CHAPTER 23

Diagnostic Reasoning

Björn Meder and Ralf Mayrhofer

Abstract

This chapter discusses diagnostic reasoning from the perspective of causal inference. The computational framework that provides the foundation for the analyses—probabilistic inference over graphical causal structures—can be used to implement different models that share the assumption that diagnostic inferences are guided and constrained by causal considerations. This approach has provided many critical insights, with respect to both normative and empirical issues. For instance, taking into account uncertainty about causal structures can entail diagnostic judgments that do not reflect the empirical conditional probability of cause given effect in the data, the classic, purely statistical norm. The chapter first discusses elemental diagnostic inference from a single effect to a single cause, then examines more complex diagnostic inferences involving multiple causes and effects, and concludes with information acquisition in diagnostic reasoning, discussing different ways of quantifying the diagnostic value of information and how people decide which information is diagnostically relevant.

Key Words: diagnostic reasoning, causal inference, uncertainty, causal structure, probabilistic inference

Diagnostic reasoning is ubiquitous in everyday life. A physician diagnoses diseases from observed symptoms. An engineer engages in diagnostic reasoning when trying to identify what caused a plane to crash. A cognitive scientist reasons diagnostically when figuring out if an experimental manipulation proved successful in an experiment that did not yield any of the expected outcomes. A judge makes a diagnostic inference when reasoning how strongly a piece of evidence supports the claim that the defendant has committed the crime. More generally, diagnostic reasoning concern inferences from observed effects to (as yet) unobserved causes of these effects. Thus, diagnostic reasoning usually involves a kind of backward inference, as people typically infer (often unobserved) conditions that existed prior to what they have observed (in contrast to predictive reasoning from causes to effects, which is a kind of forward inference from present conditions or events into the future). Diagnostic reasoning from effect to cause can, therefore, be conceptualized as a special case of inductive inference, in which a datum *e*, the observed effect, is used to update beliefs about a hypothesis *c*, the unobserved target cause of the effect.

Diagnostic reasoning, as discussed in this chapter, needs to be differentiated from other related kinds of inference. Diagnostic reasoning is tightly connected to explanatory reasoning (see Lombrozo & Vasilyeva, Chapter 22 in this volume) and abductive reasoning (Josephson & Josephson, 1996), as all are concerned with reasoning about the causes of observed effects. However, the scope and aim differ in that both explanatory and abductive reasoning are broader and less constrained. In diagnostic reasoning, as we define it here, the set of potential causes is fixed and known; the target inference is about the presence of (one of) these causes (with the potential goal of an intervention on these causes). In abductive reasoning, by contrast, the set of variables the inference operates on is often not known a priori and has to be actively constructed. In explanatory reasoning, the target of the inference

9780199399550_Waldmann290716MEDUS_Book.indb 433

is the explanation of the observed effect by means of its causes; the diagnostic inference may be part of it, but other considerations play a role as well (see Lombrozo & Vasilyeva, Chapter 22 in this volume).

In this chapter, we discuss diagnostic reasoning from the perspective of probabilistic causal inference. Pearl (2000), Spirtes, Glymour, and Scheines (1993), and Spohn (1976/1978; as cited in Spohn, 2001) laid the foundations with the development of causal Bayes nets theory, which provides a comprehensive modeling framework for a formal treatment of probabilistic inference over causal graphical models. This computational framework has been used to address several theoretical and empirical key issues within a unified account (for overviews, see Rottman & Hastie, 2014; Waldmann & Hagmayer, 2013; Waldmann, Hagmayer, & Blaisdell, 2006; see also chapters in this volume by Cheng & Lu [Chapter 5]; Griffiths [Chapter 7]; Oaksford & Chater [Chapter 19]; Rehder [Chapters 20 and 21]; and Rottman [Chapter 6]). Examples include the formal analysis of different measures of causal strength (Griffiths & Tenenbaum, 2005; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008), the distinction between inferences based on observations and interventions (Lagnado & Sloman, 2004; Meder, Hagmayer, & Waldmann, 2008, 2009; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005), categorization (Rehder, 2003, 2010; Waldmann, Holoyak, & Fratianne, 1995), causal structure learning (Bramley, Lagnado, & Speekenbring, 2015; Coenen, Rehder, & Gureckis, 2015; Mayrhofer & Waldmann, 2011, 2015a, 2015b; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003), and analogical reasoning in causal domains (Holyoak, Lee, & Lu, 2010).

The framework of probabilistic inference over causal graphical models has also provided new pathways for the formal analysis of diagnostic reasoning in causal domains. Several computational models of diagnostic inference have been proposed that differ in their theoretical assumptions, technical implementation, and empirical scope (Fernbach, Darlow, & Sloman, 2011; Meder, Mayrhofer, & Waldmann, 2014; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008).

The remainder of this chapter is structured as follows. We first consider the case of elemental diagnostic reasoning, based on a single causal relation between two events (i.e., cause and effect). We discuss different computational models of elemental diagnostic reasoning, the issues they address, and their role in empirical research as descriptive or normative models. In the second part of this chapter, we discuss more complex cases of diagnostic inferences, involving multiple causes or effects, from both a theoretical and an empirical perspective. The third section highlights different ways of quantifying the diagnostic value of information and how people decide which information is diagnostically relevant. We conclude by discussing key questions for future research and by outlining pathways for developing an empirically grounded and normatively justified theory of diagnostic causal reasoning.

Elemental Diagnostic Reasoning

()

In this section, we focus on the most basic type of diagnostic causal reasoning, which concerns an inference from a single binary effect to a single binary cause. We refer to this kind of diagnostic inference as *elemental diagnostic reasoning*. Although this most basic type of diagnostic inference seems quite simple compared with real-world scenarios involving a complex network of multiple causes and multiple effects, it highlights a number of critical questions about both how people should reason diagnostically (i.e., what would constitute an adequate *normative model*) and how people in fact do reason diagnostically (i.e., what would constitute an adequate *descriptive model*).

In the following, we provide an overview of alternative models of diagnostic inference from a single effect to a single cause and the empirical studies that have been used to test the respective models. These accounts provide computational-level models (in Marr's, 1982, terminology), in that they specify the cognitive task being solved, the information involved in solving it, and the rationale by which it can be solved (Anderson, 1990; Chater & Oaksford, 1999, 2008; for critical reviews, see Brighton & Gigerenzer, 2012; M. Jones & Love, 2011). Our goals are to highlight the ways in which a causal inference perspective provides novel insights into the computational analysis of diagnostic reasoning and to discuss how different models have informed empirical research.

Simple Bayes: Diagnostic Reasoning with Empirical Probabilities

When reasoning from effect to cause, for instance, when assessing the probability of a particular disease given the presence of a symptom, it seems natural to estimate the conditional probability of a cause given the effect. A critical question is how exactly this diagnostic probability is inferred. Many researchers have endorsed Bayes's rule applied to the empirical probabilities as the natural

(�)

normative-and potentially descriptive-model for computing the diagnostic probability.

Let random variables C and E denote a binary cause and binary effect, respectively, and let $\{c, \neg c\}$ and $\{e, \neg e\}$ indicate the presence and absence of the cause and the effect event (Figure 23.1 a). Consider a physician examining a sample of 40 patients. Each patient has been tested for the presence of a certain genetic predisposition (cause event C) and the presence of elevated blood pressure (effect event E). This set of observations forms a joint frequency distribution over C and E, which can be represented in a 2×2 contingency table (Figure 23.1 b). The conditional probability of the cause given the effect (i.e., genetic predisposition given elevated blood pressure), P(c|e), can be inferred by using Bayes's rule:

$$P(c|e) = \frac{P(e|c) \cdot P(c)}{P(e)}$$
$$= \frac{P(e|c) \cdot P(c)}{P(e|c) \cdot P(c) + P(e|\neg c) \cdot P(\neg c)}$$
(1)

where P(c) denotes the prior probability (base rate) of the cause [with $P(\neg c) = 1 - P(c)$], P(e|c)is the likelihood of the effect conditional on the presence of the cause, and $P(e|\neg c)$ is the likelihood

(�

of the effect in the absence of the cause. For the data shown in Figure 23.1 b, the corresponding (frequentist) estimates are P(c) = 20/40 = .5, $P(e|c) = 6/20 = .3, P(e|\neg c) = 2/20 = .1, \text{ and } P(e) = 8/20$ 40 = .2. Plugging these numbers into Equation 1 yields P(c|e) = .75.

An alternative way of computing the diagnostic probability is to estimate it directly from the observed joint frequencies, the number of cases in which both C and E are present, N(c, e), and the number of cases in which C is absent and E is present, $N(\neg c, e)$:

$$P(c|e) = \frac{N(c,e)}{N(c,e) + N(\neg c,e)} \quad (2)$$

For the data shown in Figure 23.1 b, this computation yields the same result as applying Bayes's rule: P(c|e) = 6/(6+2) = .75.

Under the simple Bayes account, no reference is made to the causal processes that may have generated the observed data, and no uncertainty regarding the probability estimates is incorporated in the model. This model is strictly non-causal in that it can be applied to arbitrary hypotheses and data; whether these events refer to causes or effects does not matter (Waldmann & Hagmayer, 2013).

(a) 2 × 2 Contingency table				o) Example data			
	Effect present (e)	Effect absent (¬e)			Effect present (e)	Effect absent $(\neg e)$	
Cause present (c)	N(c, e)	$N(c, \neg e)$		Cause present (c)	6	14	
Cause absent (¬c)	$N(\neg c, e)$	$N(\neg c, \neg e)$		Cause absent (¬c)	2	18	
(c) Causal structure S_1				(d) Causal structure S ₀			
A				A			
b_c C w_c E				w_a			

Figure 23.1 (a) A 2 × 2 contingency table for representing the joint frequency distribution of a binary cause, $C = \{c, \neg c\}$, and a binary effect, $E = \{e, \neg e\}$. (b) Example data set. Numbers denote frequencies of co-occurrence (e.g., cause and effect were both present in 6 of 40 cases). (c) Causal structure hypothesis S_1 , the default causal model in power PC theory (b_c = prior probability of cause C; w_c = causal strength of C; w_a = strength of background cause A). (d) Causal structure hypothesis S_0 , according to which C and E are independent variables, that there is no causal relation between candidate cause C and candidate effect $E(b_{c} = \text{prior probability of cause})$ *C*; w_a = strength of background cause *A*).

MEDER AND MAYRHOFER 435

EMPIRICAL STUDIES

The simple Bayes model has a long-standing tradition in research on elemental diagnostic reasoning in a broader sense. Starting roughly in the 1950s, psychologists began using this model as a normative, and potentially descriptive, account of sound probabilistic reasoning. The most common tasks involved book bag and poker chip (or urn) scenarios with a well-defined statistical structure (e.g., Peterson & Beach, 1967; Phillips & Edwards, 1966). A key question was whether and to what extent people's intuitive belief revision would correspond to the prescriptions of Bayes's rule. Many studies found that subjects did take into account the diagnostic impact of the observed data, but to a lesser extent than prescribed by Bayes's rule (a phenomenon referred to as conservatism; Edwards, 1968). By and large, however, the conclusion was that people have good statistical intuitions, leading to the metaphor of "man as intuitive statistician" (Peterson & Beach, 1967).

