7.3 Diagnostic Causal Reasoning

Björn Meder and Ralf Mayrhofer

Summary

Rational theories of diagnostic reasoning assume that the reasoner's goal is to infer the conditional probability of a cause given an effect from the available data. Typically, diagnostic reasoning is modeled within a statistical inference framework, with Bayes' rule applied to the obtained covariation information serving as the normative standard. This chapter analyzes diagnostic reasoning from the perspective of causal induction, using the framework of causal Bayes net theory to instantiate different accounts of rational diagnostic reasoning. These approaches elucidate the relevant kinds of inputs, computations, and outputs by differentiating between parametric causal models and observable contingency information. A particularly interesting feature of these accounts is that they can include predictions that systematically deviate from the traditional, purely statistical norm. The analyses highlight key issues for constructing a rational theory of diagnostic reasoning and the experimental study of human rationality.

1. Reasoning from Effect to Cause

Diagnostic causal reasoning pertains to inferences from effect to cause, such as reasoning from symptoms to diseases in medical diagnosis. Probabilistic diagnostic inferences can be considered a special case of Bayesian inference, where beliefs about unobserved states of the world (the cause events) are updated in light of observed data (the effect events). The question to what extent people can make appropriate diagnostic inferences has been central to debates on human rationality, with probability theory, and Bayes' rule in particular, serving as a normative or descriptive reference point (Gigerenzer, 1996; Kahneman & Tversky, 1996). Recent advances in causal modeling have provided new insights, with respect to both long-standing norms of rationality and the descriptive adequacy of models of human diagnostic reasoning. *Causal Bayes net theory*, that is, probabilistic inference over causal graphical models, provides a formal framework for representing causal dependencies and modeling different kinds of probabilistic causal inferences (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Spohn, 1976/1978, as cited in Spohn, 2001; see also chapter 4.2 by Hartmann, this handbook). The framework explicates the relations between observed covariation information and an underlying causal model that generates the data. Importantly, causal Bayes net theory has the expressive power to instantiate different rational models of diagnostic inference, thereby contesting the idea of a single normative benchmark for evaluating the rationality of human diagnostic reasoning (Meder & Mayrhofer, 2017b; Meder, Mayrhofer, & Waldmann, 2014; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008). This chapter highlights the ways in which a causal analysis of diagnostic reasoning can inform issues of theoretical rationality and guide empirical research.

2. Rational Models of Diagnostic Reasoning

Different models of diagnostic reasoning have been postulated that serve the dual purpose of providing normative standards and constituting candidate models of human cognition. Common to these approaches is the goal to infer the conditional probability of a cause given an effect. They critically differ, however, with respect to their assumptions regarding the relation between unobservable causal structures and observable covariation information, as well as the involved representations and computations.

2.1 Diagnostic Reasoning as Statistical Inference

The most elemental form of diagnostic reasoning involves inferences from a binary (present vs. absent) effect event E to a binary cause event C. For instance, the cause event C could be a disease, with c and $\neg c$ denoting the presence and absence, respectively, of the disease, and the effect event E could be a particular symptom, with e and $\neg e$ denoting the presence and absence, respectively, of

