Reasoning strategies for diagnostic probability estimates in causal contexts: Preference for defeasible deduction over abduction*

Jean-Louis Stilgenbauer[1,2], Jean Baratgin[1,3], and Igor Douven[4]

[1] CHArt (P-A-R-I-S), Université Paris 8, EPHE, 4-14 rue Ferrus, 75014 Paris – France, http://paris-reasoning.eu/
[2] Facultés Libres de Philosophie et de Psychologie (IPC), 70 avenue Denfert-Rochereau, 75014 Paris – France
jlstilgenbauer@ipc-paris.fr
[3] Institut Jean Nicod (IJN), École Normale Supérieure (ENS), 29 rue d'Ulm, 75005 Paris – France
jean.baratgin@univ-paris8.fr
[4] Sciences, Normes, Décision (SND), CNRS/Université Paris-Sorbonne, 1 rue Victor Cousin, 75005 Paris – France
igor.douven@paris-sorbonne.fr

Abstract. Recently, Meder, Mayrhofer, and Waldmann [1,2] have proposed a model of causal diagnostic reasoning that predicts an interference of the predictive probability, Pr(Effect | Cause), in estimating the diagnostic probability, Pr(Cause | Effect), specifically, that the interference leads to an underestimation bias of the diagnostic probability. The objective of the experiment reported in the present paper was twofold. A first aim was to test the existence of the underestimation bias in individuals. Our results indicate the presence of an underestimation of the diagnostic probability that depends on the value of the predictive probability. Secondly, we investigated whether this bias was related to the type of estimation strategy followed by participants. We tested two main strategies: abductive inference and defeasible deduction. Our results reveal that the underestimation of diagnostic probability is more pronounced under abductive inference than under defeasible deduction. Our data also suggest that defeasible deduction is for individuals the most natural reasoning strategy to estimate Pr(Cause | Effect).

Keywords: diagnostic inference, defeasible deduction, abduction, causal Bayes nets

1 Diagnostic Reasoning

Causal inferences can go in two opposite directions. In so-called diagnostic reasoning, they go from *Effect* to *Cause*. Specifically, in this type of reasoning, one

* All supplementary information as well as all materials, data and scripts for the statistical analyses can be downloaded from: https://osf.io/7yc92/?view_only=57085866d97d48a4bbd75938d3b4a6af.

estimates the probability of *Cause* given *Effect*, $\Pr(c \mid e)$. In so-called predictive reasoning, causal inferences go from *Cause* to *Effect*; here the aim is to estimate the probability of *Effect* given *Cause*, $\Pr(e \mid c)$. The present paper focuses on the former type of reasoning. Diagnostic reasoning is emblematic not only in the field of medicine, but also in our everyday lives. For example, we reason from effects to causes when we try to understand why our car refuses to start or why we failed the final year exam. Here, we focus on the most elementary type of diagnostic inference, which involves a single cause–effect relation between two binary events (i.e., events that either occur or do not occur).

In the experiment to be reported, we show that individuals' reasoning corroborates the predictions of the *Structure Induction Model* (SIM) recently proposed by Meder, Mayrhofer, and Waldmann [1,2]. Whereas it is usually assumed that diagnostic judgments should merely be a function of the empirical conditional probability $\Pr(c \mid e)$, the SIM predicts that diagnostic inferences are also systematically affected by the empirical predictive probability $\Pr(e \mid c)$.

We generalize this result by showing that the influence of the predictive probability in the estimation of the diagnostic probability is effective whatever the reasoning strategy followed by the participants. Two strategies have particularly caught our attention: the estimate of $\Pr(c \mid e)$ by *abduction*, on the one hand, and the estimate through *defeasible deduction*, on the other hand (see Section 3).

2 Estimate of $\Pr(c \mid e)$ via Causal Bayes Nets

Causal Bayes Nets (CBNs) are today the dominant type of model for formalizing causal inferences [3,4,5,6,7]. Applied to basic diagnostic inferences, CBNs can define basic causal structures with three parameters.

We first find P_c, which represents the *cause base rate*, considered here as the prior probability of the cause. Knowledge of this parameter generally comes from an external source, for instance, it could be reported by an expert. We then find the *causal power* of the target cause W_c. This is an unobservable parameter that can only be estimated from the data provided by nature.[1] Finally, to be fully characterized, the network must be complemented by a *nuisance parameter* that represents all possible alternative causes. Associated with the nuisance parameter is W_a, which is analogues to W_c. It corresponds to the aggregate formed from the causal power and the base rates of all possible alternative causes, and it represents the probability that the effect is present while the target cause is absent.

The activation function (*noisy-OR* type) of the causal network implies that the effect can be generated independently by the target cause, or by the amalgam of the alternative causes, or by these two variables simultaneously [3,8,10]. In

[1] According to power PC theory [8], *causal power* represents the probability of a cause, acting alone, to produce a effect: $W_c = \big(\Pr(e \mid c) - \Pr(e \mid \neg c)\big) \,/\, \big(1 - \Pr(e \mid \neg c)\big)$. Given that causes are never observed alone, this is a theoretical measure that can only be estimated from data.