With the advent of the heuristics and biases program (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974), research on probabilistic inference and elemental diagnostic reasoning continued. However, the studies conducted within this program led to a very different view of people's capacity for making sound diagnostic inferences. Findings from scenarios such as the lawyer-engineer problem (Kahneman & Tversky, 1973), the cab problem (Bar-Hillel, 1980), and the mammography problem (Eddy, 1982) seemed to indicate that people's judgments are inconsistent with Bayes's rule and generally are biased and error prone. Specifically, it was argued that people tend to neglect base rate information (i.e., the prior probability of the hypothesis) when reasoning diagnostically. In the mammography problem, for example, people were asked to give a diagnostic judgment regarding the posterior probability of breast cancer, based on a verbal description of the prior probability of the disease, P(c), the likelihood of obtaining a positive test result for a woman who has cancer, P(e|c), and the likelihood of a positive test result for a woman who does not have cancer, $P(e|\neg c)$. For instance, people were told that the prior probability of breast cancer is 1%, the likelihood of having a positive mammogram given cancer is 80%, and the probability of having a positive test result given no cancer is 9.6% (e.g., Gigerenzer & Hoffrage, 1995). Given these numbers, the posterior probability of breast cancer given a positive mammogram is about 8%. In stark contrast, a common finding was that

436 | DIAGNOSTIC REASONING

people's diagnostic judgments of the probability of breast cancer given a positive mammogram were often much higher than Bayes's theorem suggests (often around 70%–80%), which was explained by assuming that people do not take into account the low prior probability of having breast cancer in the first place.

However, the claim that people neglect base rate information on a regular basis is too strong. Koehler (1996; see also Barbey & Sloman, 2007) critically reviewed the literature, concluding that there are a variety of circumstances under which base rates are appreciated. One important factor is the way in which probabilistic information is presented (e.g., specific frequency formats vs. conditional probabilities), which can facilitate or impede people's sensitivity to base rate information when making diagnostic inferences. Gigerenzer and Hoffrage (1995; see also Sedlmeier & Gigerenzer, 2001) provided the information in the mammography problem and several other problems as natural frequencies (i.e., the joint frequencies of cause and effect, such as the number of women who have cancer and have a positive mammogram). Providing information this way facilitates derivation of the diagnostic probability because Equation 2 can be used and base rate information does not need to be introduced via Bayes's rule. These findings served as starting point for identifying and characterizing the circumstances under which base rate information is utilized and have informed more applied issues, such as risk communication in medicine (for a review, see Meder & Gigerenzer, 2014).

The question of whether and to what extent people use base rate information has been the focus of many studies on elemental diagnostic reasoning. In contrast, the relation between causal inference and elemental diagnostic reasoning has received surprisingly little attention in the literature, with respect to both normative and descriptive issues. Ajzen (1977) noted that "people utilize information, including information supplied by population base rates, to the extent that they find it possible to incorporate the information within their intuitive theories of cause and effect" (p. 312). At that time, however, the necessary tools for a formal treatment of diagnostic reasoning in terms of causal inference were not yet available, so that the exact nature of the interplay between diagnostic reasoning and causal representations was left largely unspecified (see also Tversky & Kahneman, 1982a, 1982b). Recent theoretical advances in causal modeling have made it possible to address this issue in a more rigorous way.

Power PC Theory: Diagnostic Reasoning Under Causal Power Assumptions

In contrast to the simple Bayes account, Cheng's (1997) power PC theory separates the data level (i.e., covariation information) from estimates of causal power that refer to the underlying but unobservable causal relations. The theory assumes that people aim to infer causal strength estimates because one goal of cognitive systems is to acquire knowledge of stable causal relations, rather than arbitrary statistical associations in noisy environments.

The theoretical assumptions underlying the power PC model instantiate a particular generative causal structure known as a noisy-OR gate (Glymour, 2003; Pearl, 1988): a common-effect structure with an observable effect *E* and two causes, namely an observable cause C and an amalgam of unobservable background causes A, which can independently bring about the effect (graph S_1 in Figure 23.1 c). The original version of the power PC model (Cheng, 1997) is equivalent to estimating the probability of C bringing about E (i.e., causal power) in causal structure S_1 using maximum likelihood estimates (MLEs) for the parameters derived from the relative frequencies in the data (see Griffiths & Tenenbaum, 2005, for a formal proof). An estimate for the strength of the background cause A, denoted w_{a} , is given by $P(e|\neg c)$ in the sample data, as the occurrence of E in the absence of C necessarily has to be attributed to some (unknown) background cause or causes (for mathematical convenience, A is assumed to be constantly present; Cheng, 1997; Griffiths & Tenenbaum, 2005). The observed rate of occurrence of C in the sample, P(c), provides an estimate of the base rate of C, denoted b_c . The unobservable probability with which C produces E, its generative causal power, is denoted w_c (see Cheng, 1997, for analogous derivations for preventive causal power). This estimate of causal strength is computed from P(e|c) by partializing out the influence of the background causes that may also have generated the effect (Cheng, 1997).¹ It can be estimated from the observed relative frequencies by

(�

$$w_{c} = \frac{P(e|c) - P(e|\neg c)}{1 - P(e|\neg c)} \quad (3)$$

Waldmann and colleagues (2008) showed how diagnostic inferences can be modeled in the power PC framework, that is, using the parameters of causal structure S_1 . Given the causal structure's parameters and a noisy-OR parameterization, the

diagnostic probability of candidate cause c given an effect e is given by

$$P(c|e) = \frac{P(e|c) \cdot P(c)}{P(e|c) \cdot P(c) + P(e|\neg c) \cdot P(\neg c)} \quad (4)$$
$$= \frac{w_c b_c + w_a b_c - w_c w_a b_c}{w_c b_c + w_a - w_c w_a b_c}.$$

If this diagnostic inference is based on maximum likelihood point estimates directly derived from the observed frequencies, the power PC model yields the same numeric predictions as the simple Bayes approach. For instance, for the data set shown in Figure 23.1 b, the standard power PC account predicts that $P(c|e) = .75.^2$ Thus, although the inference operates on the causal rather than the data level, the two accounts make the same prediction, namely, that diagnostic judgments should reflect the empirical conditional probability of the cause given the effect in the sample data.

A Bayesian variant of the power PC model can be implemented by associating prior distributions with the parameters of structure S_1 and updating the parameter distributions in light of the available data via Bayesian updating (Holyoak et al., 2010; Lu et al., 2008). In this case, the predictions of the power PC model do not necessarily correspond to the simple Bayes model, with the specific differences varying as a function of the prior and sample size used (see Meder et al., 2014, for a detailed discussion and example predictions). Bayesian variants of the power PC account allow it to incorporate prior knowledge and expectations of the reasoner into the diagnostic inference task via specific priors over the parameters of structure S_1 (Lu et al., 2008) and are also able to quantify (via distributions over parameters) the amount of uncertainty associated with the parameter estimates of structure S_1 .

EMPIRICAL STUDIES

Krynski and Tenenbaum (2007; see also Hayes, Hawkins, & Newell, 2015) studied the role of causal structure in elemental diagnostic reasoning tasks designed to investigate the use of base rate information, such as the mammography problem (Eddy, 1982; Gigerenzer & Hoffrage, 1995). The question they were interested in was whether people's diagnostic inferences would be mediated by the match between the provided statistics and the causal representations that people construct from the task information (e.g., a causal structure with one observed and one unobserved cause, or a causal

9780199399550_Waldmann290716MEDUS_Book.indb 437

 (\bullet)

structure with two observed causes). According to their experiments, when the given statistics can be clearly mapped onto the structure of the respective mental causal model, people's diagnostic inferences are more sensitive to normatively relevant variables, such as base rate information. For instance, if in the mammography problem an explicit cause for the false positive rate is provided (e.g., a benign cyst that can also cause a positive mammogram), people's diagnostic judgments improve substantially relative to the standard version of the problem in which no causal explanation for the false positive rate is provided.

In follow-up research, McNair and Feeney (2014; see also McNair & Feeney, 2015) explored the role of individual differences. They assessed people's numeracy, that is, the ability to perform elementary mathematical operations (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Lipkus, Samsa, & Rimer, 2001). According to their results, clarifying the causal structure among the domain variables seems helpful only for participants with high numeracy skills; the performance of participants with low numeracy did not improve.

Fernbach, Darlow, and Sloman (2010, 2011) investigated to what extent people consider the influence of alternative causes in diagnostic reasoning from effect to cause, compared with predictive reasoning from cause to effect. They used a simple causal Bayes net equivalent to the power PC model (structure S_1 ; Figure 23.1 c) as the normative benchmark for people's predictive and diagnostic inferences. To derive model predictions for different real-world scenarios, they elicited participants' existing causal beliefs about the relevant quantities, that is, the parameters associated with structure S_1 (base rate b_{j} , causal strength w_{j} , and strength of alternative causes, w_{j}). For instance, Fernbach and colleagues (2011) asked participants to estimate the prior probability that a mother of a newborn baby is drug addicted, how likely it is that the mother's drug addiction causes her baby to be drug addicted, and how likely a newborn baby is to be drug addicted if the mother is not. These estimates were then used to derive model predictions for predictive and diagnostic inferences (e.g., estimates for how likely a baby is to be drug addicted given that the mother is drug addicted, and how likely a mother is to be drug addicted given that her baby is drug addicted). Different methods were used across experiments, such as deriving posterior distributions of P(c|e) and P(e|c) via sampling from participants' estimates, or

generating predictions for each reasoner separately based on his or her individual estimates. According to their findings, people are more sensitive to the existence and strength of alternative causes when reasoning diagnostically from effect to cause than when making predictive inferences from cause to effect (but see Meder et al., 2014; Tversky & Kahneman, 1982a).

Structure Induction Model: Diagnostic Reasoning with Causal Structure Uncertainty

 (\blacklozenge)

Although the power PC model operates on causal parameters that are estimated from the observed data (in one way or another), it brings the strong assumption to the task that there is actually a causal link between C and E. The only situation in which the account assumes that there is no causal relation is when $P(c|e) = P(c|\neg e)$, and therefore $w_c = 0$. This approach lacks the expressive power to take into account the possibility that an observed contingency in the data [i.e., $P(c|e) \neq P(c|\neg e)$] is just coincidental. Consider again the data set in Figure 23.1 b. The observed data indicate that the candidate cause (e.g., genetic predisposition) raises the probability of the effect (e.g., elevated blood pressure) from $P(e|\neg c) = 2/20 = .1$ to P(e|c) = 6/20 = .3; accordingly, the estimated causal strength of C is $w_1 = 0.22$ (Equation 4). But how reliable is this estimate, given the available data? If the estimate is based on a data sample, it may well be that the observed contingency is merely accidental and not diagnostic for a causal relation. This is similar to a situation in which one tosses a fair coin 40 times-one would not be surprised if the observed number of heads was not exactly 20 but, say, 24. The important point here is that when inductive inferences are drawn based on samples, there is usually uncertainty about whether the observed contingency is indicative of a causal relation or is merely coincidental.

The structure induction model of diagnostic reasoning (Meder, Mayrhofer, & Waldmann, 2009, 2014) formalizes the intuition that diagnostic reasoning should be sensitive to the question of whether the sample data warrant the existence of a causal relation between C and E. The characteristic feature of the model is that it does not operate on a single causal structure, as the power PC model does (and its isomorphic Bayes nets representation, i.e., structure S_1 ; Figure 23.1 c). Rather, it also considers the possibility that C and E are, in fact, independent of each other (Anderson, 1990; Griffiths & Tenenbaum, 2005; see also McKenzie

 (\mathbf{r})

(�

^{438 |} DIAGNOSTIC REASONING

with structure

EMPIRICAL STUDIES

Meder and colleagues (2014) tested the structure induction model using a medical diagnosis paradigm in which participants were provided with learning data about the co-occurrences of a (fictitious) virus and a (fictitious) symptom. Given this sample data, participants were asked to make a diagnostic judgment for a novel patient who has the symptom. The studies used nine different data sets, factorially combining different levels of the (empirical) diagnostic probability, P(c|e), with different levels of the (empirical) predictive probability, P(e|c). The experimental rationale was to fix the empirical diagnostic probability but to vary other aspects of the data in order to generate predictions that distinguish the structure induction model from the simple Bayes model and different variants of power PC theory.

