoted to the particular characteristics of reasoning and learning about causal systems. In fact, the relevance of causal knowledge has now been acknowledged in virtually all fields of higher-level cognition (see Waldmann, 2017). The breadth of this literature makes it necessary to restrict the focus in the present chapter, which will primarily discuss general theoretical frameworks and empirical studies on causal reasoning and learning.

Theories of causal cognition proposed in psychology have in most cases precursors in philosophy and other normative disciplines. It is not an accident that the goals of normative and descriptive theories of causality overlap. Both scientists and laypeople intend their causal theories to be correct. Causal claims are therefore generally associated with normative force (see Spohn, 2002; Waldmann, 2011). The common goals of normative and descriptive accounts of causal cognition may be the reason why psychologists often turn to normative theories as an inspiration for psychological theories. One of the most recent examples are *causal Bayes nets*, which were first developed in philosophy and engineering (see Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 2001; Spohn, 2012) but have also been adopted by psychologists as

models of everyday causal reasoning (for reviews, see Rottman, 2017; Rottman & Hastie, 2014; Waldmann, 2017; Waldmann & Hagmayer, 2013; see also chapter 7.1 by Pearl, this handbook).

Although scientists and laypeople have common goals, it is also plausible to expect differences in their methodological approaches and the resulting types of representations of causal knowledge. Causal domains can fundamentally differ, so that a method that has been developed to acquire knowledge in economics or sociology, for example, differs from methods employed in chemistry or physics. Causal Bayes nets (see chapter 4.2 by Hartmann, this handbook), which were intended as general models of causation, are more often used in the social sciences than in physics. Similarly, laypeople use causal knowledge in various everyday domains, including intuitive psychology, biology, and sociology. Again, although there are attempts to develop abstract theories of causality, such as Bayes nets, to model knowledge in these different areas, there are also more specific theories that focus on unique domain properties (e.g., intuitive physics or belief–desire psychology). Moreover, everyday knowledge is often rudimentary, erroneous, and driven by unsubstantiated beliefs (see Rozenblit & Keil, 2002; Sloman & Fernbach, 2017). Nevertheless, causal knowledge is probably adequate more often than not; otherwise, it would not be adaptive.

Another difference between normative and descriptive approaches is that philosophers and scientists interested in methodology generally try to develop a uniform coherent account based on few fundamental principles. By contrast, laypeople are often *satisficers* (Simon, 1956): they use tools that work for a given problem, but they are rarely interested in achieving coherence and consistency (see Arkes, Gigerenzer, & Hertwig, 2016; see also chapter 8.5 by Hertwig & Kozyreva, this handbook).

2. How Rational Is Everyday Causal Cognition?

Given that most psychological theories of causal cognition have been inspired by normative accounts, most research addresses, at least implicitly, the question of whether everyday causal cognition conforms to the normative principles developed in philosophy, statistics, or machine learning. However, the normative evaluation of causal cognition is complicated by the fact that there is no unique normative account of causation that is universally accepted. The competition within the normative discourse is reflected in psychological theories. For example, associative theories, which can be traced back to the philosopher David Hume (1748/1977), model causation as statistical covariation, whereas causal Bayes nets view associative approaches as deficient and argue that they neglect important specific features of causality. A further complexity is added by the fact that even within each of these frameworks, different competing normative and descriptive theories have been developed. Thus, depending on the normative view of the researcher, findings can be seen either as confirming or as violating norms (see the following sections for examples).

When researchers discover deviations from initially held normative accounts, different strategies are pursued. Sometimes the findings lead to modifications of the theory with the primary goal of maximizing *descriptive* adequacy (e.g., different versions of associative theories). At other times, the strategy is to develop a modified *normative* theory by adding features that had previously been neglected (e.g., sensitivity to uncertainty in Bayesian theories). These attempts seek to be both descriptively adequate and normative.

Whereas the debates about which norm is appropriate address rationality in terms of coherence with specific norms, another perspective on rationality focuses on whether causal cognition corresponds to objective relations in the world (the correspondence view of rationality; Arkes et al., 2016; see also chapter 8.5 by Hertwig & Kozyreva, this handbook). Thus, one important question is whether people tend to distort observations and whether possible distortions can be rationalized as adaptive. There is evidence showing that learners are often poor in correctly encoding probability information (Rottman & Hastie, 2014) and tend to over- or underestimate covariations, especially when these are zero (for a review, see Waldmann & Hagmayer, 2013). There have been attempts to argue that some of these distortions may actually be rational because they reflect the integration of prior knowledge, but not all cases can be explained this way.