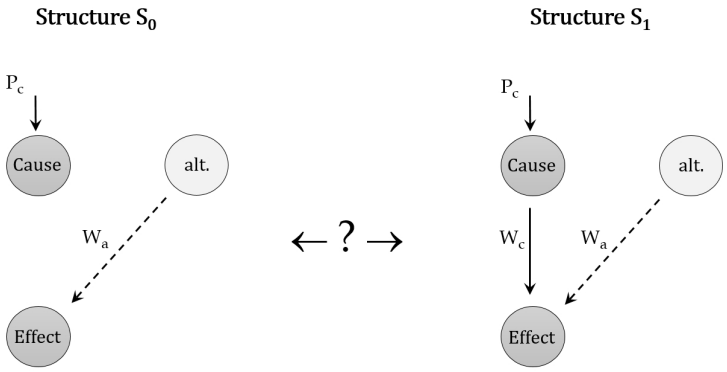

Fig. 1. Competing causal structures postulated by the SIM. The figure on the **left hand side** represents the structure S_0 that lacks a connection between the *Cause* of interest and the *Effect*. The effect can only be triggered by the alternative causes (abbreviated here as "alt."). This structure is characterized by only two parameters: P_c and W_a. The figure on the **right hand side** represents the structure S_1, which introduces a causal relation between the cause of interest and the effect. S_1 is more complex than S_0 because it is defined by three parameters: P_c, W_c, and W_a. Here, the effect can be produced by the cause of interest, or by alternative causes, or by both (activation function noisy-OR).

this context, $\Pr(c \mid e)$ is a function of the three parameters defined above, as follows:

$$\Pr(c \mid e) \;=\; 1 - (1 - P_c) \times \frac{W_a}{P_c \cdot W_c + W_a - P_c \cdot W_c \cdot W_a}. \tag{1}$$

2.1 The Structure Induction Model

Meder et al.'s previously mentioned SIM builds on the CBN literature. In their model, the diagnostic probability depends on the parameters P_c, W_c, W_a, but also on the uncertainty concerning the causal structure itself. An individual in a diagnosis situation is often uncertain about which world she inhabits. Figure 1 illustrates the situation in which she does not know whether she is in a world where there is a causal connection between the cause of interest and the effect (structure S_1) or whether she is in a world where such a link does not exist (structure S_0). In such a situation, estimating the diagnostic probability follows a three-step process:

1. Based on available data D, the first step consists in estimating, by Bayesian inference, the said parameters for each structure, S_0 and S_1, separately. We will call θ the set of parameters to be estimated. For S_0, there will be a guess of P_c and W_a (the value of W_c is set to 0). For the S_1 structure, there will be a guess of P_c, W_c, and W_a. The *posterior* probability distribution of the

parameters for a structure S_i is given by

$$\Pr(\theta \,|\, D\,; S_i) \;=\; \frac{\Pr(D \,|\, \theta\,; S_i)\;\Pr(\theta \,|\, S_i)}{\Pr(D \,|\, S_i)}, \tag{2}$$

where the *prior* probabilities of the parameters, $\Pr(\theta \,|\, S_i)$, are assumed to follow a Beta$(1, 1)$ distribution. Under a *noisy-OR* hypothesis, the likelihood of the data given the parameters values for a structure S_i, $\Pr(D \,|\, \theta\,; S_i)$, are given by Equations 3 (structure S_0) and 4 (structure S_1):[2]

$$\begin{aligned}\Pr(D \,|\, \theta\,; S_0) \;=\; & [(1-P_c)(1-W_a)]^{N(\neg c,\,\neg e)} \cdot [(1-P_c)W_a]^{N(\neg c,\,e)} \cdot \\ & [P_c(1-W_a)]^{N(c,\,\neg e)} \cdot [P_c\,W_a]^{N(c,\,e)}\,; \end{aligned} \tag{3}$$

$$\begin{aligned}\Pr(D \,|\, \theta\,; S_1) \;=\; & [(1-P_c)(1-W_a)]^{N(\neg c,\,\neg e)} \cdot [(1-P_c)W_a]^{N(\neg c,\,e)} \cdot \\ & [P_c(1-W_c)(1-W_a)]^{N(c,\,\neg e)} \cdot [P_c(W_c+W_a-W_c.W_a)]^{N(c,\,e)}. \end{aligned} \tag{4}$$

2. The next step consists in estimating the probabilities of the causal structures S_i themselves. The calculation of these probabilities is carried out for each of the structures S_0 and S_1 separately, on the basis of the available empirical data D and the parameters estimated in the previous step. This leads to the *posterior* probabilities of each structure under the data: $\Pr(S_0 \,|\, D)$ and $\Pr(S_1 \,|\, D)$.[3] These two conditional probabilities are given by

$$\Pr(S_i \,|\, D) \;=\; \frac{\Pr(D \,|\, S_i)\,\Pr(S_i)}{\Pr(D)}, \tag{5}$$

the probability of the data being given by

$$\Pr(D) \;=\; \sum_{i\in\{0,1\}} \Pr(D \,|\, S_i)\,\Pr(S_i). \tag{6}$$

$\Pr(S_i)$ is the *prior* probability of structure S_i and it was set to 0.5 for both structures. $\Pr(D \,|\, S_i)$ is the likelihood of the data given the structure and is computed by integrating over the likelihood functions of the parameters (see Equations 3 and 4) under structure S_i

$$\Pr(D \,|\, S_i) \;=\; \iiint \Pr(D \,|\, \theta\,; S_i)\,\Pr(\theta \,|\, S_i)\,d\theta, \tag{7}$$

[2] In these two equations, the terms c and e in the exponent $N(c,\, e)$, which may also occur negated, denote contingencies in the data about the target cause and the effect. For example, $N(\neg c,\, e)$ denotes the number of cases where the cause is missing and the effect present.

