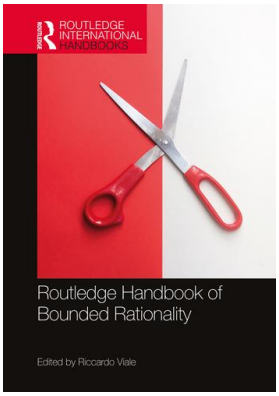

Routledge Handbook of Bounded Rationality

Riccardo Viale

Patterns of defeasible inference in causal diagnostic judgment

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-16
Jean Baratgin, Jean-Louis Stilgenbauer
Published online on: 29 Oct 2020

How to cite :- Jean Baratgin, Jean-Louis Stilgenbauer. 29 Oct 2020, *Patterns of defeasible inference in causal diagnostic judgment from:* Routledge Handbook of Bounded Rationality Routledge
Accessed on: 22 Mar 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-16

14

PATTERNS OF DEFEASIBLE INFERENCE IN CAUSAL DIAGNOSTIC JUDGMENT

Jean Baratgin and Jean-Louis Stilgenbauer

Introduction: what is diagnostic reasoning?

Diagnostic reasoning consists in going back to the causes that triggered one or multiple effects. For example, we reason from effect to cause when we notice the existence of certain symptoms (skin eruptions, sneezing, conjunctivitis, eczema, …) and try to identify the agent that triggered them (pollen allergy, food allergy, mite allergy, others …). The most basic form of diagnostic inference, which will be our main interest in this chapter, consists in estimating the probability of the cause from the knowledge of the effect. Formally this probability is written Pr(cause | effect). When physicians notice that one of their patients has conjunctivitis, they might, for example, try to find out if this symptom is linked to a pollen allergy. For this, they would need to calculate the diagnostic probability associated with the cause of interest by estimating Pr(Pollen-Allergy | Conjunctivitis).

The rational norm of causal diagnostic reasoning has evolved a lot. It went from a purely probabilistic (statistic) norm centered around the Bayes's rule to a norm centered around causal Bayesian networks. However, regardless of the rationality framework considered, people's performances systematically deviate from the predictions of these rationality norms. The *underestimation of the base rate* and the *neglect of alternative causes* are the main violations of rationality traditionally reported by psychologists. These results, which are incidentally very robust, suggest that the estimation of Pr(cause | effect) is calculated in a limited context of rationality, or of degraded rationality, and they also suggest that this estimating process relies on the use of heuristic strategies. In this chapter, we will focus on two verbal strategies of diagnostic reasoning taking the shape of probable conditional inferences (the *defeasible Modus Ponens* and the *defeasible affirming the consequent*). Before detailing these heuristics, we will first re-contextualize the basic diagnostic activity of interest here in regard to associated reasoning. We will then clarify the level on which our work is based in comparison to research in the field of causality.

Estimating Pr(cause | effect) reflects, as we mentioned, an *elementary* form of diagnostic reasoning (Tversky & Kahneman, 1974). It is a basic reasoning triggered when the diagnostic process starts and when we try to account for a new phenomenon (effect). Calculating the diagnostic probability helps people evaluate the intensity of the (statistical) link between the variable considered to be the potential cause (or cause of interest), and the effect to explain. However, this initial step is, insufficient to access the complete characterization of the causal role of the cause of interest and two other types of reasoning will allow the access to a deeper level of

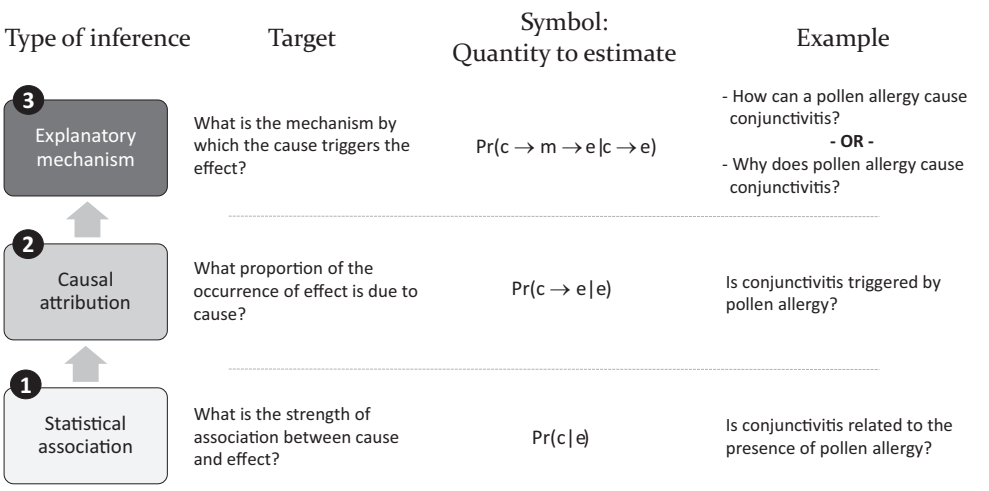

Figure 14.1 The three types of inferences involved in the production of a diagnosis. Reasoning is initiated by the observation of an effect (noted e), step (1) consists in detecting a potential cause (noted c) and in measuring the force of association between e and c. This first step corresponds to the elementary diagnostic reasoning with the estimation of Pr(c|e). Step (2) consists in estimating the contribution of c in the causal influence triggering e. This inference is usually called *causal attribution* and is represented by the expression Pr(c ® e|e). Step (3) is the most complex of all three as it is at this level that an explanatory mechanism (noted m) will be selected\conceived. This mechanism is a process (containing at least one variable) through which the influence of the cause c will be transmitted to the effect e. We can summarize this final step of the diagnostic reasoning with the expression Pr(c ® m ® e | c ® e). This step corresponds to an *inference toward the best explanation*.

understanding (see Figure 14.1). To show the effective causal role of the cause of interest, we need to extract the statistical data by producing an inference called *causal attribution* (Cheng & Novick, 2005; Hilton & Slugoski, 1986). This reasoning consists in estimating the probability of the causal link between the cause of interest and the effect. A more elaborate form of inference will then complete the diagnostic by encouraging the selection\conception of an explanatory mechanism which will explain the mode of action of the cause. This type of reasoning is usually called inference *to the best explanation* (Douven, 2016; Lipton, 2004).