3. Frameworks of Causal Cognition

The plurality of philosophical theories of causality is also reflected in psychological approaches. Various frameworks differ in how they model causality and causal cognition (see also Waldmann & Mayrhofer, 2016). I will use the term "framework" to describe classes of theories that use substantially different theoretical concepts to capture causality. Frameworks also differ in the tasks they are trying to model. In this section, I will describe different frameworks applied in causal reasoning research. The main distinguishing features of frameworks are the proposed causal *relata* (that is, the type of entities that

enter causal relations) and the way causal *relations* are modeled.

3.1 The Dependency Framework

The dependency view of causation is adopted by several psychological theories, which share central framework assumptions but otherwise often compete. Theories that can be subsumed under the dependency framework include associative theories (see Le Pelley, Griffiths, & Beesley, 2017), covariation theories (e.g., Cheng & Novick, 1992; Perales, Catena, Cándido, & Maldonado, 2017), power probabilistic contrast (PC) theory (Cheng, 1997), causal model theories (e.g., Gopnik et al., 2004; Rehder, 2017a, 2017b; Rottman, 2017; Sloman, 2005; Waldmann & Hagmayer, 2001; 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).

According to dependency theories, a variable *C* is a cause of variable *E* if *E* statistically depends upon *C*. There is an extensive debate in philosophy about the proper *causal relata* in dependency theories (e.g., events, propositions, facts, properties, or states of affairs; see Ehring, 2009; Spohn, 2012). *Causal relations* are often graphically depicted by causal arrows directed from cause to effect (see figure 7.2.1). For example, a causal model could be postulated that uses the binary variables representing the effect forest fire (present vs. absent) and as potential causes a lit match (e.g., dropped by an arsonist vs. not dropped) and a lightning (present vs. absent) (see the common-effect model in figure 7.2.1a). The dependencies, then, encode a set of hypothetical situations consistent with the causal model. Causal models can also express mechanism knowledge as chains or networks of mediating variables.

Representative research topics

Associative theories One early popular class of models of causal cognition are *associative theories*, which mainly focus on learning. Inspired by the philosopher David Hume (1748/1977), the key claim is that causal learning is based on associating repeatedly observed spatiotemporally contiguous event pairs. How exactly such associations arise, and which covariation measures they approximate, is a matter of great debate between competing associative theories (see Le Pelley et al., 2017; Perales et al., 2017).

Whereas Hume additionally assumed that causes and effects that enter associations are ordered in the temporal order of their occurrence in the world (i.e., causes precede their effects), psychological theories of association focus on the temporal order of learning cues and outcomes, which can refer either to causes or to effects. For example, for a physician, a cue could be an observed symptom and the outcome the diagnosis of a disease. In this case, the cue refers to an effect of its cause, the disease.

Associative theories versus causal model theory Whereas in associative theories, weights are acquired in the cue–outcome direction regardless of whether cues represent causes or effects, a crucial feature of causal models is that they contain arrows that represent causal directionality regardless of the order in which information about the variables is acquired (see figure 7.2.1). This crucial difference led to a debate in the 1990s about whether people are sensitive to causal directionality and use causal model representations or whether they disregard this crucial feature of causality and acquire associative knowledge in the cue–outcome direction, analogous to multiple regression analysis in statistics. Thus, this debate was about which normative principle people apply in everyday causal cognition. Waldmann and Holyoak (1992) presented evidence supporting the causal model perspective (see also Waldmann, 1996, 2000; Waldmann et al., 1995). In the meantime, much more evidence for causal model representations has been collected (e.g., Gopnik et al., 2004; Rehder, 2017a, 2017b; Sloman, 2005), although there is also some evidence showing that in complex learning tasks, learners sometimes fall back to simpler associative representations (see, e.g., López, Cobos, & Caño, 2005).

Estimating causal strength Another debate related to a normative discussion about the proper norm concerns how causal strength parameters should be estimated. Most associative theories assume that causal strength is equal to the observed covariation. A popular measure for single cause–effect relations, used both by rule-based theories and by associative theories such as the Rescorla–Wagner rule (Rescorla & Wagner, 1972), is *delta-p*, which is the difference between the probability of an effect when its cause is present and the probability of the effect

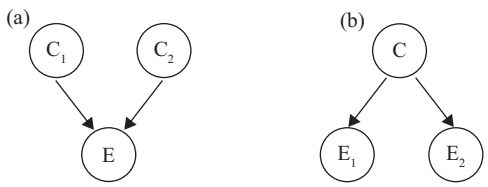


Figure 7.2.1
Two examples of causal Bayes nets: (a) common-effect model and (b) common-cause model.

when its cause is absent. Conceptually, this measure views causes as *difference makers*.