[3] A weak empirical contingency between the *Cause* of interest and the *Effect* will suggest that $\Pr(S_0 \,|\, D) > \Pr(S_1 \,|\, D)$. A strong contingency between *Cause* and *Effect* will instead suggest that $\Pr(S_0 \,|\, D) < \Pr(S_1 \,|\, D)$.

where $\Pr(\theta \mid S_i)$ denotes the joint *prior* probability over the structures' parameters. For both structures, this probability is assume to follow a $\text{Beta}(1,1)$ distribution.

3. Finally for S_0 and S_1, from the structure' parameters and from the *posterior* probabilities of the structures, one calculates two diagnostic probabilities. These probabilities are computed by integrating over the parameters' values weighted by their *posterior* probabilities:

$$\Pr(c \mid e\,; D, S_i) \;=\; \iiint \Pr(c \mid e\,; \theta, S_i) \frac{\Pr(D \mid \theta\,; S_i)\ \Pr(\theta \mid S_i)}{\Pr(D \mid S_i)}\, d\theta, \tag{8}$$

with

$$\Pr(c \mid e\,; \theta, S_0) \;=\; P_c, \tag{9}$$

and

$$\Pr(c \mid e\,; \theta, S_1) \;=\; \frac{(W_c + W_a - W_c W_a) P_c}{(W_c + W_a - W_c W_a) P_c + (1 - P_c) W_a}. \tag{10}$$

To obtain a single diagnostic probability $\Pr(c \mid e\,; D)$, the diagnostic probabilities for the two structures (see Equation 8) are weighted by the *posterior* probabilities of the corresponding structure (see Equations 5 and 6) and summed together:

$$\Pr(c \mid e\,; D) \;=\; \sum_{i \in \{0,1\}} \Pr(c \mid e\,; D, S_i) \Pr(S_i \mid D). \tag{11}$$

This *posterior* probability will thus take into account both the uncertainty concerning the parameters P_c, W_c, W_a, as well as the uncertainty related to the causal structure.[4]

In the process that leads to the estimation of the ultimate diagnostic probability $\Pr(c \mid e\,; D)$, the second step is crucial because it is the main source of the predictions of the SIM (the influence on the diagnostic probability $\Pr(c \mid e\,; D)$ of the predictive probability $\Pr(e \mid c)$). This step consists in determining the probability of each of the structures, knowing the data: $\Pr(S_i \mid D)$, for $i \in \{0, 1\}$. Since individuals have limited cognitive capacities [9,11] and do not have direct access to the parameters of the structure (e.g., the causal power of the cause of interest W_c is an unobservable parameter), they estimate $\Pr(S_i \mid D)$ by examining the contingencies between the *Cause* of interest and the *Effect* in the available data. This examination will lead them to estimate the predictive probability $\Pr(e \mid c)$. The calculation of this value functions as a kind of *heuristic* to approximate $\Pr(S_i \mid D)$ [12].

Thus, when $\Pr(e \mid c)$ is small, this suggests the absence of a link between *Cause* and *Effect*. In this case, $\Pr(S_0 \mid D)$ will be more salient than $\Pr(S_1 \mid D)$. Conversely, when $\Pr(e \mid c)$ is high, it will suggest the existence of a causal link and $\Pr(S_1 \mid D)$ will be more salient than $\Pr(S_0 \mid D)$. This mechanism constitutes

[4] See [2, p. 299] for more formal details involved in these three calculation steps.

the main novelty of the SIM, and it predicts the influence of the predictive probability on diagnostic judgments via the evaluation of the probability of the causal structure.

2.2 Key Predictions of the Structure Induction Model

The dependence of the diagnostic probability on the predictive probability is expected to introduce a bias in the estimation of the former. Specifically, the prediction is that when the predictive probability $\Pr(e \mid c)$ is low, individuals will tend to underestimate the diagnostic probability $\Pr(c \mid e)$ to a much greater extent than when $\Pr(e \mid c)$ is high.

This phenomenon can be explained intuitively by considering the situation in which an agent does not know which of the worlds S_0 and S_1 she inhabits. If the empirical predictive probability $\Pr(e \mid c)$ is small, this suggests that there may not be any link between *Cause* and *Effect*, in which case the agent will tend to believe that she is in world S_0. She will then be reluctant to attribute diagnostic virtues to the *Effect*, which in turn will lead her to underestimate the value of the diagnostic probability calculated from the data. On the other hand, if $\Pr(e \mid c)$ is high, the agent will be inclined to believe that there is a causal relationship between *Cause* and *Effect*, whence she is likely to conclude that she is in world S_1. In that case, the estimate of $\Pr(c \mid e)$ will more objectively reflect the empirical diagnostic probability of the data. This assumption has been confirmed empirically by Meder et al. [2]. In this study, we want to reproduce and generalize their important finding by testing the evolution of the bias as a function of the strategy that individuals use to estimate the diagnostic probability.

