Journal of Artificial Intelligence Research 65 (2019) 783-806 Submitted 02/2019; published 08/2019

The Mathematics of Changing One's Mind, via Jeffrey's or via Pearl's Update Rule

Bart Jacobs BART@CS.RU.NL
Institute for Computing and Information Sciences,
Radboud University, Nijmegen, The Netherlands

Abstract

Evidence in probabilistic reasoning may be 'hard' or 'soft', that is, it may be of yes/no form, or it may involve a strength of belief, in the unit interval $[0, 1]$. Reasoning with soft, $[0, 1]$-valued evidence is important in many situations but may lead to different, confusing interpretations. This paper intends to bring more mathematical and conceptual clarity to the field by shifting the existing focus from specification of soft evidence to accomodation of soft evidence. There are two main approaches, known as Jeffrey's rule and Pearl's method; they give different outcomes on soft evidence. This paper argues that they can be understood as correction and as improvement. It describes these two approaches as different ways of updating with soft evidence, highlighting their differences, similarities and applications. This account is based on a novel channel-based approach to Bayesian probability. Proper understanding of these two update mechanisms is highly relevant for inference, decision tools and probabilistic programming languages.

1. Introduction

Logical statements in a probabilistic setting are usually interpreted as *events*, that is, as subsets $E \subseteq \Omega$ of an underlying sample space Ω of possible worlds, or equivalently as characteristic functions $\Omega \to \{0, 1\}$. One typically computes the probability $\Pr(E)$ of an event E, possibly in conditional form $\Pr(E \mid D)$ where D is also an event. Events form the basic statements in probabilistic inference, where they can be used as evidence or observation. Here we shall use a more general interpretation of logical statements, namely as functions $\Omega \to [0, 1]$ to the unit interval $[0, 1]$. They are sometimes called fuzzy events or fuzzy predicates, but we simply call them predicates.

The above description $\Omega \to \{0, 1\}$ of events/evidence is standard. It is sometimes called hard or certain or sharp evidence, in contrast to soft, uncertain, unsharp, or fuzzy evidence $\Omega \to [0, 1]$. In most textbooks on Bayesian probability (see *e.g.* Barber, 2012; Bernardo & Smith, 2000; Jensen & Nielsen, 2007; Koller & Friedman, 2009; Pearl, 1988), soft evidence is missing or is only a marginal topic. For instance, in Barber's textbook (2012, §3.2) it is discussed only briefly, namely as: "In soft or uncertain evidence, the evidence variable is in more than one state, with the strength of our belief about each state being given by probabilities." The topic gets relatively much attention in Darwiche's book (2009, §3.6-3.7) with a description: "Hard evidence is information to the effect that some event has occurred ... Soft evidence, on the other hand, is not conclusive: we may get an unreliable testimony that event β occurred, which may increase our belief in β but not to the point where we would consider it certain."

©2019 AI Access Foundation. All rights reserved.

Typically, soft evidence deals with statements like: I saw the object in the dark and I am only 70% sure that its color is red. Or: my elder neighboor has hearing problems and is only 60% certain that my alarm rang. As said, we interpret such evidence as fuzzy predicates, with a degree of truth in $[0, 1]$. Somewhat confusingly, these statements may also be interpreted as a state of affairs, that is as a probability distribution with a convex combination of 0.7 red and 0.3 non-red. It seems fair to say that there is no widely accepted perspective on how to interpret and reason with such soft evidence, and in particular on how to update with soft evidence. The mathematics of such updating is the main topic of this paper, which, in the words of Diaconis and Zabell (1983), is called: the mathematics of changing one's mind.

In fact, there are two main approaches to soft updating, that is, to updating with soft evidence. They are most clearly distinguished by Chan and Darwiche (2005), but see also Darwiche (2009), Diaconis and Zabell (1982), Valtorta, Kim, and Vomlel (2002).