Although *delta-p* is an accepted covariation measure, Cheng (1997) has questioned its normative and psychological suitability as a measure of causal strength. Associative theories add up multiple associative weights when multiple causes are present, which may lead to the nonsensical prediction that an effect will occur with a probability higher than 1. Thus, possible ceiling effects are not accounted for. Another example of a problematic ceiling effect are cases in which the probability of the effect in the absence of the cause is already at maximum (i.e., 1) so that the cause cannot further increase the probability of the effect. Concluding from such an observation that the cause has zero strength seems counterintuitive.

Cheng (1997) has therefore suggested a new measure of causal strength, *causal power*, which represents the probability with which a cause, acting alone, generates or prevents a target effect. Given that causes are never observed alone, causal power is a theoretical quantity that needs to be inferred on the basis of observed data. Cheng's theory uses *delta-p* along with default background assumptions about the underlying causal model (e.g., independence of target cause and unobserved alternative causes) to infer the causal power parameter.

Cheng's (1997) theory is not the only theory of causal strength estimation that claims to be both normative and descriptively adequate. Cheng's goal was to develop a theory that generates point estimates of causal power. By contrast, Griffiths and Tenenbaum (2005) used the framework of Bayesian inference theory and analyzed causal inference in the context of their *causal support model*, which incorporates uncertainty of parameter estimates by considering distributions of parameters. Whereas Cheng's (1997) focus were judgments of causal strength, which were modeled as parameter estimation tasks, the causal support model focuses on the assessment of the likelihood of the presence of a causal link between a target cause *C* and an effect *E*. Griffiths and Tenenbaum argue that this is actually the question learners try to answer when asked to assess causal strength. The causal support model aims to contrast the evidence in favor of a causal model in which a link exists between *C* and *E* with a causal model in which these two variables are independent and *E* is only influenced by background causes. One notable feature of the model is that it can rationally explain why people sometimes offer positive numerical estimates when they have observed contingencies that are actually at zero.

Observing versus intervening A unique feature that sets causal Bayes nets apart from associative theories is their ability to predict the outcomes of interventions based on observational knowledge alone (Pearl, 2000; Spirtes et al., 2001; see also chapter 7.1 by Pearl, this handbook). Assume, for example, that we have observed that atmospheric pressure, barometer readings, and weather covary and that we have reason to believe that atmospheric pressure is the common cause of the two effects. Both causal Bayes nets and associative theories would allow us to make observational predictions between the three events. If we observe, for example, high barometer readings, we are licensed to predict good weather. However, without further observations, causal Bayes nets also allow us to correctly predict that intervening on the barometer (for example, by shaking it) and thus changing the reading would not alter the weather. According to Bayes nets accounts, we would represent such an action as the result of a *do*-operator (Pearl, 2000) that is the sole cause of the new reading, whereas the previous cause, atmospheric pressure, is now causally disconnected from the barometer. In the common-cause model, this would amount to removing the arrow between the atmospheric pressure variable and the barometer variable ("graph surgery"). Using this method, more complex planning situations can also be modeled (see Pearl, 2000). Associative theories, by contrast, can only predict that intervening on the barometer leads to outcomes that are different from those in situations in which the barometer is simply observed, when the learner has actually intervened in the past and observed the associated outcome (i.e., interventional learning); these theories cannot make correct interventional predictions based on observations alone.

The distinction between observing and intervening represents an interesting test case to distinguish between competing theories within the dependency framework. Therefore, a number of psychologists have investigated whether human subjects are capable of deriving interventional predictions from mere observations (Meder, Hagmayer, & Waldmann, 2008, 2009; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). The general finding in these studies is that participants proved capable of deriving interventional predictions from observations, although in more complex tasks, some participants seemed to confuse observational with interventional knowledge. Moreover, there are also some preliminary findings suggesting that rats, too, are capable of deriving interventional predictions from observations (Blaisdell, Sawa, Leising, & Waldmann, 2006; Leising, Wong, Waldmann, & Blaisdell, 2008).