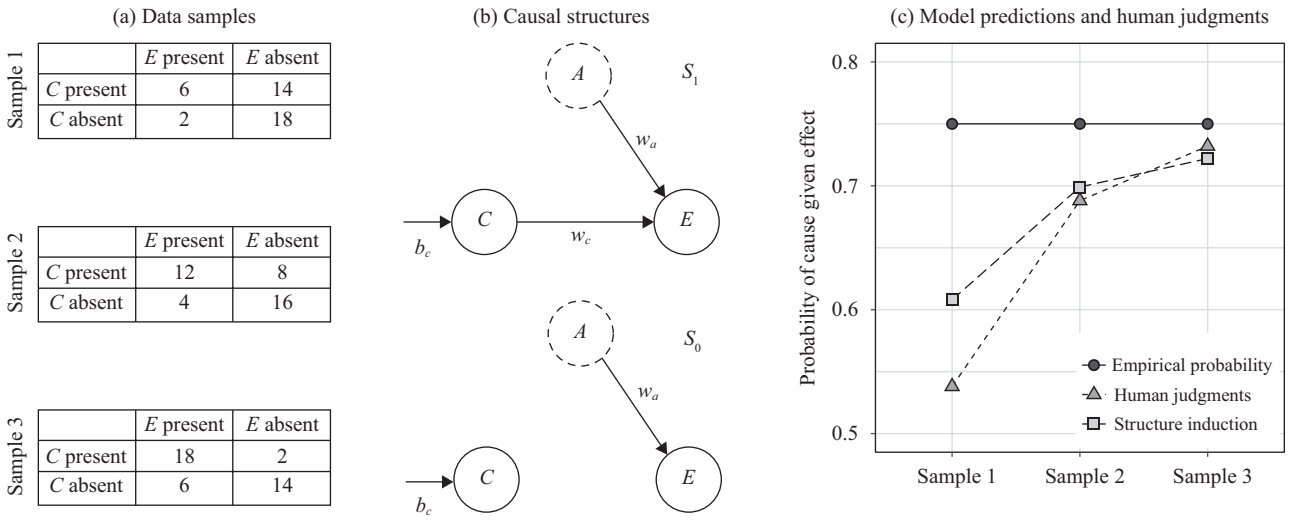


Figure 7.3.1

(a) Three joint frequency distributions over cause C and effect E. In each data sample, the diagnostic probability of cause C given effect E is 0.75. (b) Alternative causal structures that may have generated the data. According to structure S_1, there is a causal relation between C and E, as well as background causes A that can cause E independently of C. According to structure S_0, candidate cause C and effect E are independent; any observed empirical contingency is merely coincidental and caused by background causes A. (c) Model predictions and empirical results. The empirical probability of cause given effect is 0.75 in all three data samples. If uncertainty about possible causal structures is taken into account, different diagnostic probabilities result, depending on the extent to which the data warrant the existence of a causal relation (i.e., supports S_1 over S_0). Mean human judgments are not invariant across the three data sets, suggesting that people are sensitive to uncertainty about causal structure (Meder et al., 2014, experiment 2).

the symptom. Here, we consider the situation where the diagnostic inference is based on a sample of observed data. For instance, a doctor may have data on how many patients with the disease have a particular symptom, how many without the disease have the symptom, and so on. Given this covariation information, the diagnostic probability of cause given effect, $P(c \mid e)$, can be inferred using Bayes' rule:

$$P(c \mid e) = \frac{P(e \mid c) \cdot P(c)}{P(e \mid c) \cdot P(c) + P(e \mid \neg c) \cdot P(\neg c)} = \frac{P(e \mid c) \cdot P(c)}{P(e)}, \quad (1)$$

where $P(c)$ denotes the prior probability (base rate) of the cause, $P(e \mid c)$ is the likelihood of the effect given that the cause is present, and $P(e \mid \neg c)$ is the likelihood of the effect if the cause is absent. Consider the data sample shown in figure 7.3.1a (top). According to these empirical frequencies, $P(c) = 20/40 = 0.5$, $P(e \mid c) = 6/20 = 0.3$, and $P(e \mid \neg c) = 2/20 = 0.1$. Plugging these values into equation (1) yields the diagnostic probability $P(c \mid e) = 0.75$.

Whether people reason in line with Bayes' rule has been pivotal to disputes about human rationality. Studies in the 1950s and 1960s indicated that subjects updated their beliefs to a lesser extent than prescribed by Bayes' rule (Edwards, 1968; Phillips & Edwards, 1966). By and large, though, the experimental findings were considered evidence for sound probabilistic reasoning, giving rise to the metaphor of "man as intuitive statistician" (Peterson & Beach, 1967). Researchers in the heuristics-and-biases program, however, came to a rather different conclusion, arguing that typically people's diagnostic inferences are not in line with classic Bayesian norms (e.g., Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974; for a critical review, see Koehler, 1996). The discrepancy has led researchers to more precisely characterize the conditions under which people can make appropriate diagnostic inferences, for instance, when probabilistic information is conveyed in particular frequency formats (Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017; Meder & Gigerenzer, 2014).

The conditional probability of cause given effect derived by applying Bayes' rule to verbally described probabilities or sample data has been endorsed by many researchers as the normative standard, even when they otherwise debate whether and under what conditions people can solve such tasks. Next, we analyze diagnostic reasoning from the perspective of inductive causal inferences, using causal Bayes net theory to implement different candidate models for a rational account of

diagnostic reasoning. These approaches differentiate between parametric causal models and observable contingency information and can therefore lead to very different predictions than accounts that derive the conditional probability of cause given effect directly from the observed empirical data without any reference to an underlying causal model.