1. One can use *Jeffrey's rule*, see Jeffrey (1983) and also Halpern (2003) and Shafer (1981). It interprets softness as a probability distribution that represents a new state of affairs that differs from what is predicted, and that one needs to *adjust*, *adapt* or *correct* to. Adjusting to 70% probability of seeing red involves a convex combination of point updates: one takes 0.7 times the belief revision for red plus 0.3 times the revision for not-red. This approach focuses on adjustment/correction to a new state of affairs. Phrases associated with this approach are 'probability kinematics' due to Jeffrey (1983), 'radical probabilism' due to Skyrms (1996), or dealing with 'surprises' stemming from Dietrich, List, and Bradley (2016) or with 'unanticipated knowledge' from Diaconis and Zabell (1983).

2. One can also use *Pearl's method of virtual evidence*, see Pearl (1988, 1990). This approach is described operationally: extend a Bayesian network with an auxiliary node, so that soft evidence can be emulated in terms of hard evidence on this additional node, and so that the usual inference methods can be applied. We shall see that extending a Bayesian network with such a node corresponds to using a fuzzy predicate to capture the soft evidence. This approach *factors in* the soft evidence, following the basic idea: $\textit{posterior} \propto \textit{prior} \cdot \textit{likelihood}$. It involves improvement instead of correction.

This paper takes a fresh mathematical perspective on a problem that exists already for a long time in probabilistic reasoning, going back to Jeffrey (1983) and Pearl (1990). This work builds on a novel approach to Bayesian probability theory, based on programming language semantics and ultimately on category theory, started by Giry (1982), see Jacobs (2018b) for a modern overview. This approach clearly separates (fuzzy) predicates (evidence) from probability distributions (states). It is therefor well-prepared to deal with softness/uncertainty, either as fuzzy predicate or as state of affairs. In our general reformulation, Pearl's rule uses a *predicate* as soft evidence and involves backward inference via predicate transformation (see Definition 4.2), in a known manner, see the work of Jacobs and Zanasi (2016, 2019). The main (novel) mathematical observation of this paper is that Jeffrey's rule is captured via a *state* (distribution) as soft evidence and via state tranformation with the Bayesian inversion ('dagger') of the channel at hand, introduced by Clerc, Dahlqvist, Danos, and Garnier (2017), see Definition 5.2.

One fundamental problem is that the terminology in this area is confusing and is not used consistently by various authors; see Mrad, Delcroix, Piechowiak, Leicester, and Abid (2015) for a good overview of the different terminologies and their meaning (and of the literature on this topic). It uses the terminology 'likelihood evidence' or 'uncertain evidence' as 'evidence *with* certainty' for what we call a predicate; it also uses 'soft evidence' as 'evidence *of* uncertainty' for a probability distribution. We shall build on the distinction between predicates and states, since both notions are mathematically well-defined (see below); we shall use evidence and probability distribution as alternative names for predicate and state. The adjectives soft, uncertain, fuzzy will be used here only in an informal sense, without making a distinction between them. This leads to the following table.

Here	predicate evidence	state probability distribution
(Mrad et al., 2015)	likelihood evidence uncertain evidence evidence with uncertainty	soft evidence evidence of uncertainty

In accordance with this table, we shall say that Pearl's update rule is evidence-based and Jeffrey's rule is state-based.

The literature on soft updating, see esp. Chan and Darwiche (2005) and Darwiche (2009) (and references given there), focuses on the way in which softness is specified. Quoting Chan and Darwiche (2005): "The difference between Jeffrey's rule and Pearl's method is in the way uncertain evidence is specified. Jeffrey requires uncertain evidence to be specified in terms of the *effect* it has on beliefs once accepted, which is a function of both evidence strength and beliefs held before the evidence is obtained. Pearl, on the other hand, requires uncertain evidence to be specified in terms of its *strength* only." This paper shifts the emphasis from *specification* of softness to *accomodation* of softness, that is, to the precise update rules, see Definition 4.2 and 5.2, using both predicates and states to capture softness. In the end, after Lemma 6.2, we demonstrate that specification in terms of the update effect only works in the deterministic case. It is thus not a method that can be used in general.