Modeling causal queries Most studies on causal reasoning and learning have addressed how people make *predictive* inferences (from causes to effects) or assessments of causal strength. Causal knowledge representations, however, also allow for other types of queries. More recently, psychologists have developed computational models of various types of queries and have empirically tested these models. These studies are again a test case for normative theories that model how causal queries *should* be answered. For example, a set of studies on diagnostic reasoning by Meder et al. (2014) shows how deviations from previously held norms can lead to a modified normative theory that is also descriptively adequate (for overviews, see Meder & Mayrhofer, 2017; chapter 7.3 by Meder & Mayrhofer, this handbook).

Meder et al. (2014) discovered that participants who were asked to make a diagnostic judgment did not simply report the empirical probability of the cause given the effect, which has been considered the standard normative measure of diagnostic probability, but decrease the diagnostic estimates when causal strength in the predictive direction is low. Meder and colleagues showed that this behavior is in fact rational under a Bayesian inference view. They developed a modified Bayesian model that is not only sensitive to the observed diagnostic probability but also incorporates the degree of uncertainty about the underlying causal model in the judgments.

Other studies have addressed the question of how people make *singular* causal judgments (Cheng & Novick, 2005; Stephan & Waldmann, 2018; Stephan, Mayrhofer, & Waldmann, 2020). For example, it may have been observed that Peter, a heavy smoker, has contracted lung disease. It is known that generally heavy smokers tend to contract lung disease later in their lives. A singular query might in this case ask whether Peter's smoking actually caused *his* lung disease, in contrast to the possibility that in his case, the observed events are just a coincidence. The new model, which is again derived from normative accounts, tries to explain how people respond to such queries.

Acquisition of causal model knowledge One of the major engineering achievements of Bayes net research in computer science are algorithms that allow machines to induce causal networks from covariation data using minimal prior knowledge (e.g., constraint-based algorithms, Bayesian inference methods; Pearl, 2000; Spirtes et al., 2001). In psychology, there has been a debate whether these algorithms are plausible models of human learning (see Gopnik et al., 2004; Griffiths & Tenenbaum, 2009). One problem is that these algorithms typically require data sets of a size that clearly surpasses what human learners can process. Also, it is unclear whether humans are capable of making, without the aid of computers, the required precise estimates about conditional independence relations. Nevertheless, some psychologists have tried to test whether human subjects are capable of inducing causal models from covariation alone. In general, relatively simple tasks were presented in which a restricted set of causal models with a small number of variables (typically 3) were offered as candidates between which subjects were asked to choose. Nevertheless, performance was generally poor (Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). Some studies have shown that allowing subjects to intervene helps a little bit (Bramley, Lagnado, & Speekenbrink, 2015; Gopnik et al., 2004; Steyvers et al., 2003). The general conclusion from these findings is that in principle, people can induce causal models from covariation data, but performance is poor unless learning conditions are extremely favorable. This is an example of research showing deviations from what normative models can achieve.

The implausibility of domain-general algorithms of causal structure induction has led Waldmann (1996) to propose the view that people generally use prior knowledge about the structure of causal models to guide learning in a top-down fashion ("knowledge-based causal induction"; see also Lagnado, Waldmann, Hagmayer, & Sloman, 2007). Various cues, including temporal order, interventions, instructions, and prior knowledge, can guide the initial hypotheses about causal structure. Numerous studies have shown that learners are capable of using these cues and are able to prioritize them in cases of conflict.

Griffiths and Tenenbaum (2009) have proposed *hierarchical* Bayesian inference strategies to model knowledge- or theory-based causal induction. The central assumption is that probabilistic inference is carried out at multiple levels of abstraction, which influence each other and are updated simultaneously. In causal learning, these levels include the data, alternative causal models, and the theory level, which may, for example, encode physical domain knowledge. Hierarchical Bayesian models speed up learning because the range of alternative hypotheses being considered is constrained by both the data and the theory level.

Whereas theories combining top-down knowledge with bottom-up learning have initially focused on fairly abstract causal knowledge, such as knowledge about causal directionality, more recent accounts have incorporated specific domain knowledge. 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 combine (see also Griffiths & Tenenbaum, 2009).