2.2 Diagnostic Reasoning as Causal Inference

One goal of rational agents is to acquire knowledge about the causal structure of noisy environments, in order to support prediction, diagnosis, and control (Waldmann, 2017; chapter 7.2 by Waldmann, this handbook). In this view, diagnostic inferences operate on representations that preserve the directionality of causal relations (as opposed to undirected statistical or associative relations). This distinguishes the account from Bayes' rule, which is applicable to arbitrary statistically related events and does not make any reference to possible causal relations that may underlie the observed data.

The standard causal model for situations involving a binary cause event and a binary effect event is structure S_1 in figure 7.3.1b. This graph states that there is a causal relation between C and E, as well as an amalgam of unobserved background causes A that occur independently of C and can also generate E. Cheng (1997; see also Glymour, 2003; Griffiths & Tenenbaum, 2005; Novick & Cheng, 2004) showed that the generative *causal power* of a cause—the unobservable probability with which C produces E—can be estimated according to

$$w_c = \frac{P(e \mid c) - P(e \mid \neg c)}{1 - P(e \mid \neg c)}, \tag{2}$$

where w_c is the causal power of C with respect to E. The diagnostic probability of cause given effect can be inferred from the parameterized causal structure:

$$\begin{aligned} P(c \mid e) &= \frac{P(e \mid c) \cdot P(c)}{P(e \mid c) \cdot P(c) + P(e \mid \neg c) \cdot P(\neg c)} \\ &= \frac{w_c b_c + w_a b_c - w_c w_a b_c}{w_c b_c + w_a - w_c w_a b_c}, \end{aligned} \tag{3}$$

where w_c is the causal power of C with respect to E, b_c denotes the base rate (prior probability) of the cause, and w_a denotes the strength of the background cause A.

The quantities used to parameterize the causal structure are typically single-point estimates derived directly from the empirical frequencies (i.e., maximum likelihood estimates of $P(c)$, $P(e \mid c)$, $P(e \mid \neg c)$, and, hence, w_c; see, e.g., Cheng, 1997). In this case, the conditional probability of cause given effect derived from the parameterized causal graph exactly corresponds to the values obtained from directly applying Bayes' rule to the empirical frequencies, although diagnostic reasoning takes place on the causal rather than the data level (Meder et al., 2014).[1] This is not necessarily the case, though, if uncertainty about causal parameters or alternative generative causal models is incorporated in the diagnostic inference process.

2.3 Diagnostic Reasoning: Parameter Uncertainty and Causal Structure Uncertainty

Analyzing diagnostic reasoning from the perspective of causal induction enables consideration of different kinds of uncertainty in the inference process. *Parameter uncertainty* arises if estimates are derived from limited and potentially noisy data. Formally, this type of uncertainty can be modeled by using probability *distributions* for representing causal parameters, rather than single-point estimates, with the parameter distributions being updated in light of the data using Bayesian inference (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008). Importantly, the posterior distributions quantify and represent uncertainty explicitly (e.g., through the variance of the distributions).

Another type of uncertainty is *structure uncertainty*, which pertains to possible causal models that may underlie the observed data. The *structure induction model of diagnostic reasoning* (Meder et al., 2014) takes this into account by considering two causal graphs (figure 7.3.1b): instead of using only the default structure S_1, the model also considers an alternative structure S_0 according to which there is *no* causal relation between C and E. The intuition behind this is that an observed contingency between C and E may not be indicative of a causal relation but merely coincidental, and rational agents should take this into account. The data are then used to estimate the causal structures' parameters through Bayesian inference, and under each parameterized structure, an estimate of the diagnostic probability is derived. Under S_1, assuming uniform prior distributions over the parameters, the computed probability will approximate the empirically observed $P(c \mid e)$. Structure S_0, by contrast, states that C and E are independent events; therefore, observing E does not provide diagnostic evidence for C (i.e., $P(c \mid e) = P(c)$). The structure induction model computes how likely each of the two structures is, given the data, and forms a weighted average of the entailed diagnostic probabilities, with the resulting estimate taking into account both parameter and structure uncertainty.