Finally, since this chapter focuses on a form of reasoning relying on causal links, it seems legitimate to put our argument in perspective in the field of research on causality. According to Williamson (2007), three great axes of research on causality exist. The first focuses on the problem of the nature of causality, the second on learning causal links, and the third on causality-based reasoning. Our work on diagnostic inference strategies is a natural extension of research in the third axis, but also in the second one since in the field of diagnosis, learning and reasoning are often quite entangled. The first axis is in contrast to the other two, completely orthogonal to the question asked in this chapter, and as a consequence we will not comment further on contemporary theories on the nature of causality.[1]

Rationality framework: from a statistical norm to causal models

Research led by the heuristics and bias program (Tversky & Kahneman, 1974) yet shows that the estimation of the diagnostic probability of individuals deviates from the diagnostic probability calculated by Bayes's rule.[2] Different tasks like the "Lawyer-Engineer problem"

(Kahneman & Tversky, 1973), the "Taxis problem" (Bar-Hillel, 1980) and the "Mammography problem" (Eddy, 1982), show the systematic neglect of the base rate of the cause of interest when estimating the diagnostic probability (for a review, see Baratgin & Politzer, 2006; Barbey & Sloman, 2007; Koehler, 1996). In the previous example of the allergy to pollens, the base rate was then defined by the prior probability (or prevalence) of this allergy within the reference population.

Yet the pure statistical information is not enough to establish an accurate diagnosis as individuals actually infer this kind of information through a causal model of the situation (or causal structure).[3] Figure 14.1 shows an elementary causal model with only one cause of interest (pollen allergy) and only one effect (conjunctivitis). This kind of object was introduced in the field of cognitive psychology in the 1990s (Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995), it is a compact representation of the cause-effect relationships connecting the variables of interest to each other. Causal models have later been developed and updated to provide heuristics models to psychology allowing the testing of new hypotheses, and, on the other hand, these models provide a new framework of rational analysis of human causal inferences (Griffiths & Tenenbaum, 2005, 2009).

A causal model contains information that goes beyond the simple pattern of co-occurrences of the variables of interest, meaning the statistical data of the studied system. Individuals possessing a causal model will be able to understand the process that generated this data. This higher level of understanding implies attributing a new state to the variables of the system: each variable will be interpreted as either the *cause* or the *effect*. One other refinement consists in enriching a causal model through the introduction of mechanisms which will help accurately define the relationships between variables. Figure 14.2 represents an elementary causal model

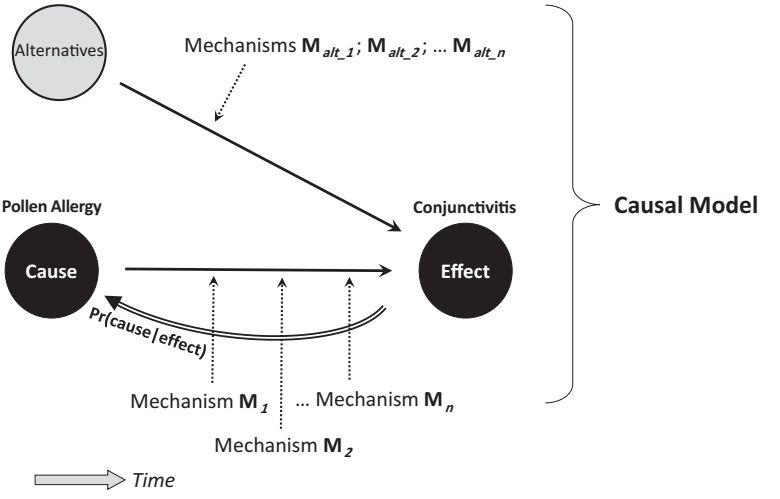

Figure 14.2 Causal model containing a single cause of interest (pollen allergy) causally linked to a single effect (conjunctivitis). The model includes a supplementary variable representing all the alternative causes potentially also triggering the effect. The elementary diagnostic reasoning shown here allows going back to the cause from the knowledge of the effect. This process consists in the estimation of the probability Pr(cause|effect) and is symbolized by the curved double arrow pointing to the cause. The causal model can also be improved through the introduction of a mechanism explaining why the cause triggers the effect (see Protzko, 2018 for a recent discussion on the concepts of cause and mechanism). In general, multiple mechanisms M_1, M_2, … M_n (dashed arrows) compete to characterize the causal link of interest. Alternative causes also possess their own explanatory mechanisms M_{alt_1}, M_{alt_2}, … M_{alt_n}.

based on a structure containing a single cause and a single effect. The arrow connecting those two variables implies a link of causality. The model also contains a supplementary variable representing all the different alternative causes able to trigger the same effect regrouped into one entity.

Today, representing causal knowledge of people through the use of *causal Bayes nets* (CBN) is quite common. The CBN are formal tools that were initially developed in the field of Artificial Intelligence (Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 2000). This formalism is commonly used to study causal cognition in psychology. For a detailed review of its uses in this field, see Rottman and Hastie (2014) and Rottman (2017). The great modularity of the CBN offers the possibility to represent causal structures like the one in Figure 14.2 and also allow the expression of the great variety of reasoning processes related to it. For example, Meder, Mayrhofer, and Waldmann (2009, 2014) recently suggested a model of diagnostic reasoning (structure induction model) establishing a new rational framework for elementary inferences produced from a causal structure similar to the one shown in Figure 14.2.[4] Despite being useful, not only to establish a standard of rationality but also for their heuristic qualities, we believe that this model remains insufficient in order to explain human diagnostic inferences. People have indeed limited cognitive resources (memory, attention, computational power) and of course, cannot process the complex computations required by CBN. In order to avoid this obstacle, people must build strategies to estimate the diagnostic probability. In the following section of this chapter, we will briefly describe some of these well-known strategies linked to a misuse of the Bayes's rule and will then detail a new type of strategies relying on conditional reasoning schemas.