Following up on earlier work demonstrating the role of temporal assumptions about cause–effect latencies (Buehner & May, 2003; Hagmayer & Waldmann, 2002), more sophisticated models of the influence of time on causal induction have been proposed (Bramley, Gerstenberg, Mayrhofer, & Lagnado, 2018; Stephan et al., 2020). Other approaches, postulating an integration of intuitive Newtonian physics with probabilistic inference in causal models ("noisy Newton"), were offered by Gerstenberg and Tenenbaum (2017) and Sanborn, Mansinghka, and Griffiths (2013; see also Bramley, Gerstenberg, Tenenbaum, & Gureckis, 2018).

Violations of Bayes net predictions Research about causal Bayes nets has initially found a surprisingly good fit between their theoretical assumptions and human causal inference (see Rottman & Hastie, 2014). However, more recently, deviations have been discovered, which led to attempts to modify the theory.

Diagnostic reasoning was already discussed above as an example of how deviations from the predictions of normative theories may lead to a revised account that claims to be a better theory of diagnostic reasoning, both descriptively and normatively. Another popular topic questioning Bayes nets as a psychological account are apparent Markov violations, first discovered by Rehder and Burnett (2005). They showed, for example, that subjects who were presented with a common-cause model (see figure 7.2.1b) and then were asked to infer the probability of an effect given its cause were influenced by how many other effects of this cause were present or absent. *Prima facie*, this finding violates the Markov constraint, which is assumed to be a central feature of Bayes nets. According to this constraint, the requested cause–effect inference should be independent of whether collateral effects are present or absent.

Whereas some philosophers use this kind of evidence as a demonstration of the inadequacy of Bayes nets (e.g., Cartwright, 2001), cognitive psychologists chose a similar strategy as Meder et al. (2014) and tried to show that Markov violations are only apparent violations of the normative account. They proposed extended Bayes net models with hidden variables, which explain the apparent Markov violations within normative Bayes nets that honor the Markov condition. Various theories have been proposed to motivate these extensions, for example, by postulating additional mechanism knowledge (Park & Sloman, 2013) or intuitions about dispositions (Mayrhofer & Waldmann, 2015). Other accounts combine Bayes net representations with associative processes (Rehder, 2014; Rehder & Waldmann, 2017).

3.2 The Disposition Framework

A completely different view answers the question of why an observed causal relation between events holds by focusing on the participants involved in the causal interaction. An often-studied example are the two animated colliding billiard balls in Michotte's (1963) task.

The normative status of dispositional theories can be debated. Some specific accounts were motivated by Aristotelian philosophy, which for centuries was viewed as rational but has been demonstrated to contradict modern physics (White, 2006). More recently, however, versions of dispositional theories have been developed that claim that these theories provide a better account of rational scientific research than competing views in philosophy of science (Cartwright & Pemberton, 2013; Mumford & Anjum, 2011).

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 may be humans or nonhuman entities. The dispositions of objects can be static (e.g., the solubility of sugar) or they can be transient and dynamic such as the sudden exertion of force when moving an object. Causal relations between events arise when objects are placed in specific situational contexts allowing them to express their dispositions or powers. Thus, dispositional theories are looking for deeper explanations of observed dependencies underlying the observed covariation. One can view this as a focus on underlying mechanisms; however, the mechanisms have different properties from mechanisms modeled within the dependency framework (e.g., as chains or networks of variables; see Stephan, Tentori, Pighin, & Waldmann, 2021).

Dispositional theories can refer to very specific dispositions (e.g., of aspirin; see Mumford & Anjum, 2011), but the accounts most popular in psychology typically distinguish between just two types of objects, often called *causal agents* and *causal patients*. A popular theory, which was initially developed in linguistic semantics, is *force dynamics* (see Gärdenfors, 2014; Talmy, 1988). For example, Gärdenfors (2014) analyzes causal scenarios as interactions between a causal agent, endowed with a force, and a causal patient, which can be an animate or inanimate, concrete or abstract object. The patient can

carry a counterforce resisting the action of the agent. Forces are primarily physical, but they can be extended metaphorically to social or mental forces (e.g., threats, commands, persuasions). Force dynamics has been used in linguistics to characterize verb semantics and argument structure. For example, in "Peter hits Mary," "hit" has two arguments, one describing an agent (Peter), the other the patient (Mary).

Wolff (2007) has developed psychological variants of force dynamics (see also Wolff, Barbey, & Hausknecht, 2010; for an overview, see Wolff & Thorstad, 2017). His force theory states that people evaluate configurations of forces attached to what Wolff calls "affectors" (i.e., agents) and "patients," which may vary in direction and degree with respect to an end state.