Consider Figure 23.2 a: in all three data sets the base rate of the cause is $P(c) = P(\neg c) = .5$ and the empirical diagnostic probability is P(c|e) = .75. In contrast, the predictive probability of effect given cause, P(e|c), and the causal strength of C, w, vary across the three data sets (from left to right the causal strength estimate increases; Equation 3). Figure 23.2 b shows the models' predictions for the three data sets. The simple Bayes and power PC model entail the same diagnostic judgment across the data sets, since the empirical diagnostic probability is invariant. The structure induction model, however, makes a very different prediction, entailing diagnostic probabilities that systematically deviate from the empirical diagnostic probability. Specifically, the model predicts an upward trend, yielding an increasing probability of target cause c given the effect e across the data sets. This upward trend results from the posterior probabilities of structures S_0 and S_1 , whose posterior Bayesian estimates vary across the data samples (see Meder et al., 2014, for details). As a consequence, the inferred diagnostic probability increases when the posterior probability of S_1 becomes higher (i.e., when it becomes more likely that the observed contingency is indicative of an underlying causal relation).

Empirically, participants' diagnostic judgments showed the upward trends predicted by the structure induction model, that is, human diagnostic judgments were not invariant for different data sets entailing the same empirical diagnostic probability P(c|e). These studies demonstrated that people's diagnostic judgments do not solely reflect the empirical probability of a cause given an effect, but systematically vary as a function of causal structure

& Mikkelsen, 2007), as illustrated with structure S_0 in Figure 23.1 d. Importantly, the two causal structures have different implications for the diagnostic inference from cause to effect. Under S_1 , observing effect E provides (probabilistic) evidence for the presence of the cause, so that P(c|e) > P(c) (except for the limiting case in which $P(c|e) = P(c|\neg e)$, and therefore $w_1 = 0$). For instance, for the data set in Figure 23.1 b, structure S_1 entails that P(c|e) = .71. Note that this value is similar but not identical to the empirical probability of .75, with the divergence resulting from the fact that the account does not use maximum likelihood but Bayesian estimates (i.e., independent uniform priors over the structures' parameters are used, which are updated in light of the available sample data). Structure S_0 , however, entails a very different value for the diagnostic probability. According to S_0 , C and E are independent events; therefore observing the presence of *E* does not increase the probability of *C*; that is, P(c|e) = P(c). Since in the data set shown in Figure 23.1 b the cause is present in 20 of the 40 cases, S_0 entails P(c|e) = P(c) = .5.

To take into account the diverging implications of the causal structures and their relative probability given the data, the structure induction model integrates out the two structures to arrive at a single estimate for the diagnostic probability. Formally, this is done by weighting the two diagnostic estimates derived from the parameterized structures by the corresponding posterior probability of structures S_0 and S_1 , respectively (i.e., Bayesian model averaging; Chickering & Heckerman, 1997), which are $P(S_0|\text{data}) = .49 \text{ and } P(S_1|\text{data}) = .51 \text{ in our example}$ (assuming a uniform prior over the structures, i.e., $P(S_0) = P(S_1) = .5$. For instance, for the data set in Figure 23.1 b the structure induction model predicts P(c|e; data) = .61, which results from weighting each structure's diagnostic estimate with the structure's posterior probability (i.e., $.49 \times .5 + .51$ $\times .71 = .61$). This diagnostic probability then reflects the uncertainty with respect to the true underlying causal structure and the uncertainty of the parameter estimates.3

In sum, the structure induction model of diagnostic reasoning takes into account uncertainty regarding possible causal models that may have generated the observed data. As a consequence, the derived diagnostic probabilities can systematically deviate from the empirical diagnostic probability of a cause given an effect in the sample data and the predictions of the simple Bayes account and power PC theory.

(

(₽)



 (\blacklozenge)

Figure 23.2 Predictions of different computational models of elemental diagnostic inference. (a) Three data sets in which the empirical diagnostic probability of a cause given an effect is P(c|e) = .75. The predictive probability, P(e|c), and the causal strength of *C*, w_c vary across the data sets (numbers are maximum likelihood estimates of causal power, based on the empirical probabilities.) (b) Predictions of the structure induction model, the simple Bayes model, and the power PC model, using maximum likelihood estimates (MLEs), for the three data sets. The latter two models predict identical diagnostic probabilities across the data sets, whereas the structure induction model predicts a systematic upward trend, resulting from different structure posteriors entailed by the data samples.

uncertainty. These findings support the idea that people's diagnostic inferences operate on the causal level, rather than on the data level, and that their diagnostic inferences are sensitive to alternative causal structures that may underlie the data.

Summary: Elemental Diagnostic Reasoning

Different computational models of elemental diagnostic inference share the assumption that the goal of the diagnostic reasoner is to infer the conditional probability of the candidate cause given the effect. However, the accounts differ strongly in their theoretical assumptions and the ways in which the diagnostic probability is computed from the available data. The simple Bayes model, which is usually presumed to provide the rational benchmark in diagnostic reasoning, prescribes that causal judgments should reflect the empirical probability of the cause given the effect in the data. Power PC theory

440 | DIAGNOSTIC REASONING

and its isomorphic Bayes net representation conceptualize diagnostic reasoning as an inference on the causal level, using structure S_1 as the default structure. The structure induction model advances this idea by considering a causal structure hypothesis according to which *C* and *E* are in fact independent events, with the inferred diagnostic probability taking into account the uncertainty about the existence of a causal relation. As a consequence, diagnostic probabilities derived from the structure induction model can systematically diverge from the empirical probability of the cause given the effect.

Diagnostic Reasoning with Multiple Causes and Effects

Our discussion thus far has centered on elemental diagnostic inferences from a single effect to a single cause. In this section, we discuss diagnostic causal reasoning with more complex causal models that

()

can involve multiple causes or effects. For instance, the same symptom could be caused by different diseases, such as a viral or bacterial infection. In this case, a single piece of evidence can have differential diagnostic implications for different possible causes. Conversely, a viral infection (cause) can generate several symptoms (effects), such as headache, fever, and nausea. In this case, different pieces of evidence need to be combined to make a diagnostic judgment about one target cause.

In the framework of probabilistic inference over graphical causal models, the causal dependencies in the graph determine the factorization of the joint probability distribution over the domain variables (Pearl, 2000; Spirtes et al., 1993). The factorization follows from applying the causal Markov condition to the graph, which states that the value of any variable in the graph is a function only of its direct causes (its Markovian parents). In other words, conditional on its direct causes, each variable in the model is independent of all other variables, except its causal descendants (i.e., its direct and indirect effects). This causally based factorization implies specific relations of conditional dependence and independence for the probability distribution associated with the graph, which facilitate and constrain inferences across multiple variables. Importantly for the present discussion, the particular dependency and independency relations entail specific diagnostic inference patterns when reasoning with different causal structures.

In the following, we discuss key issues related to diagnostic reasoning in causal models with multiple causes or effects, focusing on commoneffect and common-cause models (see also Rehder, Chapters 20 and 21, and Rottman, Chapter 6, in this volume). Subsequently, we address the relation between diagnostic reasoning and information search, which is an important aspect of diagnostic reasoning in a broader sense.

Diagnostic Reasoning with Common-Effect Structures: Explaining Away

An important property of diagnostic reasoning in common-effect structures is *explaining away* (Morris & Larrick, 1995; Pearl, 1988, 2000; Rottman & Hastie, 2014).⁴ Consider the example of a common-effect structure shown in Figure 23.3 a, according to which $C_1 = \{c_1, \neg c_1\}$ (e.g., virus present vs. absent) and $C_2 = \{c_2, \neg c_2\}$ (e.g., bacteria present vs. absent) are independent, not mutually exclusive, causes of a common effect $E = \{e, \neg e\}$ (e.g., symptom present vs. absent). Associated with the causal structure is a set of parameters: the base rates of the two cause events, their respective causal strengths, and the strength of the background cause (not shown). These parameters fully specify the joint probability distribution over the two causes and the effect.

Figure 23.3 shows an example data set for 100 cases generated from setting the base rate of each independent cause to .5 and the strength of the background cause to zero (i.e., the effect never occurs when both C_1 and C_2 are absent). The two causes, virus and bacteria, vary in their causal strength: a virus infection (C_1) generates the symptom with a probability of .8, and a bacterial infection (C_2) generates the symptom with a probability of .6 (in this example scenario, C_1 and C_2 are the sole causes of E, i.e., there are no alternative background causes; therefore these probabilities correspond to the individual causal power estimates of C_1 and C_2).⁵ Assuming a noisy-OR parameterization, the probability of the symptom is .92 when both causes are present (i.e., $P(e|c_1, c_2) = w_{c1} + w_{c2} - w_{c1}w_{c2}$ $= .8 + .6 - .8 \cdot .6 = .92$).

Explaining away occurs in common-effect structures when reasoning diagnostically from the effect to the causes. Since both C_1 and C_2 are (independent) causes of their common effect, observing the presence of the effect raises the probability of both: If we know that a patient has the symptom, this increases the probability of having a virus as well as of having a bacterial infection. The particular diagnostic probabilities depend on the causes' base rates and their causal strengths, as well as on the strength of the unobserved background causes. For instance, for the example data in Figure 23.3 b, $P(c_1|e) = 43/$ 58 = .74 and $P(c_2|e) = 38/58 = .66$: both causes are equally likely a priori, but C_1 is more likely to cause the symptom, so the diagnostic probability for C_1 is higher than for C_2 . This diagnostic inference can be modeled by Bayes's rule using a structure parameterized with conditional probability estimates (Pearl, 1988) or using estimates of causal strength, as similarly discussed in the section on elemental diagnostic reasoning. If available, the diagnostic probabilities can also be computed directly from a joint frequency distribution, as done above with the example data in Figure 23.3 b.

Explaining away with respect to some target cause occurs when conditioning not only on the effect, but also on the known presence of an alternative cause. In the present scenario, with respect to cause C_1 , explaining away corresponds to the inequality $P(c_1|e) > P(c_1|e, c_2)$. In words, the diagnostic

((()

9780199399550_Waldmann290716MEDUS_Book.indb 441

(₽)

(a) Common-effect structure (b) Example data Symptom Symptom Symptom present absent E Bacteria 23 2 present Virus .92 .8 6 present Bacteria 20 5 absent Bacteria C_1 C_2 15 10 present Virus absent Bacteria 0 25 .5 .5 absent Virus Bacteria (c) Explaining away 1.00 $_{1}|e) = .74$.75 Probability of cause ${\cal C}_1$ $P(c_1|e,c_2) = .61$.50 .25 Diagnostic probabilities $- - P(c_1|e)$ - $P(c_1|e, c_2)$ 0.00

()

۲

Figure 23.3 Explaining away. (a) A common-effect model with the independent, not mutually exclusive causes $C_1 = \{c_1, \neg c_1\}$ and $C_2 = \{c_2, \neg c_2\}$, virus and bacteria, and an effect $E = \{e, \neg e\}$, a symptom. (b) Joint frequency distribution generated from a noisy-OR parameterization of the common-effect model, assuming no background cause and $P(e|c_1, \neg c_2) = .8$ and $P(e|\neg c_1, c_2) = .6$. (c) Explaining away of c1 across different prior probabilities of the two causes, with $P(c_1) = P(c_2)$. The difference between $P(c_1|e)$ and $P(c_2|e)$ is the amount of explaining away, exemplified with the two diagnostic probabilities for $P(c_1) = P(c_2) = .5$.