Depending on the relative probability of S_0 and S_1, the inferred diagnostic probability can strongly diverge from

the empirical probability. Figure 7.3.1c illustrates this for the three data sets shown in figure 7.3.1a: in all three data sets, the empirical probability is $P(c \mid e) = 0.75$; therefore, a rational agent might conclude that the diagnostic probability of cause given effect is the same in all three situations. Consideration of alternative causal structures leads to very different inferences. For instance, for sample 1, the structure induction model entails a diagnostic probability of 0.61, much lower than the empirical probability of 0.75. This discrepancy arises from the fact that the contingency between cause and effect is relatively weak, so that structures S_0 and S_1 are almost equally likely to have generated the data. For samples 2 and 3, the discrepancy is smaller, as these data indicate a stronger contingency between C and E, making it more likely that S_1 is indeed the true generating model.

Interestingly, human diagnostic judgments mirror structure uncertainty: judgments strongly vary, although the empirical probability of cause given effect is identical across the three data sets (figure 7.3.1c; for details, see Meder et al., 2014). From the perspective of the classic statistical inference perspective, this response pattern looks irrational. By contrast, viewed as resulting from a causal inference strategy that is adapted to the uncertainties of the world outside the laboratory, the judgments reflect rational reasoning.

3. Rational Models of Causal Attribution

Theories of diagnostic reasoning typically assume that the goal of the reasoner is to infer the conditional probability of cause given effect. In many diagnostic reasoning scenarios, however, it is judgments of the *causal responsibility* (or *causal attribution*) that are of interest, that is, judgments of the probability that a candidate cause brought about the effect. For instance, instead of merely assessing the probability of a particular genetic disposition's being present, a doctor may want to find out whether that disposition is the cause of a patient's symptoms. This quantity is different from the diagnostic probability of cause given effect. For instance, if there is no causal relation (i.e., $P(e \mid c) = P(e \mid \neg c)$ and, therefore, $w_c = 0$), it holds that $P(c \mid e) = P(c)$. But, intuitively, if there is no causal dependency, then the probability that C produced E is zero.

Whereas a purely statistical account lacks the expressive power to model judgments of causal attribution, they can be formalized within a causal modeling approach (Cheng & Novick, 2005). Let $c \rightarrow e$ denote that the presence of effect E is generated by the presence of cause C. The query of whether the occurrence of effect E can be attributed to cause C translates to determining the conditional probability $P(c \rightarrow e \mid e)$. Given parameterized structure S_1, this quantity can be computed as follows (for details and for further measures of causal responsibility, see Cheng & Novick, 2005):

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

In its original formulation, parameters b_c, w_a, and w_c are maximum likelihood point estimates directly derived from the empirical data.[2] Bayesian variants of the model (Holyoak, Lee, & Lu, 2010) can incorporate parameter uncertainty, and the computations can also be incorporated into the structure induction model of diagnostic reasoning, in which case estimates of causal responsibility take into account both parameter and structure uncertainty (Meder et al., 2014). This approach has been successfully used to account for people's judgments of causal attribution about singular cases (Stephan & Waldmann, 2016, 2018).

4. Diagnostic Reasoning in Complex Causal Networks

Real-world scenarios typically involve complex causal networks relating multiple causes and multiple effects, such as diagnostic reasoning with multiple symptoms and multiple possible diseases. In the causal Bayes net framework, the factorization of the joint probability distribution over the considered variables is determined by the causal relations in the graph. This follows from applying the *causal Markov condition*, according to which the state of any variable in the graph is a function only of its direct causes, rendering it independent of all other variables except its direct and indirect effects (Hausman & Woodward, 1999; for a critique, see Cartwright, 1989, 1999). For instance, in the common-effect model shown in figure 7.3.2a, causes A and C are unconditionally independent but dependent conditional on their common effect B. In the common-cause model (figure 7.3.2b), A and C are two effects of a common cause B. If the Markov condition holds, this implies that A and C are unconditionally dependent but independent conditional on their common cause B (B "screens off" the correlation between A and C; Reichenbach, 1956). Similarly, in the causal-chain model (figure 7.3.2c), A and C are unconditionally dependent but conditionally independent given B.