3 Diagnostic Reasoning Strategies: Abduction versus Defeasible Deduction

The SIM specifies a rational computation procedure which links the diagnostic judgments of two types of uncertainty (uncertainty about the parameters and uncertainty concerning the causal structure). In the terminology of Marr [13], this procedure is at the *computational* level of the cognitive system. However, Meder et al. also showed that the execution of the rational calculation will have consequences at the *algorithmic* (i.e., psychological) level, in particular, for the influence of the predictive probability on diagnostic judgments. At the algorithmic level, however, the psychological mechanisms underlying the estimation of the diagnostic probability itself have not been precisely described by these authors. We aim to fill this lacuna by introducing the concept of *estimation strategy* of the diagnostic probability.

Bruner, Goodnow, and Austin [14, p. 54] defined the concept of strategy in a very general way as a pattern of decisions in the acquisition, retention, and use of information to achieve specific objectives. Siegler and Jenkins [15, p. 11] clarified this idea further and defined strategies as sets of procedures or possible methods put in place by individuals to accomplish a given cognitive task.

Various results suggest that individuals can estimate the diagnostic probability by following essentially different inferential strategies. In two experiments, Stilgenbauer and Baratgin [16] showed that individuals preferred to follow an abductive strategy to estimate $\Pr(c \mid e)$ when the causal rule "If CAUSE then $\mathscr{P}$(EFFECT)" was made sufficiently plausible (the operator $\mathscr{P}$ has the intended meaning "there is a chance that"). In that kind of situation, participants' diagnostic inferences followed a probabilistic Affirming the Consequent schema (henceforth AC_p). This pattern of inference is also often recognized as the basic pattern of abduction [17,18,19]:

$$\frac{\text{EFFECT} \qquad \text{If CAUSE then } \mathscr{P}(\text{EFFECT})}{\mathscr{P}(\text{CAUSE})}$$

Stilgenbauer and Baratgin [16] further showed that individuals came to prefer estimating the diagnostic probability via a defeasible deduction type of reasoning when the plausibility of the rule "If CAUSE then $\mathscr{P}$(EFFECT)" decreased. In that kind of case, individuals tended to reason in accordance with defeasible Modus Ponens (henceforth MP_d; see [20,21,22]):

$$\frac{\text{EFFECT} \qquad \text{If EFFECT then } \mathscr{P}(\text{CAUSE})}{\mathscr{P}(\text{CAUSE})}$$

There is much evidence supporting the thought that AC_p and MP_d are the two major strategies used in diagnostic reasoning. For example, Patel and Groen [23] showed that medical students naturally form rules of the form "If CAUSE then $\mathscr{P}$(EFFECT)" to evaluate the likelihood of a disease (*Cause*) from a set of symptoms (*Effect*), that, in other words, their novice participants engaged in abductive reasoning. By contrast, more experienced doctors were found to typically construct "If EFFECT then $\mathscr{P}$(CAUSE)" rules to arrive at a diagnosis. Similarly, in the field of legal reasoning, Prakken and Renooij [24] showed that it is formally possible to follow a purely abductive strategy on the one hand and a strategy which uses defeasible deduction schemes on the other hand to trace the causes in a legal case.

4 Experimental Study of Diagnostic Reasoning Strategies for Estimating $\Pr(c \mid e)$

4.1 Predictions

The objective of this research is to test the predictions of the SIM, taking into account the strategies of diagnostic reasoning followed by participants.[5] As pre-

[5] This experiment is an extension of Chapter 6 of Jean-Louis Stilgenbauer's doctoral thesis [25].

viously explained, the SIM predicts an influence on the diagnostic probability $\Pr(c \mid e)$ of the predictive probability $\Pr(e \mid c)$, a phenomenon that leads in practice to an underestimation bias of the diagnostic probability. We assume that the bias will be more pronounced for abductive estimates than for estimates obtained via defeasible deduction, given that abduction requires to construct a rule of type "If CAUSE then $\mathscr{P}$(EFFECT)," which is related to the predictive probability $\Pr(e \mid c)$.[6] In this type of situation, $\Pr(e \mid c)$ should become more salient to participants, and since this value is connected to the underestimation bias predicted by the SIM, we expect to see here the clearest evidence of interference with the diagnostic estimate.

4.2 Participants

There were 114 participants in this experiment. All were French native speakers, studying at IPC Paris (Faculty of Philosophy and Psychology). Of these participants, 27 were male ($M = 20.3$, $sd = 1.5$) and 87 female ($M = 20.5$, $sd = 2$). We eliminated 17 subjects from the protocol: 8 because they left the experiment too early, and 9 because the latency of their responses was too high (over 3 minutes).