Strategies related to the sub-optimal use of the Bayes's rule in the estimation of the diagnostic probability

The heuristics and bias program of the 1970s revealed that people's estimations of the diagnostic probability deviate from the Bayesian model. Yet, those first researchers did not take into account the influence of causal knowledge in the probabilistic reasoning. Krinski and Tenenbaum (2007) have shown that estimations of the diagnostic probability in the mammography problem were close to the Bayesian norm when the statistical information (pattern of co-occurrence of the input variables) was associated with an explicit causal model of the situation, see also Fernbach and Rehder (2012) for results of the same nature. Recent work nuances this observation (Hayes, Ngo, Hawkins, & Newell, 2018) for despite the introduction of the causal model, some people keep using non-Bayesian estimation strategies. Some participants indeed estimate the diagnostic probability Pr(cause | effect) from the rate of false positives, by taking into account only the probability Pr(effect | Øcause). Other non-Bayesian estimation strategies have been described by Cohen and Staub (2015), such as those consisting in calculating a pondered sum of the probabilities mentioned in the Bayes's rule. Some participants try to calculate the diagnostic probability from the addition of the false positives rate Pr(effect | Øcause) to the likelihood (rate of true positives) Pr(effect | cause).

The various strategies used by people in order to estimate the diagnostic probability suggest that the underlying cognitive processes can vary from one person to another, added to the fact that nothing prevents the observation of variations within the persons themselves depending on the time or context. The systematic study of estimation strategies appears to be crucial in order to improve the understanding of human diagnostic inferences. A limiting aspect of previous studies is the nature of the strategies identified. Indeed, these studies have specifically focused on strategies consisting in combining (in a sub-optimal fashion) quantities of the Bayes's rule (in

particular the rate of false positives and true positives). Yet, other strategies of diagnostic estimation can be considered like those formed from schemas of conditional inferences.

New diagnostic estimation strategies: *defeasible Modus Ponens* and *defeasible Affirming the Consequent*

In general, diagnostic probability estimation strategies are distinguished from rational models, either purely statistical like the Bayes's rule or causal like CBN framework. If we use Marr's terminology (1982), these models are at the *computational* level of the general cognitive architecture. They are functional models that define a standard of rationality to calculate the diagnostic probability. Strategies used by people are instead at the *algorithmic* level. They represent the effective cognitive processes followed by people to estimate this probability. This level being purely descriptive, we believe the influence of language and of mechanisms related to linguistic communication in the estimation process cannot easily be ignored. It is indeed natural for people to be communicating and to be using their causal beliefs in reasoning, forming propositions and organizing them in the shape of arguments. According to Mercier and Sperber (2011), it is even the essential function of reasoning. An entire literature on causal arguments exists establishing a strong link between causal cognition and argumentation (see Hahn, Bluhm, & Zenker, 2017 for a recent review). Among the causal arguments, conditional arguments are central to the human cognition. Two patterns have been of special interest to us here because of their high psychological plausibility. The first is based on the *affirming the consequent* schema (AC), the second based on the *Modus Ponens* (MP). These two schemas will be considered in particular in defeasible forms, as a consequence, we will be using the terms *defeasible affirming the consequent* (DAC) and *defeasible Modus Ponens* (DMP). These reasonings are represented in Figure 14.3.

Initially, AC and MP were studied in their causal form by Cummins, Lubart, Alksnis, and Rist (1991), Cummins (1995) and Politzer and Bonnefon (2006). In these research studies, AC was built on a conditional in which the antecedent and the consequent respectively corresponded to a *cause* and to an *effect*. The first premise of the argument consists in declaring the *effect* (we will use here the term "effect!" to mark the initial stage of reasoning that begins with the arrival of new information), the second premise corresponds to the conditional statement and the conclusion is formed by the *cause*. AC coincides with a form of reasoning called *abduction* by the philosopher C. S. Peirce (for recent reviews, see Douven, 2011; Park, 2017). Cummins (1995) and Politzer and Bonnefon (2006) also studied a specific form of MP built on an inverted conditional in which the antecedent was an *effect* and the consequent was the *cause*. The first premise and the conclusion of this form of MP remained untouched compared to AC. They were respectively instantiated by the *effect* on one hand and the *cause* on the other. We represent below the structure of AC and of MP such as they appear in the aforementioned studies:

AC:	MP:
Effect!	Effect!
If <cause> then <effect>	If <effect> then <cause>
Cause	Cause

Those two arguments constitute, in our point of view, good candidates in order to define diagnostic probability estimation strategies, as these forms of reasoning both consist in going back to the cause (conclusion) from the observation of the effect (first premise). The conditional (the second premise) defines the type of strategy. For AC, the inference is based on the trust