The causality-based factorization implies particular dependence and independence relations among the variables, which facilitate and constrain diagnostic inferences

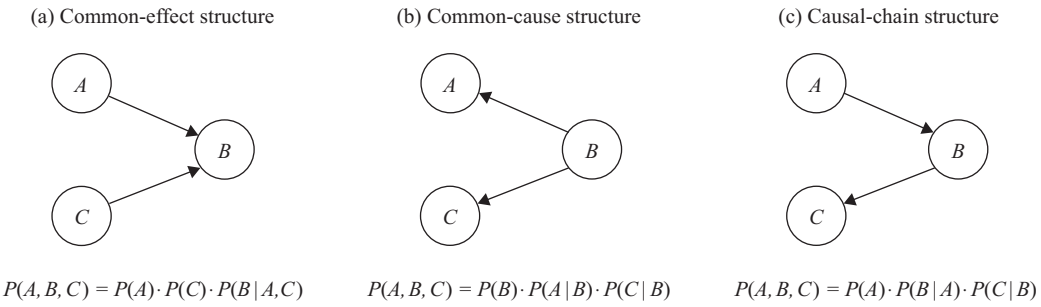


Figure 7.3.2
Basic causal structures comprising three variables A, B, and C. Applying the causal Markov condition to each causal graph entails a different factorization of the joint probability distribution $P(A,B,C)$ over domain variables A, B, and C.

across causal networks. For instance, when reasoning diagnostically in common-effect structures, *explaining away* is an intriguing inference pattern (Morris & Larrick, 1995; Pearl, 1988). Explaining away occurs when conditioning not only on the effect but also on the known presence of an alternative cause. Consider figure 7.3.2a and assume B is a symptom (e.g., fever) with two independent, not mutually exclusive, causes A and C (e.g., a virus and a bacterial infection). With respect to cause A, explaining away corresponds to the inequality $P(a\,|\,b) > P(a\,|\,b,c)$. Knowing that a patient has a fever will raise the probability of her being infected with the virus, but learning additionally that she has a bacterial infection will *lower* the probability of the virus to some extent. In other words, knowing that not only effect B but also the alternative cause C is present will explain away some of the diagnostic evidence provided by B with respect to A (for an overview, see Khemlani & Oppenheimer, 2011).

Another kind of diagnostic reasoning involves inferences from multiple effects (e.g., different symptoms) to an underlying cause (e.g., a disease), constituting a common-cause structure (figure 7.3.2b; Meder & Mayrhofer, 2017a). According to this structure, effects A and C are unconditionally dependent (e.g., lung cancer and yellow teeth correlate because of their common cause, smoking) but independent conditional on their common cause B. This property strongly simplifies diagnostic inferences, because the number of estimates required to parameterize the causal structure is greatly reduced. In particular, the joint likelihood of the effects given the cause can be computed as the product of the individual likelihoods of the effects given the cause. However, research also indicates that people's judgments do not always honor the causal Markov condition, suggesting important pathways for further investigating the rationality of human reasoning (Mayrhofer & Waldmann, 2016; Park & Sloman, 2013; Rottman & Hastie, 2016).

5. Toward a Rational Theory of Diagnostic Reasoning

Models of diagnostic reasoning fundamentally differ with respect to their assumptions, representations, and involved computations. As a consequence, they can make strongly diverging predictions in many situations. What are the implications for constructing a rational theory of diagnostic inference, and what does this mean for the experimental study of human rationality?

On the theoretical level, analyzing diagnostic reasoning from the perspective of inductive causal inference provides new insights into an old problem. Causal Bayes net theories have the expressive power to distinguish between the data level (i.e., covariation information) and parametric causal models that could underlie the observations. A causal modeling framework also enables formalizing diagnostic inferences that do not pertain to the conditional probability of cause given effect, such as judgments of causal responsibility. Some of the accounts that can be implemented within this framework are isomorphic to the classic, purely statistical model in the sense that the inferred diagnostic probability of cause given effect is identical to applying Bayes' rule to the empirical probabilities. Other models make very different predictions about what would constitute a rational solution to the inference problem, for instance, when incorporating uncertainty about alternative causal structures into the judgment process. As a result, inferences that look irrational from the perspective of one model (e.g., a statistical account operating on the data level) would be considered rational from the perspective of another model (e.g., a causal inference account that considers structure uncertainty).