4.3 Materials

To test the impact of predictive probability on diagnostic probability estimates, we created four diagnostic reasoning situations, keeping empirical $\Pr(c \mid e)$ constant at 0.75 while varying empirical $\Pr(e \mid c)$ across the situations; specifically, $\Pr(e \mid c) \in \{0.1, 0.3, 0.6, 0.9\}$. In each situation, participants were asked to estimate $\Pr(c \mid e)$. All participants were exposed to the four diagnostic reasoning contexts, which were presented in an order randomized per participant (within-subjects factor). On the basis of the SIM, we predicted that participants would tend to estimate $\Pr(c \mid e)$ below the empirical probability of 0.75, and that their estimates would depend on the value of $\Pr(e \mid c)$.

The reasoning situations were medical cases in which there were imaginary viruses (Cause) that could infect people. An infected person might or might not develop a characteristic symptom (Effect). Each situation was introduced using a population-based stimulus that summarized a set of observations. The example

[6] There is a wealth of evidence showing that people tend to interpret the probability of an indicative conditional, Pr(If A, then C), as the conditional probability $\Pr(C \mid A)$; see, e.g., [26,27,28]. Admittedly, the conditional we are considering is not "If CAUSE then EFFECT," whose probability will, for most people, equal $\Pr(e \mid c)$, but rather "If CAUSE then $\mathscr{P}$(EFFECT)," whose probability will, by the same token, equal $\Pr(\mathscr{P}e \mid c)$, which is not necessarily equal to $\Pr(e \mid c)$. But, first, we are only claiming that "If CAUSE then $\mathscr{P}$(EFFECT)" makes $\Pr(e \mid c)$ salient, and that it can do by making $\Pr(\mathscr{P}e \mid c)$ salient (given the similarity between the two expressions). Second, although the two conditional probabilities are formally distinct, anyone who has taught a course in modal logic knows that people have a tendency to collapse iterated or even mixed modalities. As a result, many may fail to distinguish between $\Pr(e \mid c)$ and $\Pr(\mathscr{P}e \mid c)$ in the first place [29].

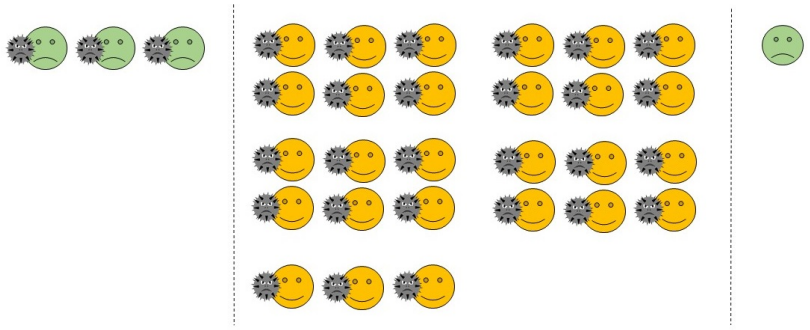

Fig. 2. Stimulus for empirical $\Pr(e \mid c) = 0.1$ and empirical $\Pr(c \mid e) = 0.75$. The green smileys on the left are infected by the virus (called *Igorusphère* in this example). They represent people who have developed the characteristic symptom (nausea, in this case). The yellow smileys in the center represent infected people without the characteristic symptom. Finally, the green smiley on the right represents an uncontaminated person with the symptom.

in Figure 2 shows the stimulus corresponding to the values $\Pr(e \mid c) = 0.1$ and $\Pr(c \mid e) = 0.75$. The other three stimuli were defined by the following probability pairs: $\Pr(e \mid c) = 0.3$ and $\Pr(c \mid e) = 0.75$; $\Pr(e \mid c) = 0.6$ and $\Pr(c \mid e) = 0.75$; and $\Pr(e \mid c) = 0.9$ and $\Pr(c \mid e) = 0.75$.

To test the impact of reasoning strategy ($\mathrm{AC}_p/\mathrm{MP}_d$) on estimates of $\Pr(c \mid e)$, we created four non-overlapping groups of participants (between-subjects factor). Each group was encouraged to estimate the diagnostic probability through a particular type of reasoning. In addition, we created a situation of "free" reasoning to determine the spontaneous and natural estimates of the participants. Participants were randomly assigned to one of four groups:

1. In the first group, participants were encouraged to reason in accordance with the abductive schema AC_p. To this end, we introduced with the stimuli the following conditional rule: "If CAUSE then $\mathscr{P}$(EFFECT)." The *Cause* is a fictitious virus and the *Effect* is its characteristic symptom. For example, with the material shown in Figure 2, the following rule appeared at the top of the picture: IF A PATIENT IS INFECTED WITH THE IGORUSPHÈRE, THERE IS A CHANCE THAT THE PATIENT HAS NAUSEA.

2. In the second group, the participants were encouraged to reason in accordance with the defeasible Modus Ponens schema (MP_d). Now, at the top of the picture, there appeared the default rule: "If EFFECT then $\mathscr{P}$(CAUSE)." For the example of Figure 2, the specific instance was: IF A PATIENT HAS NAUSEA, THERE IS A CHANCE THAT THE PATIENT HAS BEEN INFECTED WITH THE IGORUSPHÈRE.