Representative research topics Numerous studies on force dynamics have been conducted in the context of linguistic semantics. My focus here will be on selected findings from psychology.

Understanding causal terms Wolff (2007) used force theory to analyze the meaning of abstract causal concepts, such as *cause*, *prevent*, *enable*, and *despite* (see also Wolff & Song, 2003). For example, when we say, "High winds caused the man to move toward the bench," we mean that the patient (the man) had no tendency to move toward the bench, the affector (i.e., agent; the wind) acted against the patient, and the resultant of the forces acting on the patient was directed toward the result of moving toward the bench. This constellation of forces describes how we understand the term *cause*. Related analyses were proposed for the other abstract causal terms. Wolff and colleagues tested their theory, for example, by presenting visual scenes to subjects depicting stationary or moving people or objects and asking subjects to choose between alternative causal terms (see Wolff, 2007; Wolff & Thorstad, 2017).

Initially, the theory was applied to individual causal relations between agents and patients. Later, the theory was extended to more complex scenarios, such as causal chains or complex preventive relations, such as double preventions. For example, consider somebody hitting a pole that prevents a tent from falling. The action (hitting the pole) in this case prevents a preventer (the pole keeping the tent upright), which leads to the falling of the tent. Similarly, causation by omission can be modeled by considering the forces that are in play when an action is not executed (see Wolff et al., 2010; Wolff & Barbey, 2015).

Causal asymmetry Another example of a dispositional account (although different from force dynamics) is White's (2006) theory. Many of White's studies have focused on the Michotte (1963) task, in which subjects observe two-dimensionally rendered animated moving balls. 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* begins moving, eliciting a causal impression. White (2009) has shown that subjects tend to view the launching ball (*X*) as the agent and the launched ball (*Y*) as the patient (or "cause" and "effect object," in his terminology). Subjects typically describe the launching scene as an event in which "*X* launched *Y*" instead of using the equally valid description that "*Y* stopped *X*." Moreover, force estimates for *X* tend to be higher than force estimates for *Y*. According to White, both findings are indicators of the underlying dispositional distinction between the two types of objects. 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 (i.e., "*X* launched *Y*" and "*Y* stopped *X*") should be equally appropriate (see also Mayrhofer & Waldmann, 2014, 2016, for follow-up studies).

White (2006, 2009) has proposed a dispositional theory of causal asymmetry that links perceived scenes to stored representations of sensomotoric experiences of our actions on objects. According to White, we experience our own agency along with its associated forces during the course of our ontogenetic development. When we perceive a scene, we compare the movements of the objects with these stored representations. We tend to overestimate the force of the causal agent (i.e., cause object) relative to the counterforce of the manipulated patient (i.e., effect object) because this asymmetry corresponds to our experience of resistance that is overcome by our action.

4. Conclusion

Despite many attempts to develop a unitary theory of causal reasoning, this review of the literature shows that fundamentally different frameworks compete. Each framework has philosophical precursors that provide an initial normative basis, but within each framework, these accounts have been further developed with the goal to provide both descriptively more adequate theories and, at least in some cases, also improved normative accounts. Given that both the descriptive and the normative theories developed within and across the

frameworks compete, no uniform normative or descriptive theory of causal cognition has emerged.

The frameworks differ in terms of the causal relata they employ and the way causal relations are construed. These differences make them more or less suitable for modeling different tasks. For example, dependency theories are particularly good at modeling how people predict events within complex causal models, whereas dispositional theories are typically applied to linguistic phrases or visual scenarios involving a limited number of participants. These specializations may be the reason why the different frameworks rarely compete directly with each other. Most of the debates in the causal cognition literature concern competing theories *within* each framework. There are some limited attempts to cross the aisle and develop unitary frameworks that also apply to the tasks of the competitor (Cheng, 1993; Wolff, 2014), but these attempts have so far not been widely adopted by the research community.

As a consequence, other proposals have been presented, which try to address the issue of competing frameworks. One position is *causal pluralism*, which accepts that different tasks may be modeled best by different types of theories (e.g., Lombrozo, 2010). Another approach is to develop *hybrid* theories that focus on how different kinds of causal representations interact in specific tasks. An example of a hybrid theory that combines dependency with dispositional representations has been presented by Mayrhofer and Waldmann (2015; see also Waldmann & Mayrhofer, 2016). They claim that verbal instructions, whose causal content can be best modeled by dispositional accounts, influence how causal Bayes nets are construed that are used to make causal inferences. These are just initial attempts. Future research will have to address in greater detail whether unitary, pluralistic, or hybrid accounts are best suited to model causal cognition.