Some more technical background: within the compositional programming language perspective, a Bayesian network is a (directed acyclic) graph in the Kleisli category of the distribution monad $\mathcal{D}$ — or the Giry monad $\mathcal{G}$ for continuous probability theory — see Fong (2012). The maps in these Kleisli categories are also called *channels*; they carry lots of useful algebraic structure that forms the basis for a compositional approach to probability. Along these channels one can do state transformation and predicate transformation, like in programming language semantics. These transformations are of direct relevance in Bayesian inference as shown by Jacobs and Zanasi (2016, 2019), giving rise, for instance, to a new inference algorithm (see Jacobs, 2018a). This paper builds on this 'channel-based' approach to give a novel precise account of Jeffrey's and Pearl's update rules. However, no familiarity with category theory is assumed and all the relevant concepts are introduced here.

The paper starts by elaborating a standard Bayesian example of a disease, with a prior probability, and a test for the disease that has a certain sensitivity; the question is: what can we infer about the disease if we are 80% sure the test comes out positive? We illustrate how to compute the different outcomes of Jeffrey and Pearl (12% versus 3% disease probability). We postpone the mathematical analysis and first go into reflective mode in Section 3.

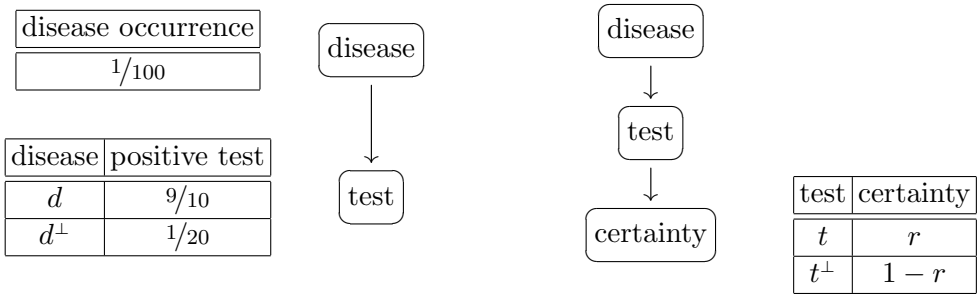


Figure 1: A Bayesian network for testing for a disease, on the left, and an extension of this network with a certainty node, where the relevant probability is a parameter $r \in [0, 1]$.

There we consider the question how to understand and when to use Jeffrey's or Pearl's approach. This leads to a terminological table (3). The mathematics itself is precise and clear, see from Section 4 onwards, but it often remains unclear when to use which approach. Our terminology of 'adjusting to a new state of affairs' (Jeffrey) versus 'factoring in new evidence' (Pearl) is meant to provide some guidance, but we are the first to admit that this remains vague — see also Example 6.4, originally introduced by Dietrich et al. (2016), where both approaches are used, for different reasons. Section 3 briefly mentions some further perspectives. For instance, if you perform Jeffrey's updating of your belief with what you can predict you learn nothing new; if you do Pearl's updating with no information (a uniform likelihood), you learn nothing new. Both make sense, but they are clearly different. This reflective section is meant chiefly to generate further discussion on this fundamental and intriguing topic, but not to provide a decision mechanism for the 'right' form of updating.

In Section 4 the mathematical analysis starts. First, background information is given about states, predicates, updating, channels, and transformation along channels. This allows us to identify Pearl's rule as backward inference. Section 5 first explains the Bayesian inversion of a channel and then uses this construction to capture Jeffrey's rule. Subsequently, Section 6 reviews some standard examples from the literature in terms of the new channel-based framework, and then shows how the earlier methods focused on specification of soft evidence — also known as "all things considered" and "nothing else considered" in the words of Goldszmidt and Pearl (1996) — fit naturally in the new framework.

2. A simple Illustration

Consider a simple Bayesian network involving a test for a disease, as on the left in Figure 1. There is an a priori disease probability of 1%. The test has a sensitivity as given by the table on the lower-left in the figure: in presence of the disease, written as d, the likelihood of a positive test outcome is 90%; in absence of the disease, there is still a chance of 5% that the test comes out positive.