 (\bullet)

.50 Prior probability of causes C_1 and C_2

.25

0.00

probability of cause c_1 conditional on effect e alone is higher than when conditioning on both the effect *e* and the alternative cause c_3 ; thus, the presence of c_2 explains away some of the diagnostic evidence of e with respect to c_1 . Consider again the medical scenario: if a patient has the symptom, reasoning diagnostically increases the probability of the virus being present. Now imagine you also learn that the patient has a bacterial infection, which is the other of the two possible causes that could have produced the symptom. Intuitively, if we learn that the patient

DIAGNOSTIC REASONING 442

has a bacterial infection this "explains away" (some of) the diagnostic evidence of the symptom regarding the presence of the virus; that is, it reduces the probability of the virus being present relative to a situation in which we only condition on the effect.

1.00

.75

Consider the example data set shown in Figure 23.3 b: Given this joint frequency distribution, what are the diagnostic probabilities $P(c_1|e)$ and $P(c_1|e, c_2)$? In other words, how likely is the virus to be present if the symptom is present, and how likely is the virus to be present given

the presence of both the symptom and the bacteria? Based on the joint frequency distribution, $P(c_1|e) = 43/58 = .74$; that is, the virus is present in about 74% of the cases in which the symptom is present. The probability of c_1 given both *e* and c_2 can be computed analogously, yielding $P(c_1|e_2)$ c_2 = 23/38 = .61; that is, the virus is present in about 61% of the cases in which both the symptom and the bacteria are present-the presence of the alternative cause c_2 has "explained away" some of the diagnostic evidence of e with respect to c_1 . The amount of explaining away is the difference between the two diagnostic probabilities; that is, $P(c_1|e) - P(c_1|e, c_2) = .13$. Note that the probability of c_1 does not reduce to zero: Because the virus and the bacterial infection are independently occurring causes, the presence of the bacterial infection does not rule out that the patient also has a viral infection-it only makes it less likely than before (see Morris & Larrick, 1995, for a detailed analysis of the conditions of explaining away). In fact, the diagnostic probability is still higher than the base rate of the virus, which is .5 in this example.

Figure 23.3 c (cf. Figure 5 in Morris & Larrick, 1995) illustrates a more general case, showing the amount of explaining away for different base rates of the two cause events, under the constraint that $P(c_1) = P(c_2)$. The causal strengths are fixed to the values as above (i.e., the individual likelihoods are .8 for c_1 and .6 for c_2 , no background cause, noisy-OR parameterization). The curves correspond to the two diagnostic probabilities $P(c_1|e)$ and $P(c_1|e, c_2)$ across different base rates of the two causes, showing how the amount of explaining away varies as a function of the causes' prior probability. The two dots are the data points from the preceding example, in which $P(c_1) = P(c_2) = .5.$

EMPIRICAL STUDIES

(�

Empirical research on explaining away in diagnostic causal reasoning with common-effect structures has yielded mixed findings. While there are many studies on discounting in a broader sense (see Khemlani & Oppenheimer, 2011, for an overview), there are few studies that have directly investigated explaining away from the perspective of inductive causal inference.

Morris and Larrick (1995; Experiment 1) investigated whether and to what extent people demonstrate explaining away in a social inference scenario. They used a paradigm by E. E. Jones and Harris (1967), in which the task was to infer the political attitude of the writer of an essay E. For instance,

 (\mathbf{r})

the potential causes of a positive essay about Fidel Castro were a pro-Castro attitude (A) of the writer or the instruction (I) to write a positive essay. This situation can be conceptualized as a common-effect model $A \rightarrow E \leftarrow I$. The independence and base rate of I were instructed through a cover story; quantitative model predictions were derived by eliciting participants' subjective judgments of the other relevant probabilities (e.g., base rates of causes A and I, the prevalence of pro-Castro attitudes and probability of having been instructed to write a pro-Castro essay, and corresponding likelihoods). Explaining away can be tested by comparing judgments for P(A|E), the probability that the writer has a positive attitude given a pro-Castro essay, with P(A|E, I), the probability that the writer has a positive attitude given a pro-Castro essay and given that the writer was instructed to write a positive essay. Consistent with explaining away, lower judgments for P(A|E, I) were obtained than for P(A|E): given a pro-Castro essay, participants increased their judgment of the probability that the writer had a pro-Castro attitude but lowered their judgments when informed that the writer had been instructed to write a positive essay.

More recent research has tested explaining away in the context of causal Bayes net theories. Rehder (2014; see also Rehder & Waldmann, in press) used common-effect structures with two binary causes and one binary effect in different domains, such as economics, meteorology, and sociology. Participants were taught qualitative causal models based on described causal relations between binarized variables, such as "a low amount of ozone causes high air pressure" or "low interest rates cause high retirement savings." The instructions also explicated the causal mechanisms underlying these relations (see Rehder, 2014, for details). No quantitative information on the exact parameters of the instructed causal networks was provided; the studies focused on the qualitative diagnostic inference patterns. The studies used a forced-choice task in which participants were presented with a pair of situations, corresponding to judgments about $P(c_1|e)$ and $P(c_1|e, c_2)$. The task was to choose in which situation a target cause C_1 was more likely to take a particular value: when only the state of the effect was known, or when both the effect and the alternative cause were known. If people's inferences exhibit explaining away, they should prefer the former over the latter, corresponding to the inequality $P(c_1|e) > P(c_1|e, c_2)$. Human behavior was at variance with explaining away; in fact, participants tended to exhibit the opposite pattern [i.e., choosing $P(c_1|e, c_2)$ over $P(c_1|e)$].

> MEDER AND MAYRHOFER 443

9780199399550_Waldmann290716MEDUS_Book.indb 443

()

Rottman and Hastie (2015; see also Rottman & Hastie, 2016) investigated explaining away using a learning paradigm in which participants observed probabilistic data generated from a parameterized common-effect model with binary variables. Quantitative predictions for patterns of explaining away were derived from the parameterized causal model. However, people's inferences were inconsistent with the model predictions, and most of the diagnostic judgments did not exhibit explaining away.

SUMMARY

(�)

The currently available evidence on explaining away in human reasoning with common-effect models is limited. While some studies observed explaining away, others found diagnostic inference patterns at variance with explaining away. These are critical findings for adopting causal Bayes net theories as a modeling framework for human causal induction and diagnostic inference. Further empirical research is needed to identify and characterize the circumstances under which human diagnostic reasoning is sensitive to explaining away.

Diagnostic Inference in Common-Cause Structures: Sequential Diagnostic Reasoning

In many diagnostic-reasoning situations, such as medical diagnosis, several pieces of evidence (e.g., results of different medical tests) are observed sequentially at different points in time. In this case, multiple effects are used to reason about the presence of an underlying cause (e.g., a disease) constituting a common-cause structure (Figure 23.4). Sequential diagnostic inferences also raise the question of possible order effects (Hogarth & Einhorn, 1992), such as those resulting from temporal weighing of the sequentially acquired information (e.g., primacy or recency effects).

Hayes, Hawkins, Newell, Pasqualino, and Rehder (2014; see also Hayes et al., 2015), drawing on the work of Krysnki and Tenenbaum (2007) discussed earlier, explored sequential diagnostic reasoning in the mammography problem (Eddy, 1982). In the standard version of the problem, participants are presented with a single piece of evidence, a positive mammogram, and are asked to make an inference about the probability of the target cause, breast cancer. In the studies by Hayes and colleagues, diagnostic judgments based on one versus two positive test results from two different machines were elicited. The crucial manipulation concerned information on possible causes of false-positive results. In the non-causal condition, participants were merely informed about the relative frequency of false positives (e.g., that 15% of women without breast cancer had a positive mammogram). In this situation, the false-positive rates of the two machines are assumed to be independent of each other, so that the second mammogram provides additional diagnostic evidence (i.e., participants' diagnostic judgments regarding the target cause, breast cancer, should further increase relative to diagnostic judgments based on a single test result). In the causal condition, participants received the same statistical information but were also told about a possible alternative cause that can lead to false positives, a benign cyst. The underlying rationale was that the benign cyst would constitute a stable common cause within a tested person, so that a second positive mammogram provides little diagnostic value over the first one. Participants' diagnostic judgments closely resembled these predictions: in the non-causal condition the second mammogram was treated as providing further diagnostic evidence, raising the probability of the target cause relative to the situation with just a single positive test result. By contrast, in the causal condition the second positive mammogram had very little influence on diagnostic judgments. These findings show that people are sensitive to the causal underpinnings of different situations and their implications for probabilistic diagnostic inferences.

Meder and Mayrhofer (2013) investigated sequential diagnostic reasoning with a commoncause model consisting of a binary cause (two chemicals) and four binary effects (different symptoms, e.g., fever and headache). They presented participants with a series of three symptoms, one after the other, with a diagnostic judgment required after each piece of evidence. Information on the individual cause-effect relations was given either in a numerical format (e.g., "Chemical X causes symptom A in 66% of the cases") or in verbal frequency terms (e.g., "Chemical X frequently causes symptom A"). Diagnostic probabilities for the verbal reasoning condition were derived using the numerical equivalents of the used verbal terms from an unrelated study (Bocklisch, Bocklisch, & Krems, 2012; see Mosteller and Youtz, 1990, for an overview). The diagnostic task for participants was to estimate the posterior probabilities of the two causes, given all observed effects so far. In this study, people's sequential diagnostic inferences were remarkably accurate, with judgments closely



(a) Parameterized causal structure; numbers denote unconditional and conditional probabilities. (b) Joint frequency distribution of 100 cases generated from the parameterized causal model. (c) Diagnostic tree. On the first step, the diagnostic reasoner has to decide which symptom to query (i.e., fever or nausea). On the next step, information about the state of the symptom is obtained, with the numbers referring to the probability of the different states. For instance, the probability that a patient has fever is .55, and the probability that a patient has nausea is .17. The bottom of the tree shows the diagnostic probabilities entailed by the symptom status. For instance, if a patient has fever, the probability of the virus being present is .49; if the patient has no fever, the probability of the virus being present is .07.

()

tracking the diagnostic probabilities derived from the parameterized common-cause model. This was the case regardless of whether information on the cause-effect relations was provided numerically or through rather vague verbal frequency terms. This finding is also interesting with respect to studies showing Markov violations (see the following discussion), because participants' diagnostic judgments were very close to the predictions of a common-cause model in which the effects are independent given the cause. Finally, the study points to interindividual differences regarding the temporal weighting of evidence in sequential diagnostic reasoning. For instance, when previously observed

symptoms had to be recalled from memory, the judged diagnostic probabilities reflected a stronger influence of the current evidence, relative to earlier observed symptoms.