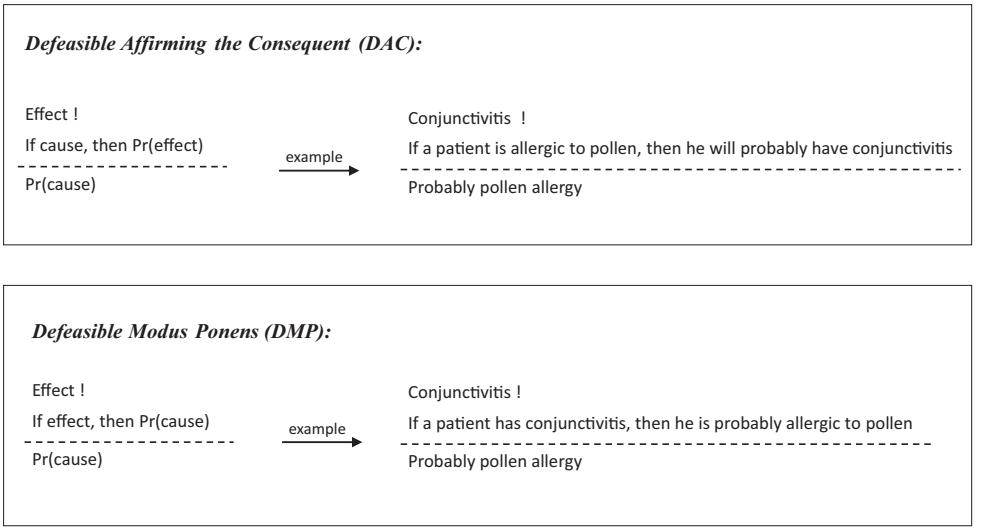

Figure 14.3 These two reasoning patterns represent two formally different strategies for estimating the diagnostic probability Pr(cause|effect). They correspond respectively in the A.I. and philosophical literature to two modes of inference: abduction for DAC and deduction defeasible for DMP. First, the defeasible deduction is a weakened form of deduction that permits the production of provisionally true and/or probable conclusions. Second, abduction is a heuristic reasoning that serves to identify explanatory causes or hypotheses.

put in the causality link between the antecedent and the consequent. This trust depends on the existence of alternative causes that are, just like the cause of interest, capable of triggering the effect. For MP, the mechanism which the inference relies on is, in our point of view, different, for it is rather based on the explanatory qualities of the cause of interest. Here, it's the strength of the mechanism by which the effect is produced by the cause of interest that will guide the inference.

Despite AC and MP being *a priori* interesting in order to define diagnostic reasoning strategies, those arguments are not, as they are, completely satisfying. Indeed, as we've explained above, diagnostic reasoning is an uncertain reasoning allowing the estimation of the probability of the cause of interest, given the knowledge of the effect, or formally the estimation of the probability Pr(cause|effect). AC and MP schemas must be generalized and considered in a probable form as is shown in Figure 14.3. As is shown in Figure 14.3, the probabilistic aspect is defined by the degree of probability placed, on one hand, on the consequent of the conditionals, and on the other, on the reasoning conclusions. A good diagnostic strategy should also possess the quality of being defeasible, for it is natural for people to draw a conclusion or to update its probability in the case of the arrival of new information (taking an additional premise into account for example). The defeasible aspect is shown through the dashed inference line in Figure 14.3.

In order to clearly differentiate the standard schemas AC and MP from their probable and defeasible generalization, we will from now on be talking respectively about DAC and DMP. Those two schemas, as causal diagnostic reasoning strategies, have recently been studied by Stilgenbauer and Baratgin (2018, 2019) who have shown their psychological relevance in estimating Pr(cause|effect). The authors submitted participants to an experimental paradigm of rule production, where people had to rebuild the conditional rule of DAC and DMP schemas.

Participants were only able to use two pieces of information. They received the first premise of the schema (which was certain) corresponding to the phase of the declaration of the effect. This step is symbolized in Figure 14.3 by the term "effect!." Participants then were informed of the (probable) conclusion of the reasoning, represented in Figure 14.3 by the term "Pr(cause)." The task consisted in the production of a conditional rule allowing the inference of the conclusion from the first premise. To answer, participants had at their disposal jumbled words from which they could form the rule of their choice: [1] *if cause then Pr(effect)* or the reversed rule [2] *if effect then Pr(cause)*. The kind of rule produced revealed the diagnostic estimation strategy the participants preferred. The production of a rule similar to [1] signaled a preference for the DAC estimation strategy. Similarly, the production of a rule similar to [2] revealed a preference for the DMP strategy.

The results of Stilgenbauer and Baratgin (2018, 2019) show that the participants preferred DMP, as they built rules similar to [2] to infer the conclusion "Pr(cause)" from the premise "effect!." The use of a rule like *if effect then Pr(cause)* is indeed quite natural in this situation since a direct correspondence exists between that conditional and the diagnostic probability Pr(cause | effect).[5] This is not the case with DAC inferences though, which rely on rules similar to [1] and which instead make the predictive probability Pr(effect | cause) stand out, since in this case the antecedent of the rule is made of the *cause*, and the consequent is made of the *effect*.

The results of Stilgenbauer and Baratgin (2018) also show that the participants' preferred strategy was not always DMP. Indeed, in certain situations, the DAC strategy can strongly compete with the DMP strategy. This change of preference is linked to the perceived value of the predictive probability Pr(effect | cause).[6] This result has been obtained while controlling the relative levels of the diagnostic and predictive probabilities. In a first *probabilistic context*, participants received information indicating that the diagnostic probability was higher than the predictive probability: Pr(cause | effect) > Pr(effect | cause). In a second condition, it was the opposite: the information given indicated that the diagnostic probability was lower than the predictive probability Pr(cause | effect) < Pr(effect | cause). The results show that conditional rules built by the participants depended on the probabilistic context (meaning the relative levels of the diagnostic and predictive probabilities). People, in the main, build rules similar to [2] *if effect then Pr(cause)* (DMP strategy) in the context Pr(cause | effect) > Pr(effect | cause). Yet, in the context of Pr(cause | effect) < Pr(effect | cause), the proportion of rules similar to [1] *if cause then Pr(effect)* (DAC strategy) significantly increases and does not differ from the proportion of rules similar to [2]. These results confirm the idea that people can estimate the diagnostic probability following strategies taking the form of defeasible schemas of inferences. In the following section we will report new experimental results confirming the robustness of the data we just exposed.