3. In the third group, no rules were proposed, leaving participants completely free to estimate $\Pr(c \mid e)$ in whichever way they preferred. For each stimulus,

an accompanying sentence merely emphasized the existence of an association between *Cause* (disease) and *Effect* (the symptom). In this group, we first introduced the disease and only then the symptom. With Figure 2 appeared the sentence: THERE IS A CHANCE THAT IGORUSPHÈRE INFECTION IS ASSOCIATED WITH NAUSEA.

4. The fourth group was like the third—so this was also a free reasoning group—except that the order in which disease and symptom were introduced was reversed. For example, the sentence that now went with the situation depicted in Figure 2 was: THERE IS A CHANCE THAT NAUSEA IS ASSOCIATED WITH IGORUSPHÈRE INFECTION.

4.4 Procedure

The experiment was implemented on the *SoSci Survey* website (https://www.soscisurvey.de/). Participants were recruited via a group email. Once connected to the experiment, they were informed that they were supposed to answer all four questions. It was emphasized that their answers should be spontaneous and provided fairly quickly. Then, participants were asked to provide some demographic information: gender, age, native language, type and level of the university course. After this information had been recorded, the general experimental instructions were presented. In each of the four situations, participants were asked to estimate $\Pr(c \mid e)$, for the relevant c and e. More specifically, in each situation a new patient with the characteristic symptom of the virus displayed in the stimulus was presented to the participants. They were then asked to assess the chances that this new patient had been infected by the virus. Participants were asked to give their responses on a scale from 0 % to 100 %, which appeared beneath the stimulus, with the cursor initially set at the 0 % end of the scale.

4.5 Results

The results are summarized in Figure 3. A repeated measures analysis of variance (ANOVA) was carried out on the estimates of $\Pr(c \mid e)$ recorded in the free reasoning groups 3 and 4. There was no order effect related to the terms VIRUS/SYMPTOMS and SYMPTOMS/VIRUS: $F(1, 46) = 0.083$, $p > .05$. Accordingly, we merged the data from groups 3 and 4 for the remainder of the analysis.

Impact of predictive probability. A repeated measures ANOVA revealed a main effect of the within-subjects factor predictive probability. The estimates of diagnostic probability $\Pr(c \mid e)$ were found to depend on the predictive probability value $\Pr(e \mid c)$. The graph shows the $\Pr(c \mid e)$ estimates to increase (and to approximate the empirical diagnostic probability $\Pr(c \mid e) = 0.75$) as the predictive probability values $\Pr(e \mid c)$ increase. The effect was highly significant: $F(3, 276) = 27.27$, $p < .001$.

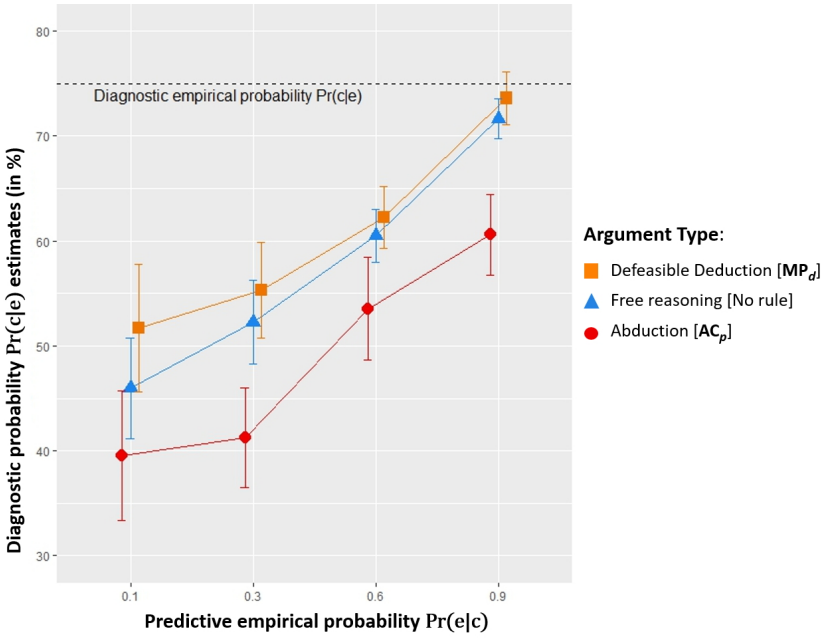

Fig. 3. Mean of diagnostic probability estimates $\Pr(c \mid e)$ performed under defeasible deduction, free reasoning, and abduction (error bars indicate standard errors). The diagnostic empirical probability fixed by the stimuli was constant throughout the experiment: $\Pr(c \mid e) = 0.75$. It is represented on the graph by the dotted horizontal line. Conversely, the predictive empirical probability $\Pr(e \mid c)$ varied for the four different stimuli: $\Pr(e \mid c) \in \{0.1, 0.3, 0.6, 0.9\}$.