This also raises critical methodological issues for the experimental study of human rationality, which all too often involves comparing human behavior to a single,

supposedly normative, yardstick. Rarely, if ever, will there be a single rational theory for a particular phenomenon (Cohen, 1981; Gigerenzer, 1996; cf. chapter 2.4 by Fiedler, Prager, & McCaughey, this handbook). First, different frameworks exist that can—and should—be used to construct rational theories, including probability theory (Anderson, 1990; chapter 4.5 by Chater & Oaksford, this handbook), logic-based theories (Ragni & Knauff, 2013), ranking theory (Spohn, 2012; chapter 5.3 by Kern-Isberner, Skovgaard-Olsen, & Spohn, this handbook), and theories of bounded and ecological rationality (Chase, Hertwig, & Gigerenzer, 1998; chapter 8.5 by Hertwig & Kozyreva, this handbook). Comparing different types of models can provide insights that could not be gained when restricting the analysis to one particular theoretical viewpoint. Moreover, within a particular methodological framework, different models can be implemented and defended as rational, challenging the common approach of comparing human behavior to a single, supposedly normative, standard (Meder et al., 2014). In the case of elemental diagnostic reasoning, all the models discussed in this chapter rely on some form of probabilistic inference, but their scope, assumptions, and predictions strongly differ. From the perspective of the behavioral sciences, these models should be considered candidate theories, not standards, of human behavior (McKenzie, 2003). In this view, the influence runs both ways: rational theories can inform empirical research, but if there is a stable behavioral pattern that is inconsistent with a particular rational model, one should also revise one's beliefs about the appropriateness of the presumed normative yardstick.

Acknowledgments

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 Anita Todd for editing the manuscript.

Notes

1. Consider sample 1 in figure 7.3.1a (top). The empirical probabilities derived from the frequency data are $P(c) = 0.5$, $P(e \mid c) = 0.3$, and $P(e \mid \neg c) = 0.1$. Accordingly, $w_c = 0.22$ (equation (2)), with $P(c)$ and $P(e \mid \neg c)$ serving as estimates for b_c and w_a, respectively. Plugging these values into equation (3) yields $P(c \mid e) = 0.75$; the same value results when applying Bayes' rule to the empirical probabilities (equation (1)).

2. Consider again sample 1 in figure 7.3.1a (top). Using these data to parameterize structure S_1 with maximum likelihood point estimates yields $b_c = 0.5$, $w_c = 0.22$, and $w_a = 0.1$. Plugging these values into equation (3) yields $P(c \mid e) = 0.75$, whereas the estimate for the probability that C produced E derived from equation (4) yields $P(c \rightarrow e \mid e) = 0.56$.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford, England: Oxford University Press.

Cartwright, N. (1999). Causal diversity and the Markov condition. *Synthese*, *121*, 3–27.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*, 206–214.

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

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.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*, 317–331.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Science*, *7*, 43–47.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 334–384.

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

Holyoak, K. J., 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.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982) *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, *103*, 582–591.

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.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–54.

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

Mayrhofer, R., & Waldmann, M. R. (2016). Sufficiency and necessity assumptions in causal structure induction. *Cognitive Science*, *40*, 2137–2150.

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*, 1273–1312.

McKenzie, C. R. M. (2003). Rational models as theories—not standards—of behavior. *Trends in Cognitive Sciences*, *7*, 403–406.

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., & Mayrhofer, R. (2017a). Diagnostic causal reasoning with verbal information. *Cognitive Psychology*, *96*, 54–84.

Meder, B., & Mayrhofer, R. (2017b). Diagnostic reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 433–458). New York, NY: Oxford University Press.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*, 277–301.

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.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal power. *Psychological Review*, *111*, 455–485.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.

Park, J., & Sloman, S. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. *Cognitive Psychology*, *67*, 186–216.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco, CA: Morgan-Kaufmann.

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

Peterson, C. R., & Beach, L. R. (1967). Man as intuitive statistician. *Psychological Bulletin*, *68*, 29–46.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*, 561–588.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

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.

Spirtes, P., Glymour, C., & Scheines, P. (1993). *Causation, prediction, and search*. New York, NY: Springer.

Spohn, W. (1978). *Grundlagen der Entscheidungstheorie* [Foundations of decision theory]. Kronberg/Taunus, Germany: Scriptor.

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.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford, England: Oxford University Press.

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.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Waldmann, M. R. (Ed.). (2017). *The Oxford handbook of causal reasoning*. New York, NY: Oxford University Press.

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, England: Oxford University Press.

This is a section of doi:10.7551/mitpress/11252.001.0001

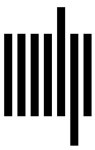
The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

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

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



The MIT Press

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

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

Library of Congress Cataloging-in-Publication Data

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