New experimental data: a test of diagnostic strategies through a rule evaluation paradigm

We propose here a new experiment expanding on the logic of Stilgenbauer and Baratgin (2018). The task no longer consists in *producing* a conditional rule, but to *evaluate* (and compare) the two rules that DAC and DMP are made of. The experiment procedure was the following: participants begin by receiving the first premise of the diagnostic strategy "effect!" as well as the conclusion "Pr(cause);" we then explicitly and simultaneously show them the conditional rules [1] *If cause then Pr(effect)* and [2] *If effect then Pr(cause)*. The task of the participants consists in judging if those rules are *judicious* to infer the conclusion "Pr(cause)" from the premise "effect!." Each rule is evaluated on a seven points scale (1 = not very judicious; 7 = very judicious). We assume that the rule judged to be the most judicious will indicate the preferred

diagnostic estimation strategy. If the rule [1] is judged to be more judicious than rule [2], this will show that DAC is the preferred strategy to estimate the diagnostic probability. But if [2] is judged more judicious than [1], this will show a preference for DMP. In this experiment, we also recreated two probabilistic contexts identical to those in the study of Stilgenbauer and Baratgin (2018). The judgment of the participants was recorded in the context of Pr(cause | effect) > Pr(effect | cause) and also in the context Pr(cause | effect) < Pr(effect | cause).

A scenario of *fault detection* has been used to cover the situation of diagnostic reasoning. This scenario involved the operation of an industrial machine producing auto parts. The probabilistic context was introduced in a between-subject design, using pictures as shown in Figure 14.4. No matter the probabilistic context, a machine composed of multiple components was introduced to the participants. Components of this machine can malfunction, occasioning the production of faulty parts. The *components* of the machine represent the *causes* and the *defects* on the parts represent the *effects*. In Figure 14.4, the picture on the left refers to the context Pr(cause | effect) > Pr(effect | cause) and the picture on the right to the context Pr(cause | effect) < Pr(effect | cause). For example, for the picture on the left, the machine has two components that can malfunction resulting in parts produced with four defects. The malfunction of components X_1 or X_2 can cause the defects X_a, X_b, X_c, or X_d. For the two probabilistic contexts, it is impossible to have multiple components malfunctioning at the same time, neither is it possible to have multiple defects at the same time. For the left picture, the diagnostic and predictive probabilities are respectively equal to Pr(cause | effect) = ½ and Pr(effect | cause) = ¼.

Participants of the experiment were then told of the malfunction of the machine, and we used the conversation between the repairmen attempting to solve the issue to introduce simultaneously the conditional rules [1] and [2]. Before taking apart the machine, the repairmen

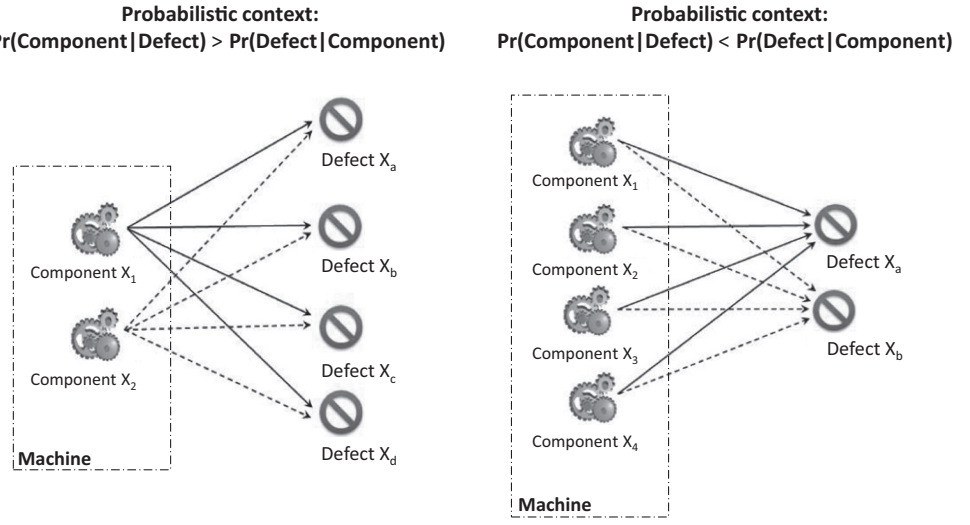

Figure 14.4 Experimental set-up introducing the situation of causal diagnostic reasoning. In each probabilistic context, an industrial machine is symbolized by dashed rectangles. The machines are made up of a certain number of components and their malfunction (causes) can trigger the production of faulty parts with defects (effects). The figure on the left defines the probabilistic context Pr(cause | effect) > Pr(effect | cause). For example, if we're interested in the defect X_a (*effect*) and to the component X_1 (*cause*), we have $Pr(X_1 | X_a) = 1/2$ and $Pr(X_a | X_1) = 1/4$. The figure on the right defines the opposite probabilistic context Pr(cause | effect) < Pr(effect | cause) with $Pr(X_1 | X_a) = 1/4$ and $Pr(X_a | X_1) = 1/2$.