Rebitschek, Bocklisch, Scholz, Krems, and Jahn (2015; see also Jahn & Braatz, 2014; Jahn, Stahnke, & Rebitschek, 2014; Rebitschek, Krems, & Jahn, 2015) investigated order effects in sequential diagnostic reasoning more closely. They used a medical diagnosis task with four chemicals as possible causes and six symptom categories, with each category including two symptoms (e.g., "twinge" and "sting" belonged to the category "pain"). Participants were presented with four sequentially presented

MEDER AND MAYRHOFER 445

9780199399550_Waldmann290716MEDUS_Book.indb 445

()

symptoms, with the symptom sequences designed to examine possible order effects (e.g., whether it matters which of two hypotheses was supported more strongly by the first symptom, even if the total diagnostic evidence supported them equally). The diagnostic task was to choose the chemical that was most likely to have caused the symptom(s). Diagnostic judgments were obtained either after participants saw the full sequence of symptoms, or judgments were obtained after each symptom (see Hogarth & Einhorn, 1992, for a discussion of different elicitation methods with respect to order effects). Diagnostic judgments were not invariant with respect to presentation order, with the diagnoses often being influenced by the initially presented piece of evidence. This primacy effect was mediated by the testing procedure: diagnostic judgments after the full symptom sequence showed a strong primacy effect, whereas when participants were asked to rate their diagnostic beliefs after each symptom, the final diagnosis was only weakly influenced by the initially observed symptom. Moreover, the influence of late symptoms was revealed (i.e., recency effects).

SUMMARY AND DISCUSSION

Diagnostic reasoning in common-cause models has been primarily investigated from the perspective of order effects. The exact nature of order effects, the conditions under which they occur, and how they can be formally modeled from the perspective of causal inference remain important issues for future research (see also Trueblood & Busemeyer, 2011).

In common-cause models, it is assumed that the effects are conditionally independent of each other given their common cause (i.e., Markov property), such that they provide independent evidence for the cause. (In the machine-learning literature, this property is referred to as *class-conditional indepen*dence of features, implemented in the naïve Bayes classifier; see Domingos & Pazzani, 1997; Jarecki, Meder, & Nelson, 2016.) Making this assumption strongly simplifies the diagnostic inference process, because the number of estimates required to parameterize the causal structure is greatly reduced. However, a growing body of research on human causal reasoning shows that people's inferences in related tasks, such as (conditional) predictive causal reasoning, do not honor the Markov condition (Mayrhofer & Waldmann, 2015a; Park & Sloman, 2013; Rehder, 2014; Rehder & Burnett, 2005; Rottman & Hastie, 2015; Walsh & Sloman, 2008; but see Jarecki, Meder, & Nelson, 2013; von Sydow, Hagmayer, & Meder, 2015): typically people seem

to expect a stronger correlation between effects of a common cause than normatively justified. These findings raise the question to what extent and under what conditions human causal reasoning is consistent with the Markov condition and the entailed dependency and independency relations that should guide and constrain diagnostic inferences.

Diagnostic Reasoning and Information Search

How do people decide what information is diagnostically relevant? So far our discussion has focused on situations in which the reasoner makes diagnostic inferences from one or more effects to possible causes. In many circumstances, however, diagnostically relevant information needs to be actively acquired before making a diagnostic inference, such as when deciding which medical test to conduct.

A key theoretical question is how to quantify the diagnostic value of possible information queries (Nelson, 2005). Different models of the value of information have been proposed in the literature, based on a probabilistic framework. The models entail different types of informational utility functions that quantify the diagnostic value of a datum (e.g., the outcome of a medical test; Benish, 1999) according to some formal metric, such as expected reduction in uncertainty or expected improvement in classification accuracy. In the following, we introduce key ideas pertaining to diagnostic causal reasoning and discuss the application of information-theoretic concepts in empirical research.

Quantifying Diagnostic Value

Consider a medical scenario in which a virus (binary cause event *C*) probabilistically generates two symptoms, fever (E_1) and nausea (E_2) . This scenario can be represented as a common-cause structure (Figure 23.4 a). The parameters associated with the causal structure are unconditional and conditional probabilities.⁶ The virus has a base rate of P(virus) = .3 and generates fever and nausea with likelihoods P(fever|virus) = .9 and P(nausea|virus) = 1/3. The symptoms can also occur in the absence of the virus, with $P(\text{fever}|\neg \text{virus}) = .4$ and $P(\text{nausea}|\neg \text{virus}) = .1$. Figure 23.4 b shows an example data set of 100 cases, generated from the parameterized common-cause model.

Now imagine a physician diagnosing a new patient. It is unknown if the patient has fever or nausea, but the doctor can acquire information about the symptoms. Is it more useful to find out

 (\mathbf{r})

about the presence or absence of fever or nausea, respectively? Note the crucial difference in the diagnostic reasoning scenarios considered so far, where the diagnostic inference was based on knowing the state of the effect. In the present scenario, the critical question is which query is more useful to conduct, with the outcome being uncertain. For instance, when testing the patient for fever there are two possible outcomes, namely, fever or no fever. Both states have implications for the diagnostic inference about the virus, but prior to gathering information the state of the effect is uncertain.

Since the virus is causally related to both fever and nausea, learning about either of them provides diagnostic information about the presence of the virus. This is illustrated in the diagnostic tree in Figure 23.4 c, which shows the probability of observing the different symptom states, as well as the resulting posterior probabilities of the cause. For instance, if testing for the presence of fever (left branch), the probability that the patient has fever is .55, in which case the probability of the virus being present will increase to .49. Conversely, if the patient does not have fever, which happens with probability .45, the posterior probability of the virus being present is .07. (These probabilities can be computed from the parameterized causal model via Bayes's rule or directly from the joint frequencies in Figure 23.4 b.)

But which query has higher diagnostic value: Is it better to test for the presence of fever or for the presence of nausea? The answer to this question crucially depends on how we value a query's outcome. Different measures for quantifying the usefulness of a datum (e.g., outcome of a medical test) have been suggested in statistics, philosophy of science, and psychology (for reviews, see Crupi & Tentori, 2014; Nelson, 2005). Typically, the different measures are based on a comparison of the prior versus posterior probability distributions, for each possible outcome of a query (a pre-posterior analysis, in the terminology of Raiffa & Schlaifer, 1961). The expected usefulness of a query Q (e.g., a medical test) is computed by weighting the usefulness of each possible query outcome by its probability of occurrence. In the present example there are two queries, referring to gathering information about whether the patient has fever or nausea, with each query having two possible outcomes (e.g., fever present or absent).

Importantly, alternative measures of the value of information are not formally equivalent, as they rank the usefulness of possible diagnostic queries differently (Nelson, 2005). To illustrate, we here focus on two prominent measures: information gain (Lindley, 1956), which values queries according to the expected reduction in uncertainty, measured via Shannon (1948) entropy, and probability gain, which values queries according to the expected improvement in classification accuracy (Baron, 1985).

Information gain quantifies the usefulness of a datum by the expected reduction in Shannon entropy.7 (Note that in the expectation, information gain is equivalent to Kullback-Leibler, [1951], divergence, although the usefulness of individual outcomes may differ.) In the current scenario, to compute the information gain of, say, testing a patient for the presence of fever, the posterior entropy of the cause's distribution given the two possible test outcomes (fever vs. ¬fever) is considered. The information gain of a test outcome (which can be positive or negative) is the difference between the entropy of the prior distribution and the entropy of the posterior distribution, conditional on the status of the effect. The expected information gain is then computed by weighting the (positive or negative) gain of each possible outcome of the query by the probability of observing the outcome. Given the parameters of the common-cause model, the expected information gain of testing for fever is 0.172 bits. In other words, learning whether the patient has fever will, in the expectation, reduce the diagnostic reasoner's uncertainty about the virus by 0.172 bits.8 The analogous calculation for the alternative effect, nausea, yields an expected information gain of 0.054 bits. Thus, from the perspective of uncertainty (entropy) reduction, testing a patient for the presence of fever is more useful than testing for the presence of nausea, because the former entails a higher reduction in Shannon entropy.

A different model for quantifying the usefulness of diagnostic tests is probability gain (Baron, 1985), which values information by the expected improvement in classification accuracy (Nelson, McKenzie, Cottrell, & Sejnowski, 2010). Formally, this measure is based on the difference in accuracy prior to conducting a query versus accuracy after conducting a query. Consider a patient drawn randomly from the data sample in Figure 23.4 b. If the goal is classification accuracy, one should predict the most likely hypothesis, namely, that the patient does not have the virus, because the virus is present in only 30% of the cases (see Meder & Nelson, 2012, for analyses of scenarios with situation-specific payoffs). In other words, the probability of making a correct classification decision is .7 prior to obtaining any information about the effects (symptoms).

MEDER AND MAYRHOFER 447

9780199399550_Waldmann290716MEDUS_Book.indb 447

(�)

()

Can a higher accuracy be expected if testing the patient for fever or nausea? The basic rationale is the same as with the information gain model. First, the posterior distribution of the cause given each state of the effect is considered. For instance, when fever is present, accuracy decreases to .51, because 51% of patients with fever do not have the virus. By contrast, when the patient does not have fever, accuracy increases to .93, because in 93% of the cases the patient does not have the virus. To compute the overall probability gain of the query, an expectation is computed by weighting each outcome's gain by the probability that a patient does or does not have fever. Interestingly, the expected probability gain of testing a patient for the presence of fever in our example is zero.9 Thus, from the perspective of the probability gain model this query is useless. By contrast, the same computations for the second effect, nausea, give a probability gain of .03, that is, testing a patient for the presence of nausea will, on average, increase classification accuracy by 3%. Thus, a diagnostic reasoner who aims to increase classification accuracy should find out whether the patient has nausea. By contrast, a diagnostic reasoner who aims to reduce uncertainty should find out whether a patient has fever, because this query entails the higher expected reduction in Shannon entropy.

This divergence between different models of the value of information is critical because it highlights that the usefulness of possible queries depends on which metric is used to quantify the diagnostic value of information. In the scenario considered here, if the goal is to reduce uncertainty (measured via Shannon entropy) about the virus, the diagnostic reasoner should test for the presence of fever. By contrast, if the goal is classification accuracy, the diagnostic reasoner should test for the presence of nausea. (Similar divergences hold for other models of the value of information; see Nelson et al., 2010.)

EMPIRICAL STUDIES

Different models of the value of information have been used to explain human behavior on a variety of cognitive tasks involving active information acquisition (Austerweil & Griffiths, 2011; Baron & Hershey, 1988; Markant, Settles, & Gureckis, 2015; Meder & Nelson, 2012; Meier & Blair, 2013; Nelson et al., 2010; Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014; Rusconi & McKenzie, 2013; Wells & Lindsay, 1980). For instance, Oaksford and Chater (1994) re-analyzed Wason's (1968) selection task from the perspective of inductive probabilistic inference,

 (\mathbf{r})

448 | DIAGNOSTIC REASONING

arguing that human behavior is inconsistent with the classic logico-deductive analysis but constitutes rational behavior from the perspective of active information sampling (Oaksford & Chater used Shannon entropy to quantify the usefulness of queries; Nelson, 2005, showed that alternative models of the value of information yield similar predictions). Crupi, Tentori, and Lombardi (2009) provided an analysis of the pseudodiagnosticity paradigm (Doherty, Mynatt, Tweney, & Schiavo, 1979)-a task that has been interpreted to demonstrate flawed human thinking regarding the diagnostic value of information. Crupi and colleagues showed that this interpretation relies on a specific model for computing diagnostic value, and that participants' behavior is, in fact, consistent with seeking high-probability-gain information.