Acknowledgments

This research was supported by Deutsche Forschungsgemeinschaft (DFG) Grant WA 621/24-1. I am grateful to Simon Stephan for helpful comments.

References

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

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*, 1020–1022.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., & Lagnado, D. A. (2018). Time in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1880–1910.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, *105*, 9–38.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 708–731.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology Section A*, *56*, 865–890.

Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist*, *84*, 242–264.

Cartwright, N., & Pemberton, J. M. (2013). Aristotelian powers: Without them, what would modern science do? In R. Groff & J. Greco (Eds.), *Powers and capacities in philosophy: The new Aristotelianism* (pp. 93–112). New York, NY: 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, NY: 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.

Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review*, *112*, 694–706.

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

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

Gerstenberg, T., & Tenenbaum, J. (2017). Intuitive theories. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 515–547). New York, NY: 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.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*, 1128–1137.

Hume, D. (1977). *An enquiry concerning human understanding*. Indianapolis, IN: Hackett. (Original work published 1748)

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. E. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford, England: Oxford University Press.

Leising, K. J., Wong, J., Waldmann, M. R., & Blaisdell, A. P. (2008). The special status of actions in causal reasoning in rats. *Journal of Experimental Psychology: General*, *137*, 514–527.

Le Pelley, M. E., Griffiths, O., & Beesley, T. (2017). Associative accounts of causal cognition. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 13–28). New York, NY: Oxford University Press.

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

López, F. J., Cobos, P. L., & Caño, A. (2005). Associative and causal reasoning accounts of causal induction: Symmetries and asymmetries in predictive and diagnostic inferences. *Memory & Cognition*, *33*, 1388–1398.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955–984.

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 & Review*, *23*, 789–796.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & 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 & Cognition*, *37*, 249–264.

Meder, B., & Mayrhofer, R. (2017). Diagnostic reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 433–457). New York, NY: 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, NY: Basic Books.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. New York, NY: 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.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.

Perales, J., Catena, A., Cándido, A., & Maldonado, A. (2017). Rules of causal judgment: Mapping statistical information onto causal beliefs. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 29–51). New York, NY: Oxford University Press.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, *72*, 54–107.

Rehder, B. (2017a). Categories as causal models: Categorization. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 347–375). New York, NY: Oxford University Press.

Rehder, B. (2017b). Categories as causal models: Induction. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 377–413). New York, NY: Oxford University Press.

Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264–314.

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*, *45*, 245–260.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rottman, B. M. (2017). The acquisition and use of causal structure knowledge. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 85–114). New York, NY: Oxford University Press.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*, 109–139.

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*, 411–437.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.

Sloman, S., & Fernbach, P. (2017). *The knowledge illusion: Why we never think alone*. New York, NY: Riverhead Books.

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

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction and search*. New York, NY: 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, England: Oxford University Press.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2020). Time and singular causation—a computational model. *Cognitive Science*, *44*, e12871.

Stephan, S., Tentori, K., Pighin, S., & Waldmann, M. R. (2021). Interpolating causal mechanisms: The paradox of knowing more. *Journal of Experimental Psychology: General*. Advance online publication.

Stephan, S., & Waldmann, M. R. (2018). Preemption in singular causation judgments: A computational model. *Topics in Cognitive Science*, *10*, 242–257.

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. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (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.). (2017). *The Oxford handbook of causal reasoning*. New York, NY: Oxford University Press.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, *82*, 27–58.

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.), *The Oxford handbook of cognitive psychology* (pp. 733–752). New York, NY: 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., & Mayrhofer, R. (2016). Hybrid causal representations. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 65, pp. 85–127). New York, NY: Academic Press.

White, P. A. (2006). The causal asymmetry. *Psychological Review*, *113*, 132–147.

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

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

Wolff, P. (2014). Causal pluralism and force dynamics. In B. Copley & F. Martin (Eds.), *Causation in grammatical structures* (pp. 100–118). New York, NY: Oxford University Press.

Wolff, P., & Barbey, A. K. (2015). Causal reasoning with forces. *Frontiers in Human Neuroscience*, *9*, 1–21.

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

Wolff, P., & Thorstad, R. (2017). Force dynamics. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 147–167). New York, NY: Oxford University Press.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455