Impact of estimation strategy. The ANOVA also showed a main effect of the between-subjects factor diagnostic probability estimation strategy. The effect was significant: $F(2, 92) = 3.44$, $p < .05$. Multiple comparisons with Bonferroni correction showed that the $\Pr(c \mid e)$ estimates under defeasible deduction did not vary significantly from estimates made in the free reasoning condition ($p > .05$). However, $\Pr(c \mid e)$ estimates made under abduction did vary significantly from estimates made under defeasible deduction ($p < .01$) as well as from estimates made under free reasoning ($p < 0.05$). The graph shows that $\Pr(c \mid e)$ estimates were more strongly underestimated compared to the empirical value $\Pr(c \mid e) = 0.75$ under abductive strategy than under defeasible deduction and free reasoning.

Interaction between predictive probability and estimation strategy. The statistical analysis did not reveal any interaction between the predictive probability $\Pr(e \mid c)$ and the diagnostic estimation strategy: $F(6, 276) = 0.17$, $p > .05$.

4.6 Discussion

Our results are clear evidence that participants systematically underestimated $\Pr(c \mid e)$, and that this bias was strongly related to the predictive empirical probability $\Pr(e \mid c)$. For small values of $\Pr(e \mid c)$, the underestimation of $\Pr(c \mid e)$ was maximal, and it decreased as the predictive probability increased. This result is important because it strongly supports the key calculation step of the causal structure probability $\Pr(S_i \mid D)$ that was expected in light of Meder, Mayrhofer, and Waldmann's Structure Induction Model (SIM).

Our data also show that the best estimates of $\Pr(c \mid e)$ were given under defeasible Modus Ponens and under free reasoning (there was no significant difference between these two conditions). This result suggests that defeasible deduction is the natural inference mode to estimate the diagnostic probability. In any case, diagnostic estimates made through abduction deviated much more from the empirical value $\Pr(c \mid e) = 0.75$. This confirms our initial hypothesis: when the salience of predictive probability $\Pr(e \mid c)$ is increased by the introduction of a causal rule of the form "If CAUSE then $\mathscr{P}$(EFFECT)," the underestimation bias predicted by the SIM gets worse.

Jointly, these results shed interesting new light on the reasoning process underlying defeasible Modus Ponens in causal contexts. In our opinion, this reasoning can be considered as an elementary form of *inference to the best explanation* [30,38] because the major premise of the MP_d schema (If EFFECT then $\mathscr{P}$(CAUSE)) is an *explanation-evoking rule* or *evidential rule* which, according to Pearl [32], suggests the activation and search for explanation. Coupled with the operating principle of the SIM—in particular the computation step of $\Pr(S_i \mid D)$—this inference tends to support the significant role played by explanatory considerations in the process of determining the diagnostic estimate [33,34,35].

In this work, we only tested a minimal type of explanatory consideration, one corresponding to the predictive probability $\Pr(e \mid c)$. In actuality, however, people may well exploit more complex predictive probability-based measures, such as Popper's measure [36] or Good's [37].[7] Whether this is so, we leave as a topic for future research.

References

1. Meder, B., Mayrhofer, R., Waldmann, M. R.: 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 (2009).

[7] According to Popper's measure, the explanatory power of c in light of e (plus background knowledge) is given by $\big(\Pr(e \mid c) - \Pr(e)\big) / \big(\Pr(e \mid c) + \Pr(e)\big)$. According to Good, that power is given by $\ln\big(\Pr(e \mid c) / \Pr(e)\big)$. See [33,38] for discussion of these and other probabilistic measures of explanatory power. See [39,40] for discussion of human (heuristic and analytic) cognitive process in causal reasoning.

2. Meder, B., Mayrhofer, R., Waldmann, M. R.: Structure induction in diagnostic causal reasoning. Psychological Review, 121(3), 277–301 (2014). doi.org/10.1037/a0035944
3. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Kaufmann, San Mateo: CA (1988). doi.org/10.1016/0004-3702(91)90084-W
4. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, New York (2000).
5. Glymour, C.: The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology. MIT Press, Cambridge MA (2001).
6. Hagmayer, Y.: Causal Bayes nets as psychological theories of causal reasoning: Evidence from psychological research. Synthese, 193(4), 1107–1126 (2016). doi.org/10.1007/s11229-015-0734-0
7. Rottman, B. M., Hastie, R.: Reasoning about causal relationships: Inferences on causal networks. Psychological Bulletin, 140, 109–139 (2014). doi.org/10.1037/a0031903
8. Cheng, P. W.: From covariation to causation: A causal power theory. Psychological Review, 104, 367–405 (1997). doi.org/10.1037/0033-295X.104.2.367
9. Simon, H.: Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting. Wiley, New York (1957).
10. Glymour, C.: Learning, prediction and causal Bayes nets. Trends in Cognitive Sciences, 7, 43–48 (2003). doi.org/10.1016/S1364-6613(02)00009-8
11. Simon, H.: Bounded rationality and organizational learning. Organization Science, 2, 125–134 (1991). doi.org/10.1287/orsc.2.1.125
12. Meder, B., Gerstenberg, T., Hagmayer, Y., Waldmann, M. R.: Observing and intervening: Rational and heuristic models of causal decision making. The Open Psychology Journal, 3(2), 119–135 (2010). doi.org/10.2174/1874350101003020119
13. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco, Freeman (1982).
14. Bruner, J. S., Goodnow, J. J., Austin, G. A.: A Study of Thinking. Wiley, New York (1956).
15. Siegler, R. S., Jenkins, E. A.: How Children Discover New Strategies. Lawrence Erlbaum Associates, Hillsdale (1989).
16. Stilgenbauer, J.-L., Baratgin, J.: Étude des stratégies de raisonnement causal dans l'estimation de la probabilité diagnostique à travers un paradigme expérimental de production de règle [Study of causal reasoning strategies in diagnostic probability estimation through an experimental rule production paradigm]. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale (2017). doi.org/10.1037/cep0000108
17. Magnani, L.: Abduction, Reason, and Science: Processes of Discovery and Explanation. Kluwer, New York (2001).
18. Aliseda, A.: Abductive Reasoning: Logical Investigations into Discovery and Explanation. Springer, Berlin (2006).
19. Schurz, G.: Patterns of abduction. Synthese, 164(2), 201–234 (2007). doi.org/10.1007/s11229-007-9223-4
20. Toulmin, S.: The Uses of Argument. Cambridge University Press, Cambridge (1958).
21. Reiter, R.: A logic for default reasoning. Artificial Intelligence, 13(1–2), 81–132 (1980). doi.org/10.1016/0004-3702(80)90014-4
22. Pollock, J. L.: Defeasible reasoning. Cognitive Science, 11(4), 481–518 (1987). doi.org/10.1207/s15516709cog1104_4