Most of these studies have not explicitly adopted a causal modeling framework, but there are important connections between key theoretical ideas. Nelson and colleagues (2010; see also Meder & Nelson, 2012) examined information search in a classification task. First, participants learned about the statistical structure of the environment in a trial-by-trial learning procedure, categorizing artificial biological stimuli into one of two classes based on two binary features. The generative model underlying the task environment corresponds to a common-cause structure, in which the likelihoods of the features are conditionally independent given the true class. This situation is analogous to the preceding common-cause scenario, with the class corresponding to the cause variable and the stimuli's features corresponding to its effects. In a subsequent search task, learners could query one of the two features to obtain information before making a classification decision. The structure of the environment was such that one query would improve classification accuracy (i.e., had higher probability gain), whereas the alternative query was more useful from the perspective of information gain (or some other model of the value of information; Nelson and colleagues considered several models from the literature). Across several experiments, participants' search behavior was best accounted for by probability gain. The studies also highlight the importance of how information about the relevant probabilities is conveyed. A clear preference for the diagnostic query with the higher probability gain was only obtained when people learned about the statistical structure of the environment through experience, whereas conveying probability information (base rates and likelihoods) through words and

numbers was not very helpful for identifying the higher-probability-gain query, with search decisions often being close to chance level (see also Meder & Nelson, 2012).

Summary

(�)

There is a rich theoretical literature on quantifying the diagnostic value of information queries. Different models have been suggested, based on different assumptions about what makes information valuable with respect to the goals of the diagnostic reasoner. An important insight is that different models can make similar predictions in many statistical environments (Nelson, 2005), which highlights the need for carefully designed experiments that allow researchers to disentangle competing models (Meder & Nelson, 2012; Nelson et al., 2010). This is also an important issue for the normative analysis of human search behavior and people's sensitivity to the diagnostic value of queries (e.g., Crupi et al., 2009; Oaksford & Chater, 1994).

Most empirical studies on information search have not explicitly adapted a causal modeling framework, but there are relations in terms of the generative models that have been used (e.g., the close relation between the independency relations in common-cause models and the notion of class-conditional independence, which can be considered a special case of the Markov condition). More recently, empirical studies on causal structure induction have applied different models of the value of information (see also Rottman, Chapter 6 in this volume). Steyvers and colleagues (2003) explored different variants of models based on information gain to predict intervention decisions on causal networks. Bramley et al. (2015) considered different models besides entropy reduction for quantifying the usefulness of interventions in causal structure learning. Coenen and colleagues (2015) contrasted information gain with a positive test strategy (e.g., Klayman & Ha, 1987) in structure induction. These studies provide pathways for future research by bringing together information-theoretic ideas of the diagnostic value of information with studies on human causal reasoning.

General Discussion

The goal of this chapter was to discuss diagnostic reasoning from the perspective of causal inference. The computational framework that provides the foundation for our analyses, probabilistic inference over graphical causal models, makes it possible to implement a variety of different models that

share the assumption that diagnostic inferences are guided and constrained by causal considerations. The first part of this chapter highlighted that causalbased models of diagnostic inference can make systematically different predictions from purely statistical accounts, such as the simple Bayes model. This is a critical insight for both the normative and descriptive analysis of human diagnostic reasoning, regardless of whether computational (or "rational") models of cognition (in the sense of Marr, 1982, and Anderson, 1990) are treated as normative standards or psychological theories of human behavior (McKenzie, 2003). In the second part, we discussed more complex diagnostic inferences involving multiple causes or multiple effects. A causal-modelbased factorization of probability distributions entails specific relations of conditional dependence and independence among the domain variables, which constrain diagnostic inferences when reasoning with more complex causal models. The third section considered the question of how to quantify the diagnostic value of information. Deciding what information is diagnostically relevant is a key issue in diagnostic reasoning, and future research should aim to explicate the relations between models of diagnostic inference, measures of the value of information, and human information-acquisition strategies in the context of diagnostic causal reasoning.

Key Issues for Future Research

The analysis of diagnostic reasoning from the perspective of causal inference has provided a number of novel theoretical insights and guided empirical research on people's diagnostic reasoning. In the following, we discuss theoretical and empirical key issues that should be addressed in future work.

THE INDETERMINACY OF RATIONAL MODELS

The development of the framework of probabilistic inference over graphical causal models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993) has advanced research on human causal reasoning, from both a theoretical and an empirical perspective. One way to think about the relation between the general modeling framework and particular models is in terms of the "building blocks" that can be used to characterize existing models or develop new accounts. One differentiating feature concerns the question of parameter estimation and representation, that is, whether a particular model uses maximum likelihood estimates (e.g., Cheng, 1997; Fernbach et al., 2011) or distributions over parameters (e.g., Holyoak et al., 2010; Lu et al., 2008;

> MEDER AND MAYRHOFER 449

()

Meder et al., 2014). In the case of elemental diagnostic reasoning, a power PC model based on maximum likelihood estimates directly derived from the data makes the same predictions as the simple Bayes model. A Bayesian power PC model, in contrast, leads to different predictions, depending on what kinds of priors are used. Examples include uniform priors, which the structure induction model uses, the "sparse-and-strong" prior suggested by Lu and colleagues (2008), and a "sufficiency prior" formalizing a tendency to assume that causal relations are almost deterministic (i.e., high causal strengths; see Mayrhofer & Waldmann, 2011, 2015b; Yeung & Griffiths, 2015). Another key issue concerns the question of structure uncertainty. One idea is to use a single causal structure whose parameters are estimated from data (in one way or another); another idea is to consider multiple causal structures that may have generated the data. These considerations can lead to quite different model behavior, as exemplified by the diverging predictions of the power PC account and the structure induction model. Another issue that we have not discussed so far concerns the functional form of the considered causal structures, that is, how multiple causes combine to generate an effect. We focused on a noisy-OR parameterization as a default functional form for how multiple causes produce the effect (Cheng, 1997; Pearl, 1988), but different functional forms are plausible in other circumstances (Novick & Cheng, 2004; Waldmann, 2007). For instance, a causal model of a food allergy may state that two ingredients (e.g., peanuts and raisins) are jointly necessary to produce an allergic shock. Thus, the assumed functional form constitutes another building block, with the question being whether the functional form is fixed or is assumed to be part of the inference problem (Lucas & Griffiths, 2010).

The upshot is that it is important to distinguish a computational modeling framework such as probabilistic inference over graphical causal models from the specific model instantiations, which can strongly differ in their scope and predictions. The framework supports the development of different computational models that can be tested empirically, but the framework itself is not subject to direct empirical tests. A possible exception might be to test the psychological validity of central theoretical assumptions, such as people's sensitivity to particular dependency and independency relations, based on the Markov condition (e.g., conditional independence in common-cause models or explaining away in common-effect models). In sum, research should be guided by competitive model testing. Instead of comparing human behavior to a single "rational" model (Anderson, 1990), multiple (rational or otherwise) models should be considered and evaluated with respect to their psychological validity and normative desirability.

FROM DIAGNOSTIC PROBABILITIES TO ESTIMATES OF CAUSAL RESPONSIBILITY

Models of diagnostic reasoning typically assume that the computational goal is to infer the probability of a cause given an effect. Another plausible goal of a reasoner might be a judgment of causal responsibility (or causal attribution). Such a diagnostic judgment refers to the probability that the occurrence of effect *E* was in fact brought about by target cause *C*, which is different from the diagnostic conditional probability P(c|e) (Cheng & Novick, 2005).

Consider a medical diagnosis scenario regarding the causal relation between a genetic predisposition and elevated blood pressure. Assume that a study tests 100 patients and finds that 50 have a genetic predisposition, so the cause's empirical base rate in the sample is $P(c) = P(\neg c) = .5$. Of the 50 patients with a genetic predisposition, 30 have elevated blood pressure, so P(e|c) = .6. On the other hand, 30 of the 50 patients without the genetic predisposition also have elevated blood pressure; that is, $P(e|\neg c) = .6$. These estimates suggest that having a genetic predisposition does not raise the probability of elevated blood pressure, which implies that the causal strength is zero (Equation 3). In this case, the probability that a patient from the sample with elevated blood pressure has the genetic predisposition is 50%, as P(c|e) = P(c) = .5.

A different diagnostic inference concerns the probability that the genetic predisposition is causally responsible for the elevated blood pressure. Intuitively, the answer to this question is very different: if there does not exist a causal relation, then the probability that the genetic predisposition is causally responsible for the elevated blood pressure is zero. (Note that the difference between estimates of diagnostic probability and estimates of causal responsibility holds not only when the data indicate that there is no causal relation from *C* to *E*, but also in situations in which there is a relation, i.e., w > 0.)

The difference between estimates of conditional probability and causal responsibility is intuitively plausible, but a purely statistical account of diagnostic reasoning (i.e., the simple Bayes model) lacks

 (\mathbf{r})

(�)

the expressive power for providing a formal treatment of causal responsibility. Cheng and Novick (2005; see also Holyoak et al., 2010) showed how to formally derive different estimates of causal responsibility within power PC theory. Their analyses can also be incorporated into the structure induction model of diagnostic reasoning, which allows for deriving estimates of causal responsibility that take into account structure uncertainty (Meder et al., 2014). For instance, in the case of elemental diagnostic reasoning, an estimate of causal responsibility is computed separately under structures S_0 and S_1 . According to structure S_0 there is no causal relation between C and E; therefore this structure entails that estimates of causal responsibility are zero. Under structure S_1 , the model's parameters are used to derive an estimate of causal responsibility, as in power PC theory. The final step is to integrate out the two causal structures, with the resulting estimate of causal responsibility depending on the relative posterior probabilities of the two structures.

The critical point is that depending on the goal of the diagnostic reasoner, different quantities are of interest, and these can systematically diverge from each other. A common assumption in the normative and descriptive analysis of diagnostic reasoning is that the computational goal is to assess the conditional probability of a cause given an effect. However, sometimes it might be more appropriate to ask for the probability that the effect was indeed brought about by the target cause.

(�

With respect to empirical studies, little is known about the extent to which human diagnostic reasoners are sensitive to the distinction between different types of diagnostic inferences. Holyoak and colleagues (2010) used a Bayesian variant of the power PC causal attribution account to model predictive inferences and causal attribution judgments in the context of analogical reasoning. Meder and colleagues (2014) examined the power PC model of causal responsibility in the context of elemental diagnostic reasoning, in which the goal is to infer the conditional probability of a cause given an effect, but their studies were not specifically designed to investigate different types of diagnostic inferences. Stephan and Waldmann (2016) tested which model of causal responsibility best accounts for human judgments, pitting the standard power PC model and a Bayesian variant of it against the structure induction model. The results of three studies supported the structure induction model of causal responsibility, showing that people's judgments of causal responsibility are sensitive to causal

structure uncertainty. These findings provide pathways for future research on different kinds of diagnostic inferences.

DIAGNOSTIC HYPOTHESIS GENERATION

Throughout this chapter, we discussed diagnostic reasoning in situations in which the set of variables and their causal roles (i.e., causes vs. effects) were predefined and well specified. While this assumption eases theoretical analysis as well as experimental consideration, in most real-world situations diagnostic inferences are embedded in a complex web of causally related and often unknown variables. This naturally raises the question of how diagnostic inferences might be carried out under such circumstances and how the causal structures on which the inferences operate are learned or determined.

The most unconstrained way is to infer the relevant causal structure (i.e., causal roles of the variables, as well as the relations between them) directly from data, such as patterns of co-occurrences, as the causal structure of the world imposes constraints on the data that can potentially be observed (e.g., Gopnik et al., 2004; Steyvers et al., 2003). However, the number of possible causal structures that have to be taken into account grows exponentially with the number of variables considered, which poses several computational challenges. In addition, humans have been shown to fail in contingency-based structure induction, even in quite simple cases (Fernbach & Sloman, 2009; White, 2006), unless specific constraints or assumptions are met (e.g., determinism; Deverett & Kemp, 2012; Mayrhofer & Waldmann, 2015b; Rothe, Deverett, Mayrhofer, & Kemp, 2016).