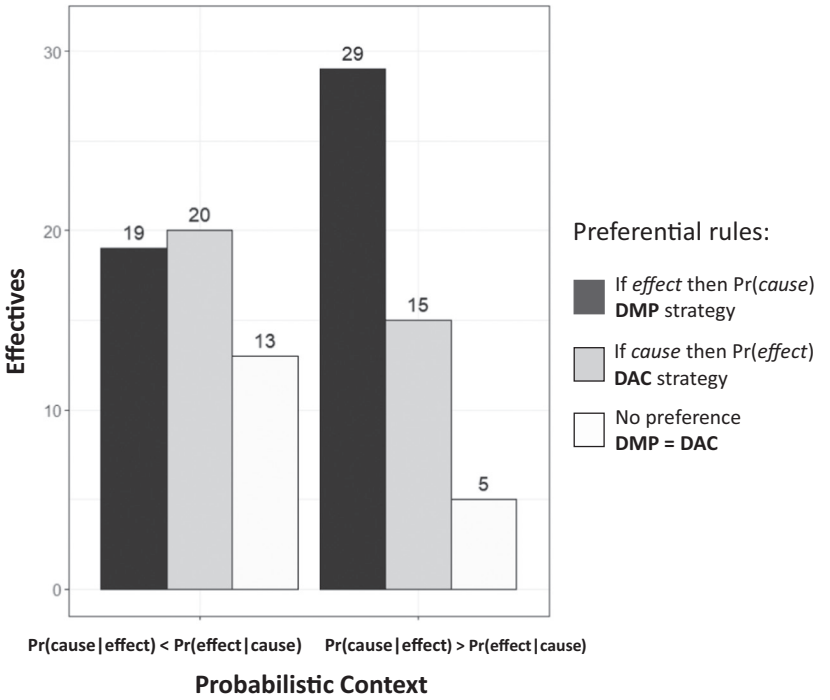

Figure 14.5 Experimental results obtained in rule evaluating. Distribution of the preferred rules of participants depending on the probabilistic context defined by the relative levels of the diagnostic and predictive probabilities.

thought out loud and one of them suggested rule [1] *If component* X_1 *then Pr(defect* X_a*)*. Another repairman then contradicted him and suggested the reversed rule [2] *If defect* X_a *then Pr(component* X_1*)*.

We recruited 101 participants for the experiment and the results are shown in Figure 14.5. We categorized participants depending on the grades they attributed to each rule. When the grade of the rule [1] was higher than the grade of the rule [2], participants were categorized in DAC. They were categorized in DMP if they graded [2] higher than [1]. In the case where participants graded the two rules the same, they were categorized in "No preference."

Calculating a χ^2 revealed a significant link between the preferred rule and the probabilistic context. In the situation Pr(cause|effect) > Pr(effect|cause), participants believe rules similar to [2] to be more judicious than rules similar to [1]. Yet, in the situation of Pr(cause|effect) < Pr(effect|cause), there was no particular preference visible between the two kinds of rules: 20 participants believed rule [1] to be more judicious than rule [2] and 19 participants believed the opposite. In the context Pr(cause|effect) > Pr(effect|cause), there are also 5 participants grading both rules as equally judicious, and 13 in the context Pr(cause|effect) < Pr(effect|cause).

These results obtained with a paradigm of rule evaluation are interesting for they corroborate those of Stilgenbauer and Baratgin (2018) who used a paradigm of rule production. This new data seems to validate the idea that it is natural for people to use defeasible strategies like DMP and DAC in order to estimate the diagnostic probability. This experiment also confirms that these two strategies are not equally preferred by individuals. Our data indicate that DMP

seems to constitute the strategy by default, but in situations where the perceived value of the predictive probability increases, DAC strongly competes with this strategy.

Conclusion: the diversity of diagnostic reasoning shapes

In this chapter, we have been interested in an elementary form of diagnostic reasoning consisting in estimating the probability of a *cause* from the knowledge of its *effect*: Pr(cause | effect). We have limited ourselves to the study of inferences produced from the causal structure known *a priori* and given to participants. This structure was composed of a single cause of interest and a single effect (see Figure 14.2), we will refer to it by using the expression *cause* → *effect*.[7] In the literature, the majority of studies consist in the evaluation of the performance of participants compared to the rational norm that today is made up of *causal Bayesian networks*. The aim of this work was different. It intended to remain at the level of psychological processes (in other words, processes that are at the *algorithmic* level according to the terminology of D. Marr) to test the plausibility of two strategies of diagnostic inferences. The first was based on a *defeasible affirming the consequent* and the second was based on a *defeasible Modus Ponens*. Our results show that participants understand and can follow these patterns of inference to estimate the probability of a *cause* from the knowledge of its *effect*. These two strategies are important for they are natural for people, although it does not mean that they constitute the only way possible when trying to estimate Pr(cause | effect). Other ways can be imagined, for example, we can think about the use of strategies based on negated components in the shape of a *defeasible Modus Tollens* (effect!, If ¬ cause, then Pr(¬ effect) ⇒ Pr(cause))[8] or of a *defeasible denying the antecedent* (effect!, If ¬ effect, then Pr(¬ cause) ⇒ Pr(cause)). It belongs to future research to test the psychological reality of these kinds of strategies and to report their potential usage by people.

In this chapter, we studied the estimation strategies of Pr(cause | effect) for basic diagnostic inferences produced from a *cause* → *effect* structure known *a priori*. Yet, we believe that two other kinds of diagnostic reasoning seem to be of importance and, to our knowledge, no study exists on their estimation strategies. Figure 14.1 at the beginning of this chapter suggests an articulation between the basic diagnostic reasoning and these two complementary forms of inferences. The first is called *causal attribution* (Cheng & Novick, 2005; Hilton & Slugoski, 1986) and intervenes in the context where the causal structure is unknown. In this situation, the diagnosis relies on the *link* between the cause and the effect, and the diagnostic probability of interest will be written Pr(cause → effect | effect). The second kind of diagnostic inference that seems to us to be of importance is a bit more complex. It consists in inferring the *mechanism* explaining the nature of the causal link. A minimalistic mechanism can be reduced to a single *mediator* (in Figure 14.2, the concept of mechanism is represented by dashed arrows without actually being clarified). A mediator is a variable allowing the transmission of the influence of the cause to the effect. For example, we know today that *smoking* (cause) can trigger *lung cancer* (effect), through the *tar* (mediator) accumulating in lung alveoli (the role of tar hasn't always been clear in the past). In the context where the structure *cause* → *effect* is known *a priori*, we can, for example, ask ourselves if a mediator transmitting the influence of the cause to the effect exists. In this situation, the probability Pr(cause → mediator → effect | cause → effect) constitutes the diagnostic probability of interest. In statistics, a very interesting literature exists on the topic of mediation. For a reference, we can read VanderWeele (2015). Many examples of mediation can also be found in MacKinnon (2008).