23. Patel, V. L., Groen, G. J.: Developmental accounts of the transition from medical students to doctor: Some problems and suggestions. Medical Education, 25, 527–535 (1991). doi.org/10.1111/j.1365-2923.1991.tb00106.x
24. Prakken, H., Renooij, S.: Reconstructing causal reasoning about evidence: a case study. In Legal Knowledge and Information Systems. JURIX: The Fourteenth Annual Conference (pp. 131–142). IOS Press, Amsterdam (2001).
25. Stilgenbauer, J.-L.: Étude expérimentale des stratégies de raisonnement causal dans l'estimation de la probabilité diagnostique : Stratégie abductive versus stratégie par déduction rétractable [Experimental Study of Causal Reasoning Strategies in the Estimate of Diagnostic Probability: Abductive Strategy versus Defeasible Deduction Strategy]. Thèse de doctorat, École Pratique des Hautes Études (EPHE), Paris - France (2016). doi.org/10.13140/RG.2.2.28698.03523
26. Baratgin, J., Politzer, G.: 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). MIT Press, Cambridge MA (2016).
27. Baratgin, J., Ocak, B., Bessaa, H., Stilgenbauer, J.-L.: Updating context in the Equation: An experimental argument with eye tracking. In M. B. Ferraro, P. Giordani, B. Vantagi, M. Gagolewski, M. Angeles Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Data Science, Advances in Intelligent Systems and Computing. Vol. 456 (pp. 25–33). Springer, Warsaw (2017). doi.org/10.1007/978-3-319-42972-4_4
28. Douven, I.: The Epistemology of Indicative Conditionals. Cambridge University Press, Cambridge (2016).
29. Over, D., Douven, I., Verbrugge, S.: Scope ambiguities and conditionals. Thinking & Reasoning, 19(3), 284–307 (2013). doi.org/10.1080/13546783.2013.810172
30. Lipton, P.: Inference to the Best Explanation. Routledge, London and New York (2004).
31. Douven, I.: Abduction. In E. N. Zalta (Ed.), Stanford Encyclopedia of Philosophy (2011). https://plato.stanford.edu/entries/abduction/
32. Pearl, J.: Embracing causality in default reasoning. Artificial Intelligence, 35(2), 259–271 (1988). doi.org/10.1016/0004-3702(88)90015-X
33. Douven, I., Schupbach, J. N.: The role of explanatory considerations in updating. Cognition, 142, 299–311 (2015). doi.org/10.1016/j.cognition.2015.04.017
34. Douven, I., Schupbach, J. N.: Probabilistic alternatives to Bayesianism: The case of explanationism. Frontiers in Psychology, 6(459), 1–9 (2015). doi.org/10.3389/fpsyg.2015.00459
35. Douven, I., Wenmackers, S.: Inference to the Best Explanation versus Bayes's Rule in a Social Setting. The British Journal for the Philosophy of Science, in press. doi.org/10.1093/bjps/axv025
36. Popper, K. R.: The Logic of Scientific Discovery. Hutchinson, London (1959).
37. Good, I. J.: Weight of evidence, corroboration, explanatory power, information and the utility of experiment. Journal of the Royal Statistical Society. Series B (Methodological), 22(2), 319–331 (1960). http://www.jstor.org/stable/2984102
38. Douven, I.: 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 University Press, Oxford, in press.
39. Hattori, M., Over, D. E., Hattori, I., Takahashi, T., Baratgin, J.: Dual frames in causal reasoning and other types of thinking. In N. Galbraith, E. Lucas, D. E. Over (Eds.), The Thinking Mind: A Festschrift for Ken Manktelow (pp. 15-28). London: Routledge (2016).

40. Hattori, I., Hattori, M, Over, D. E., Takahashi, T., Baratgin, J.: Dual frames for causal induction: The normative and the heuristic. Thinking & Reasoning, in press. doi.org/10.1080/13546783.2017.1316314.