From a psychological perspective, it seems plausible that humans consider only a subset of possible causal structures, with different cues to causality constraining the hypothesis space, such as temporal information (Lagnado & Sloman, 2004, 2006), hierarchical event structures (Johnson & Keil, 2014), linguistic markers in causal language (Mayrhofer & Waldmann, 2015a), and prior knowledge (Waldmann, 1996). In line with this idea, Griffiths and Tenenbaum (2007; Tenenbaum, Griffiths, & Niyogi, 2007) proposed a "causal grammar" that specifies the variables that form the causal structure, the possible relations between the domain variables (i.e., their causal roles), and the functional form of the considered relations. This knowledge is at a higher level of abstraction than a specific causal structure hypothesis, much like a grammar in language constrains the set of possible sentences. This approach can be formalized as a hierarchical

Bayesian model, in which abstract knowledge about the domain generates and constrains the hypothesis space over the causal structures, thereby addressing the problem of combinatorial explosion (Griffiths & Tenenbaum, 2009).

A very different approach addresses the question of diagnostic hypothesis generation and evaluation from the perspective of memory processes and cue recall. According to the HyGene model (Thomas, Dougherty, Sprenger, & Harbison, 2008), diagnostic hypotheses in long-term semantic memory are activated, matched against sampled probes (i.e., previously encountered and stored diagnosis-cue sets) from episodic memory, and then potentially placed in working memory (constituting the set of leading contending hypotheses) in an iterative fashion. In the end, the diagnostic judgment for the target hypothesis is computed relative to the memory strengths of the alternatives in working memory. This model is non-causal in nature, as the causal relations between hypothesis (to-bediagnosed causes) and cues are not relevant for judgments; essentially it can be applied-just like the simple Bayes model—to any arbitrarily related set of hypothesis and data.

Concluding Remarks

(�)

Research on diagnostic reasoning has a long tradition in psychology. Much of the literature on judgment and decision-making has focused on the conditions under which people utilize base rate information and make judgments in accordance with a simple statistical model, Bayes's rule. We think it is time to take a fresh look at the problem of diagnostic reasoning from the perspective of causal inference under uncertainty. The framework of probabilistic inference over graphical causal models provides a strong formal foundation for modeling diagnostic reasoning, and a variety of empirically testable models can be realized within this computational framework. The discussion on elemental diagnostic reasoning illustrates that there is not a single normative benchmark for diagnostic reasoning under uncertainty against which human behavior can be evaluated, but that different ideas exist about what may constitute an appropriate standard of rational inference from effect(s) to cause(s). Importantly, these models make diverging predictions, for instance, on whether a diagnostic judgment should reflect solely the observed probability of a cause given an effect. A key goal for future research is to systematically investigate the descriptive validity of the alternative accounts in

different circumstances, as well as their theoretical behavior under different conditions. This will support the development of a comprehensive theory of human diagnostic reasoning that is informed and constrained by normative considerations and empirical data.

Author Note

This research was supported by grants ME 3717/2-2 and Ma 6545/1-2 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516). We thank Michael Waldmann for helpful feedback on this chapter and Anita Todd for editing the manuscript.

Notes

 (\mathbf{r})

- The noisy-OR gate assumes independent causes C and A; thus, the probability P(e|c) is given by w_c + w_a - w_cw_a, because when C is present E is present, when either C produced it (with probability w_c) or the background A generated it (with probability w_a); the last term corrects for double counting cases in which both causes brought about E (with probability w_cw_a).
- For the data set in Figure 22.1 b, the maximum likelihood estimates (MLEs) for the parameters of structure S₁ (Figure 22.1C) are b_c = 0.5, w_c = 0.222, and w_a = 0.1. Plugging these number into Equation 4 yields P(c|e) = .75.
- 3. From a mathematical perspective, the structure induction model contains Bayesian variants of power PC theory (e.g., Holyoak et al., 2010; Lu et al., 2008) as special cases. If the prior for structure S_1 is set to 1, structure S_0 plays no role when integrating out the structures to obtain a single estimate for the diagnostic probability P(c|e). Note, however, that certain technical differences exist between proposed Bayesian variants of power PC theory and the structure induction model, such as the used priors (e.g., so-called sparse-and-strong priors in Lu et al., 2008, vs. uniform priors in the structure induction model in Meder et al., 2014).
- 4. A related term is *discounting*, which in the literature has sometimes been used interchangeably with the notion of explaining away, but also has been used to describe different empirical phenomena, such as variations in causal strength judgments of a target cause in the presence of alternative causes (e.g., Goedert, Harsch, & Spellman, 2005). See Khemlani and Oppenheimer (2011) for a review of the use of both terms and an overview of different models and findings. We here focus on explaining away as conceptualized in the context of diagnostic inference over graphical causal models.
- 5. More generally, estimating individual causal powers in situations with multiple causes according to Equation 2 requires conditioning on the absence of the alternative causes to determine the relevant strength estimates. See Cheng and Novick (1990, 1992; Melz, Cheng, Holyoak, & Waldmann, 1993) for details; see also Novick and Cheng (2004) for a formal analysis of strength estimates when causes are interacting.
- 6. An alternative way would be to parameterize the model with causal strength estimates, analogous to the cases

discussed earlier. The computations for deriving the information value of different queries as discussed later can then be based on these parameters (e.g., derivation of diagnostic probabilities via causal strength estimates; e.g., Equation 4). To simplify matters, we here consider only the case in which the causal structure is parameterized by conditional probabilities.

- Other entropy measures besides Shannon (1948) could be used, for instance, the Rényi (1961) or Tsallis (1988) families of entropy measures (for a detailed discussion, see Crupi, Nelson, Meder, Cevolani, & Tentori, 2016).
- 8. If the patient has fever, the entropy increases, because the posterior probability distribution over the cause (virus) is close to uniform (.49 vs. .51, respectively; Figure 22.4 c). Therefore, this datum entails a negative information gain (i.e., an increase in uncertainty about the true state of the cause variable). (The entropy of a binary random variable is maximal when the distribution is uniform, i.e., both states are equiprobable.) Conversely, if the patient has no fever, this datum decreases the entropy, because conditional on the absence of fever it is very likely that the virus is not present (.07 vs. .93; Figure 22.4 c). To compute the expected information gain of testing for fever, the individual gains are integrated by weighing each gain with the probability of observing each of the two states, fever vs. ¬fever (.55 vs. .45; Figure 22.4 c). (Values in the text are not based on the rounded values in the tree in Figure 22.4 c but are the exact values.)
- 9. The probability gain model assumes that the diagnostic reasoner always selects the more likely hypothesis, i.e., uses an *arg-max* decision rule. Accuracy based on the prior distribution of the cause is .7. If fever is present, accuracy decreases to .51; the probability gain relative to the prior is negative. By contrast, when the patient does not have fever, accuracy increases to .93; accordingly, the probability gain is positive. To compute the overall probability gain of the query, an expectation is computed based on taking into account the probability of each state (i.e., that a patient does or does not have fever).

References

- Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality* and Social Psychology, 35, 303–314. doi:10.1037/ 0022-3514.35.5.303
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science*, 35, 499–526. doi:10.1111/j.1551-6709.2010.01161.x
- Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual process. *Behavioral and Brain Sciences*, 30, 241–297. doi:10.1017/S0140525X07001653
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–233. doi:10.1016/ 0001-6918(80)90046-3
- Baron, J. (1985). Rationality and intelligence. Cambridge, UK: Cambridge University Press.
- Baron, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: I. Priors, error costs, and test accuracy. Organizational Behavior and Human Decision Processes, 41, 259–279. doi:10.1016/0749-5978(88)90030-1
- Benish, W. A. (1999). Relative entropy as a measure of diagnostic information. *Medical Decision Making*, 19, 202–206. doi:10.1177/0272989X9901900211

- Bocklisch, F., Bocklisch, S. F., & Krems, J. F. (2012). Sometimes, often, and always: Exploring the vague meanings of frequency expressions. *Behavior Research Methods*, 44, 144–157. doi:10.3758/s13428-011-0130-8
- Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 708–731. doi:10.1037/xlm0000061
- Brighton, H., & Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds? In S. Okasha & K. Binmore (Eds.), *Evolution and rationality: Decisions, cooperation, and strategic behaviour* (pp. 84–109). Cambridge, UK: Cambridge University Press.
- Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3, 57–65. doi:10.1016/S1364-6613(98)01273-X
- Chater, N., & Oaksford, M. (2008). (Eds.). The probabilistic mind. New York: Oxford University Press.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405. doi:10.1037//0033-295X.104.2.367
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58, 545–567. doi:10.1037/0022-3514.58.4.545
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365–382. doi:10.1037/0033-295X.99.2.365
- Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal reasoning: Reply to White (2005) and to Luhmann & Ahn (2005). *Psychological Review*, 112, 694–707. doi:10.1037/0033-295X.112.3.694
- Chickering, D. M., & Heckerman, D. (1997). Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 29, 181–212. doi:10.1023/A:1007469629108
- Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, 79, 102–133. doi:10.1016/j.cogpsych.2015.02.004
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7, 25–47.
- Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2016). Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. Manuscript in preparation.
- Crupi, V., & Tentori, K. (2014). State of the field: Measuring information and confirmation. *Studies in History and Philosophy of Science*, 47(C), 81–90. doi:10.1016/j.shpsa.2014.05.002
- Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review*, 116, 971–985. doi:10.1037/a0017050
- Deverett, B., & Kemp, C. (2012). Learning deterministic causal networks from observational data. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 288–293). Austin, TX: Cognitive Science Society.
- Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43, 111–121. doi:10.1016/0001-6918(79)90017-9
- Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29, 103–130. doi:10.1023/A:1007413511361

MEDER AND MAYRHOFER | 45

453

9780199399550_Waldmann290716MEDUS_Book.indb 453

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). New York: Cambridge University Press.

- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: John Wiley & Sons.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, 21, 329–336. doi:10.1177/ 0956797610361430
- Fernbach, P. M., Darlow A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, 140, 168–185. doi:1037/ a0022100
- Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 678–693. doi:10.1037/a0014928
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704. doi:10.1037/ 0033-295X.102.4.684
- Glymour, C. (2003). Learning, prediction and causal Bayes nets. Trends in Cognitive Science, 7, 43–47. doi:10.1016/ S1364-6613(02)00009-8
- Goedert, K. M., Harsch, J., & Spellman, B. A. (2005). Discounting and conditionalization: Dissociable cognitive processes in human causal inference. *Psychological Science*, 18, 590–595. doi:10.1111/j.1467-9280.2005.01580.x
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111, 3–32. doi:10.1037/0033-295X.111.1.3
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 334– 384. doi:10.1016/j.cogpsych.2005.05.004
- Griffiths, T. L., & Tenenbaum, J. B. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 323– 346). Oxford, UK: Oxford University Press.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116, 661–716. doi:10.1037/ a0017201
- Hayes, B. K., Hawkins, G. E., & Newell, B. R. (2015, November 16). Consider the alternative: The effects of causal knowledge on representing and using alternative hypotheses in judgments under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 723–739. doi:10.1037/xlm0000205
- Hayes, B. K., Hawkins, G. E., Newell, B. R., Pasqualino, M., & Rehder, B. (2014). The role of causal models in multiple judgments under uncertainty. *Cognition*, 133, 611–620. doi:10.1037/0096-3445.134.4.596
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1–55. doi:10.1016/0010-0285(92)90002-J
- Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, 139, 702–727. doi:10.1037/a0020488