Finally, we will note that the causal attribution Pr(cause→ effect | effect) consists in the elaboration of a causal structure by definition unknown and that the inference of a mediator Pr(cause → mediator → effect | cause → effect) consists in the modification of the causal

structure known *a priori*. In the literature dedicated to causal cognition, these processes are generally understood as learning processes allowing on the one hand the elaboration of the structure (defining the variables as *cause* and *effect*), and on the other, the estimation of its parameters (defining the probability distributions of the structure's variables). There is a third class of processes traditionally distinguished in the literature regrouping all the causal reasoning that can be done from a completely defined and parameterized structure (we can refer to Hastie, 2015; Rottman, 2017; and Rottman & Hastie, 2014 for a general review). Yet, we believe that the strategies used by people in order to estimate the diagnostic probability cannot be classified in the traditional way in effect in the field of causal cognition. The main reason for this is that in the field of diagnosis (at least) learning a causal link is often concomitant to the diagnostic reasoning itself. This is also what Meder et al. (2014) suggest with the model of *structure induction* combining learning of the causal structure with the estimation of Pr(cause | effect) (see also Meder et al., 2009). Having said that, we believe that the result or the product of the process of diagnostic inference (the potential cause, the causal link or the mediator) constitutes the central element from which a typology of diagnostic reasoning strategies can be elaborated. A future research program dedicated to the use of diagnostic estimation strategies by people will come, we hope, to bring answers to this important question.

Acknowledgments

The authors would like to thank Frank Jamet and Baptiste Jacquet for much discussion and other help in their research.

Notes

1 For readers interested in the different conceptions of causality, we recommend reading Beebee, Hitchcock, and Menzies (2009), especially Parts 2 and 3.

2 The Bayes's rule is a rule of belief updating when new convincing information is learned. It is the only possible rule of revision of beliefs in the situation of revision called focusing (the situation where the message concerns an object drawn at random from a population of objects which constitutes a certain stable universe, see Baratgin & Politzer, 2010; Walliser & Zwim, 2011). In the context of causal diagnostic, if we write c = *cause* and e = *effet*, the rule would be written in the following way: $\Pr(c|e) = \frac{Pr(e|c) \times Pr(c)}{Pr(e|c) \times Pr(c) + Pr(e|\neg c) \times Pr(\neg c)}$. The term $\Pr(c|e)$ corresponds to the diagnostic probability, which is the probability of the cause when the effect is present. The quantity $\Pr(e|c)$ is generally called *likelihood*, yet in the situation of causal diagnostic, we would rather use the rate of *true positives*, which is the probability of observing the effect when the cause is present. The quantity $\Pr(c)$ is the *prior* probability of the cause. From an objective point of view, it corresponds to the *base rate* (or prevalence) of the cause within the reference population. The term $\Pr(e|\neg c)$ refers to the influence of alternative causes, in other words, the probability of observing the effect without observing the cause of interest. It can also be called the rate of *false positives*. Finally, the quantity $\Pr(\neg c)$ corresponds to the *base rate* of all of the alternative causes.

3 From the historical point of view, from the very beginning of the rise of modern statistics during the nineteenth century, has started to emerge a radical separation between this discipline, on the one hand and causal concepts, on the other (such as those of strength, of mechanisms or of a causal model). Chapter 2 of Pearl and Mackenzie (2018) tells the history of this divorce which, beyond statistics, has produced considerable negative effects for all data-driven sciences. Unmistakably, this separation also had consequences for research on causal cognition. The excessive focus of the *heuristics and bias* school on statistical concepts to the detriment of the introduction of causal concepts likely stems from this divorce. Fortunately, we see today the return of causality. The "causal revolution," according to Pearl is on the way and has started planting its seeds into all sciences and in particular in cognitive psychology.

Waldmann and Hagmayer (2013) wrote an excellent review describing the evolution of causal cognition theories and in particular the shift from purely statistical models to models integrating causal concepts.

4 Meder and Mayrhofer (2017) have recently shown an interest in the more sophisticated forms of diagnostic reasoning that might be produced from causal structures containing multiple causes and effects.

5 Today, many results demonstrating that people interpret the probability of an indicative conditional such as Pr(if A, then C) as the conditional probability Pr(C|A) are available, see for example Baratgin and Politzer (2016). Yet in this work, we study strategies relying on conditionals in the shape of "if A then Pr(C)" and this type of rule constitutes the natural interpretation of Pr(if A, then C), see for this specific point Over, Douven, and Verbrugge (2013).

6 Predictive probability plays an important part in the estimation of the diagnostic probability. It influences, in particular, the accuracy of the estimations of the diagnostic probability. For example, when we ask participants to estimate from a set of data the value of the diagnostic probability, they will weight their estimations with the perceived value of the predictive probability. The lower the predictive probability, the more under-estimated the diagnostic probability is, compared to its real value contained in the data (Meder et al., 2009, 2014; see also Meder & Mayrhofer, 2017; Stilgenbauer & Baratgin, 2019; Stilgenbauer, Baratgin, & Douven, 2017).

7 Yet, in the experiment described in the previous section, we have defined different values of Pr(cause|effect) by changing the number of alternative causes (see Figure 14.4).

8 The symbol "$\Longrightarrow$" represents a defeasible consequence relationship.

References

Baratgin, J., & Politzer, G. (2006). Is the mind Bayesian? The case for agnosticism. *Mind and Society*, 5(1), 1–38. http://doi.org/10.1007/s11299–006–0007–1.