- Jahn, G., & Braatz, J. (2014). Memory indexing of sequential symptom processing in diagnostic reasoning. *Cognitive Psychology*, 68, 59–97. doi:10.1016/j.cogpsych.2013.11.002
- Jahn, G., Stahnke, R., & Rebitschek, F. G. (2014). Parallel belief updating in sequential diagnostic reasoning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), Proceedings of the 36th annual conference of the Cognitive Science Society (pp. 2405–2410). Austin, TX: Cognitive Science Society.
- Jarecki, J., Meder, B., & Nelson, J. D. (2013). The assumption of class-conditional independence in category learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 2650–2655). Austin, TX: Cognitive Science Society.
- Jarecki, J., Meder, B., & Nelson, J. D. (2016). Naïve and robust: Class-conditional independence in human classification learning. Manuscript submitted for publication.
- Johnson, S. G. B., & Keil, F. C. (2014). Causal inference and the hierarchical structure of experience. *Journal of Experimental Psychology: General*, 143, 2223–2241. doi:10.1037/ a0038192
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1–24. doi:10.1016/0022-1031(67)90034-0
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34, 169–188. doi:10.1017/ S0140525X10003134
- Josephson, J. R., & Josephson, S. G. (1996). Abductive inference: Computation, philosophy, technology. Cambridge, UK: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454. doi:10.1016/0010-0285(72)90016-3
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. doi:10.1037/ h0034747
- Khemlani, S. S., & Oppenheimer, D. M. (2011). When one model casts doubt on another: A levels-of-analysis approach to causal discounting. *Psychological Bulletin*, 137, 195–210. doi:10.1037/a0021809
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228. doi:10.1037/0033-295X.94.2.211
- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, 19, 1–54. doi:10.1017/ S0140525X00041157
- Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, 136, 430–450. doi:10.1037/ 0096-3445.136.3.430
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22, 79–86. doi:0.1214/aoms/1177729694
- Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 856–876. doi:10.1037/ 0278-7393.30.4.856
- Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 451–460. doi:0.1037/0278-7393.32.3.451

454 | DIAGNOSTIC REASONING

- Lindley, D. V. (1956). On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27, 986-1005. doi:10.1214/aoms/1177728069
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. Medical Decision Making, 21, 37-44. doi:10.1177/ 0272989X0102100105
- 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. doi:10.1037/ a0013256
- Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. Cognitive Science, 34, 113-147. doi:10.1111/ j.1551-6709.2009.01058.x
- Markant, D. B., Settles, B. and Gureckis, T. M. (2015). Selfdirected learning favors local, rather than global, uncertainty. Cognitive Science, 40, 100-120. doi:10.1111/cogs.12220
- Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco, CA: Freeman.
- Mayrhofer, R., & Waldmann, M. R. (2011). Heuristics in covariation-based induction of causal models: Sufficiency and necessity priors. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), Proceedings of the 33rd annual conference of the Cognitive Science Society (pp. 3110-3115). Austin, TX: Cognitive Science Society.
- Mayrhofer, R., & Waldmann, M. R. (2015a). Agents and causes: Dispositional intuitions as a guide to causal structure. Cognitive Science, 39, 65-95. doi:10.1111/cogs.12132
- Mayrhofer, R., & Waldmann, M. R. (2015b). Sufficiency and necessity assumptions in causal structure induction. Cognitive Science. doi:10.1111/cogs.12318

(4)

- McKenzie, C. R. M. (2003). Rational models as theories-not standards-of behavior. Trends in Cognitive Sciences, 7, 403-406. doi:10.1016/S1364-6613(03)00196-7
- McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. Cognitive Psychology, 54, 33-61. doi:10.1016/j.cogpsych.2006.04.004
- McNair, S., & Feeney, A. (2014). When does information about causal structure improve statistical reasoning? Quarterly Journal of Experimental Psychology, 67, 625-645. doi:10.1080/17470218.2013.821709
- McNair, S., & Feeney, A. (2015). Whose statistical reasoning is facilitated by a causal structure intervention? Psychonomic Bulletin & Review, 22, 258-264. doi:10.3758/ s13423-014-0645-y
- Meder, B., & Gigerenzer, G. (2014). Statistical thinking: No one left behind. In E. J. Chernoff & B. Sriraman (Eds.), Advances in mathematics education: Probabilistic thinking: Presenting plural perspectives (pp. 127-148). Dordrecht, Netherlands: Springer.
- Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. Psychonomic Bulletin & Review, 15, 75-80. doi:10.3758/ PBR.15.1.75
- 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. doi:10.3758/MC.37.3.249
- Meder, B., & Mayrhofer, R. (2013). Sequential diagnostic reasoning with verbal information. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th

annual conference of the Cognitive Science Society (pp. 1014-1019). Austin, TX: Cognitive Science Society.

- Meder, B., Mayrhofer, R., & Waldmann, M. R. (2009). A rational model of elemental diagnostic inference. In N. A. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st annual conference of the Cognitive Science Society (pp. 2176-2181). Austin, TX: Cognitive Science Society.
- Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. Psychological Review, 121, 277-301. doi:10.1037/a0035944
- Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. Judgment and Decision Making, 7, 119-148.
- Meier, K. M., & Blair, M. R. (2013). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. Cognition, 126, 319-325. doi:10.1016/ j.cognition.2012.09.014
- Melz, E. R., Cheng, P. W., Holyoak, K. J., & Waldmann, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla-Wagner Learning Rule? Comment on Shanks (1991). Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1398-1410. doi:10.1037/0278-7393.19.6.1398
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. Psychological Review, 102, 331-355. doi:10.1037/0033-295X.102.2.331
- Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. Statistical Science, 5, 2-12. doi:10.1214/ss/ 1177012251
- Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. Psychological Review, 112, 979-999. doi:10.1037/ 0033-295X.112.4.979
- Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. Cognition, 130, 74-80. doi:10.1016/j.cognition.2013.09.007
- Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. Psychological Science, 21, 960-969. doi:10.1177/0956797610372637
- Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal power. Psychological Review, 111, 455-485. doi:10.1037/0033-295X.111.2.455
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631. doi:10.1037/0033-295X.101.4.608
- Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. Cognitive Psychology, 67, 186-216. doi:10.1016/j.cogpsych.2013.09.002
- Pearl, J. (1988). Probabilistic reasoning in intelligent systems. San Francisco, CA: Morgan-Kaufmann.
- Pearl, J. (2000). Causality: Models, reasoning and inference. Cambridge, UK: Cambridge University Press.
- Peterson, C. R., & Beach, L. R. (1967). Man as intuitive statistician. Psychological Bulletin, 68, 29-46. doi:10.1037/h0024722
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. Journal of Experimental Psychology, 72, 346-354. doi:10.1037/h0023653
- Raiffa, H., & Schlaifer, R. O. (1961). Applied statistical decision theory. Cambridge, MA: Division of Research, Graduate School of Business Administration, Harvard University.

MEDER AND MAYRHOFER 455

9780199399550_Waldmann290716MEDUS_Book.indb 455

Rebitschek, F. G., Bocklisch, F., Scholz, A., Krems, J. F., & Jahn, G. (2015). Biased processing of ambiguous symptoms favors the initially leading hypothesis in sequential diagnostic reasoning. *Experimental Psychology*, 62, 287–305. doi:10.1027/ 1618-3169/a000298

- Rebitschek, F. G., Krems, J. F., & Jahn, G. (2015). Memory activation of multiple hypotheses in sequential diagnostic reasoning, *Journal of Cognitive Psychology*, 27, 780–796. doi:10.1080/20445911.2015.1026825
- Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, 27, 709–748. doi:10.1016/ S0364-0213(03)00068-5
- Rehder, B. (2010). Causal-based classification: A review. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 39–116). San Diego, CA: Academic Press.
- Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, 72, 54–107. doi:10.1016/j.cogpsych.2014.02.002
- Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, 50, 264– 314. doi:10.1016/j.cogpsych.2004.09.002
- Rehder, B., & Waldmann, M. R. (in press). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*.
- Rényi, A. (1961). On measures of entropy and information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 547–561.
- Rothe, A., Deverett, B., Mayrhofer, R., & Kemp, C. (2016). Successful structure learning from observational data. Unpublished manuscript.
- Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences in causal networks. *Psychological Bulletin*, 140, 109–139. doi:10.1037/a0031903

(�

- Rottman, B. M., & Hastie, R. (2015). Do Markov violations and failures of explaining away persist with experience? In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th* annual conference of the Cognitive Science Society (pp. 2027– 2032). Austin, TX: Cognitive Science Society.
- Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology*, 87, 88–134. doi:10.1016/j.cogpsych.2016.05.002
- Rusconi, P., & McKenzie, C. R. M. (2013). Insensitivity and oversensitivity to answer diagnosticity in hypothesis testing. *The Quarterly Journal of Experimental Psychology*, 66, 2443– 2464. doi:10.1080/17470218.2013.793732
- Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130, 380–400. doi:10.1037// 0096-3445.130.3.380
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Stephan, S., & Waldmann, M. R. (2016). Answering causal queries about singular cases. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2795–2801). Austin, TX: Cognitive Science Society.
- Sloman, S. A., & Lagnado, D. A. (2005). Do we "do?" Cognitive Science, 29, 5–39. doi:10.1207/s15516709cog2901_2
- Spirtes, P., Glymour, C., & Scheines, P. (1993). Causation, prediction, and search. New York: Springer.

(

- Spohn, W. (1976/1978). Grundlagen der Entscheidungstheorie, Dissertation at the University of Munich 1976, published: Kronberg/Ts.: Scriptor, 1978.
- Spohn, W. (2001). Bayesian nets are all there is to causal dependence. In M. C. Galavotti, P. Suppes, & D. Costantini (Eds.), *Stochastic dependence and causality* (pp. 157–172). Stanford, CA: CSLI Publications.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27, 453–489. doi:10.1207/ s15516709cog2703_6
- Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 301–322). Oxford, UK: Oxford University Press.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, 115, 155–185. doi:10.1037/ 0033-295X.115.1.155
- Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, 35, 1518–1552. http://doi.org/10.1111/ j.1551-6709.2011.01197.x
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52, 479–487. doi:10.1007/BF01016429
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. doi:10.1126/science.185.4157.1124
- Tversky, A., & Kahneman, D. (1982a). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117–128). New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1982b). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153– 160). New York: Cambridge University Press.
- von Sydow, M., Hagmayer, Y., & Meder, B. (2015). Transitive reasoning distorts induction in causal chains. *Memory* & Cognition, 44, 469–487. http://doi.org/10.3758/ s13421-015-0568-5
- 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, pp. 47–88). San Diego, CA: Academic Press.
- Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science*, 31, 233–256. doi:10.1080/15326900701221231
- Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. In N. Chater & M. Oaksford (Eds.), *The* probabilistic mind: Prospects for Bayesian cognitive science (pp. 453–484). Oxford, UK: Oxford University Press.
- Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 216–227. doi:10.1037/0278-7393.31.2.216
- 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.

456 | DIAGNOSTIC REASONING

- OUP UNCORRECTED PROOF REVISES, Thu Feb 09 2017, NEWGEN
- Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, 15, 307–311.
- 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.
- Walsh, C., & Sloman, S. (2008). Updating beliefs with causal models: Violations of screening off. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A Festschrift for Gordon H. Bower* (pp. 345–357). New York: Erlbaum.
- Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20, 273–281. doi:10.1080/ 14640746808400161
- Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88, 776–784. doi:10.1037/0033-2909.88.3.776
- White, P. (2006). How well is causal structure inferred from co-occurrence information? *European Journal of Cognitive Psychology*, 18, 454–480. doi:10.1080/09541440500264861
- Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, 76, 1–29. doi:10.1016/j.cogpsych.2014.11.001

()

۲