Baratgin, J., & Politzer, G. (2010). Updating: A psychologically basic situation of probability revision. *Thinking & Reasoning*, 16(4), 253–287. http://doi.org/10.1080/13546783.2010.519564.

Baratgin, J., & Politzer, G. (2016). Logic, probability and inference: A methodology for a new paradigm. In L. Macchi, M. Bagassi, & R. Viale (Eds.), *Cognitive unconscious and human rationality* (pp. 119–142). Cambridge, MA: MIT Press.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30(3), 241–254. http://doi.org/10.1017/S0140525X07001653.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211–233. http://doi.org/10.1016/0001–6918(80)90046–3.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421–441.

Beebee, H., Hitchcock, C., & Menzies, P. (Eds.) (2009). *The Oxford handbook of causation*. Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199279739.001.0001.

Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review*, 112(3), 694–707. http://doi.org/10.1037/0033–295X.112.3.694.

Cohen, A. L., & Staub, A. (2015). Within-subject consistency and between-subject variability in Bayesian reasoning strategies. *Cognitive Psychology*, 81, 26–47. http://doi.org/10.1016/j.cogpsych.2015.08.001.

Cummins, D. (1995) Naïve theories and causal cognition. Memory and Cognition, 23(5), 646–659.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, 19(3), 274–282.

Douven, I. (2011). Abduction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2011 ed.). Available at: https://stanford.library.sydney.edu.au/entries/abduction/

Douven, I. (2016). Inference to the best explanation: What is it? And why should we care? In T. Poston & K. McCain (Eds.), *Best explanations: New essays on inference to the best explanation*. Oxford: Oxford University Press.

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). Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511809477.019.

Fernbach, P. M., & Rehder, B. (2012). Cognitive shortcuts in causal inference. *Argument & Computation*, 4(1), 64–88. http://doi.org/10.1080/19462166.2012.682655.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51(4), 334–384.. http://doi.org/10.1016/j.cogpsych.2005.05.004.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661–716. http://doi.org/10.1037/a0017201.

Hahn, U., Bluhm, R., & Zenker, F. (2017). Causal argument. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*. Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199399550.013.26.

Hastie, R. (2015). Causal thinking in judgments. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 590–628). New York: Blackwell. http://doi.org/10.1002/9781118468333.ch21.

Hayes, B. K., Ngo, J., Hawkins, G. E., & Newell, B. R. (2018). Causal explanation improves judgment under uncertainty, but rarely in a Bayesian way. *Memory and Cognition*, 46(1), 112–131. http://doi.org/10.3758/s13421–017–0750-z.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution. The abnormal conditions focus model. *Psychological Review*, 93(1), 75–88. http://doi.org/10.1037/0033–295X.93.1.75.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. http://doi.org/10.1037/h0034747.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19(1), 1–17. http://doi.org/10.1017/S0140525X00041157.

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology General*, 136(3), 430–450. http://doi.org/10.1037/0096–3445.136.3.430.

Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.

Machamer, P., Darden, L., & Carver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25. http://doi.org/10.1086/392759.

MacKinnon, D. (2008). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Meder, B., & Mayrhofer, R. (2017). Diagnostic reasoning. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 433–458). Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199399550.013.23.

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(3), 277–301. http://doi.org/10.1037/a0035944.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–111. http://doi.org/10.1017/S0140525X10000968.

Over, D., Douven, I., & Verbrugge, S. (2013). Scope ambiguities and conditionals. *Thinking & Reasoning*, 19(3), 284–307. http://doi.org/10.1080/13546783.2013.810172.

Park, W. (2017). *Abduction in context: The conjectural dynamics of scientific reasoning*. Berlin: Springer. http://doi.org/10.1007/978–3-319–48956–8.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Kaufmann. http://doi.org/10.1016/0004–3702(91)90084-W.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press. http://doi.org/10.1017/CBO9780511803161.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.

Politzer, G., & Bonnefon, J.-F. (2006). Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind and Language*, 21(4), 484–503.

Protzko, J. (2018). Disentangling mechanisms from causes: And the effects on science. *Foundations of Science*, 23(1), 37–50. http://doi.org/10.1007/s10699–016–9511-x.

Rottman, B. M. (2017). The acquisition and use of causal structure knowledge. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 1–55). Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199399550.013.10.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, 140, 109–139. http://doi.org/doi.org/10.1037/a0031903.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. New-York: Springer-Verlag.

Stilgenbauer, J.-L., & Baratgin, J. (2018). Étude des stratégies de raisonnement causal dans l'estimation de la probabilité diagnostique à travers un paradigme expérimental de production de règle. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 72(1), 58–70. http://doi.org/doi.org/10.1037/cep0000108.

Stilgenbauer, J.-L., & Baratgin, J. (2019). Assessing the accuracy of diagnostic probability estimation: Evidence for *defeasible modus ponens*, *International Journal of Approximate Reasoning*, 105, 229–240. https://doi.org/10.1016/j.ijar.2018.11.015.

Stilgenbauer, J.-L., Baratgin, J., & Douven, I. (2017). Reasoning strategies for diagnostic probability estimates in causal contexts: Preference for defeasible deduction over abduction. In *Proceedings of the 4th International Workshop on Defeasible and Ampliative Reasoning (DARe-17) Co-located with the 14th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2017)* (pp. 29–43). Espoo, Finland. Available at: http://ceur-ws.org/Vol-1872/.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Van der Weele, T. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York: Oxford University Press.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121(2), 222–236. Available at: www.ncbi.nlm.nih.gov/pubmed/1534834.

Walliser, B., & Zwirn, D. (2011). Change rules for hierarchical beliefs. *International Journal of Approximate Reasoning*, 52(2), 166–183. http://doi.org/10.1016/j.ijar.2009.11.005.

Williamson, J. (2007). Causality. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (vol. 14, pp. 95–126). Dordrecht: Springer.