In this situation we can compute the predicted positive-test likelihood $\Pr(t)$ via the law of total probability, as:

$$\begin{aligned}\Pr(t) &= \Pr(t \mid d) \cdot \Pr(d) + \Pr(t \mid d^{\perp}) \cdot \Pr(d^{\perp}) \\ &= \tfrac{9}{10} \cdot \tfrac{1}{100} + \tfrac{1}{20} \cdot \tfrac{99}{100} \;=\; \tfrac{117}{2000} \;\sim\; 6\%.\end{aligned}$$

The probability of the disease d, given a positive test t, is computed via Bayes' rule:

$$\Pr(d \mid t) \;=\; \frac{\Pr(t \mid d) \cdot \Pr(d)}{\Pr(t)} \;=\; \frac{{}^{9}/_{10} \cdot {}^{1}/_{100}}{{}^{117}/_{2000}} \;=\; \frac{18}{117} \;\sim\; 15\%$$

Similarly one obtains the conditional probability $\Pr(d \mid t^{\perp}) = \frac{{}^{1}/_{1000}}{{}^{1883}/_{2000}} = \frac{2}{1883} \sim 0.1\%$ of the disease given a negative test $t^{\perp}$.

This paper focuses on *soft* evidence. It arises for instance in a situation where the test outcome is observed in the dark, and that there is, say, only 80% certainty when the test is positive (and 20% certainty when it is negative).

There are two ways in the literature for handling soft evidence, called Jeffrey's rule and Pearl's method for virtual evidence. Jeffrey's rule says that we should take the convex combination, with factors 0.8 and $0.2 = 1 - 0.8$, of the "point updates", for the point evidence t and $t^{\perp}$. Thus one takes the convex combination of the above outcomes $\Pr(d \mid t)$ and $\Pr(d \mid t^{\perp})$, resulting in the probability:

$$0.8 \cdot \Pr(d \mid t) + 0.2 \cdot \Pr(d \mid t^{\perp}) \;\sim\; 12\%. \tag{1}$$

Thus, the certainties — 80% for a positive test and thus 20% for a negative test — are used as weights for the two corresponding conditional probabilities $\Pr(d \mid t)$ and $\Pr(d \mid t^{\perp})$. This makes sense.

In contrast, Pearl's rule involves extending the Bayesian network with an additional binary node for 'certainty', as on the right in Figure 1. One can then compute the probability of the disease if the test is positive with 80% certainty in the usual Bayesian way — by taking $r = 0.8$ in the lower-right table in Figure 1:

$$\Pr(d \mid c) \;=\; \Pr(d \mid t) \cdot \Pr(t \mid c) + \Pr(d \mid t^{\perp}) \cdot \Pr(t^{\perp} \mid c) \;=\; \tfrac{148}{4702} \;\sim\; 3\%. \tag{2}$$

This approach also makes sense. But its outcome differs substantially from Jeffrey's outcome of 12% in (1). Which rule is the right one here: Jeffrey's or Pearl's?

In order to get a better picture we now take the soft evidence probability for a positive test as a parameter $r \in [0, 1]$. Thus the number r represents the certainty of a positive test. The resulting a posteriori disease probabilities are plotted in Figure 2, on the left for the a priori disease probability of 1%, and on the right for the higher prior of 10%. We see that the two lines coincide at the extremes, for $r = 0$ and $r = 1$, corresponding to 100% certainty of a negative test and 100% certainty of a positive test. Inbetween these extremes, when $0 < r < 1$, the outcomes differ. Thus, making a distinction between the use of a Jeffrey's and Pearl's rule really only makes sense for soft evidence.

We also see that Jeffrey's rule yields a straight line. This is because it is defined by the linear (convex) function:

$$r \longmapsto r \cdot \Pr(d \mid t) + (1 - r) \cdot \Pr(d \mid t^{\perp}).$$

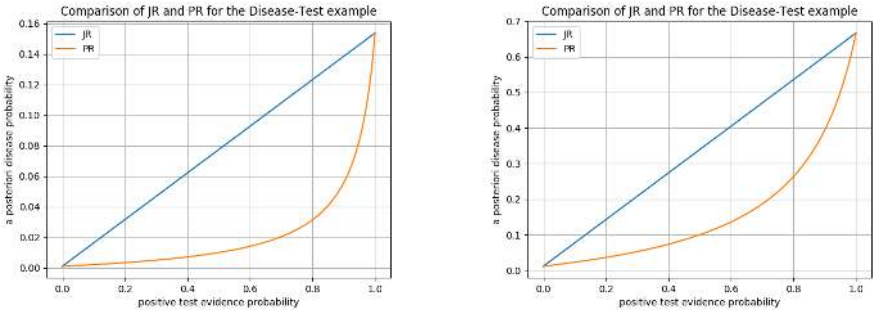


Figure 2: Probabilities obtained by applying both Jeffrey's rule (JR) and Pearl's rule (PR) to the disease-test network of Figure 1. The plot on the left captures the outcomes for an a priori disease probability of 1%, whereas the plot on the right uses 10% instead.

Pearl's rule gives a non-linear outcome, according to the familiar formula: $\textit{posterior} \propto \textit{prior} \cdot \textit{likelihood}$.

A final observation is that both rules do take the prior into account: the range of outcomes is quite different on the left and on the right.

3. Some Observations about Jeffrey's and Pearl's Rules

Before focusing on a mathematical analysis we like to make some remarks about the delicate question which form of updating — Jeffrey's or Pearl's — is the 'right' one. It is an important issue, for instance in the implementation of inference tools, see the overview of Mrad et al. (2015), or decision support systems, since as we have seen in the previous section, the two approaches give radically different outcomes.

The question 'which rule is the right one' may be refined to: under which circumstances should we use which rule, with which interpretation of softness?

Here we propose the following intuitive explanation of the two approaches, applied to the disease-test example from the previous section, where, recall, we had 80% certainty about a postive test outcome.

- Jeffrey's approach is state-based and uses the 80% positive-test certainty as a probability distribution (state), for which we shall use the following notation: $0.8|t\rangle + 0.2|t^{\perp}\rangle$. This means that we interpret it as a given *state of affairs* in which the test has a positive outcome t with a likelihood of 80% and a negative test outcome $t^{\perp}$ with 20% probability. When we see this state of affairs as a new situation — a 'surprise' as suggested by Dietrich et al. (2016) — and we wish to *adjust*, *adapt* or *correct* to this state of affairs, we use Jeffrey's rule as a form of backtracking.

- Pearl's approach is evidence-based: the 80% certainty is used as uncertain evidence that is *factored in*, via a suitable multiplication with the prior information (plus normalisation). The evidence is not treated as surprising, but as additional information

that is smoothly taken into account, in the regular Bayesian manner. This can be described either via an extra variable, or via a predicate, see Section 4.

The suggestion here is that Jeffrey's rule is for *correction* and Pearl's rule for *improvement*. The following table summarises the terminology.

rule	**uncertainty via**	**form of updating**
Jeffrey's	state of affairs proability distribution	state-based adjusting to correction
Pearl's	predicate evidence	evidence-based factoring in improvement

(3)

The remainder of this section contains some general observations and questions for further research.

1. From a mathematical perspective, Pearl's update rule is most well-behaved. In particular, iterated applications of the constructive rule commute, see Proposition 4.3 (3), whereas multiple usages of Jeffrey's rule do not commute. This is in line with the idea that the Jeffrey's approach involves abrupt adjustments/adaptations.

2. Pearl's rule makes classical use of Bayesian networks, as is illustrated via the additional binary node in Figure 1, on the right. In inference in such networks one factors in the evidence by propagating it through the network — and then marginalising.

3. In certain (other) cases one may explicitly wish to have an alternative rule for updating. For instance, Valtorta et al. (2002) describe a model of multi-agent systems, each with their own knowledge represented via a local Bayesian network. It is explicitly required that: "The mechanism for integrating the view of the other agents on a shared variable is to replace the agent's current belief in this variable with that of the communicating agent." Such replacements are obtained via Jeffrey's rule.

4. One can try to think of experimental verifications of the rules of Jeffrey/Pearl. A frequentist approach involves computing ratio's via counting and seems to support Jeffrey's form of updating. After all, Jeffrey's rule involves taking a convex sum of updates with individual point observations.

5. If probabilistic updating is seen as a mathematical model (or approximation) of cognitive priming, see *e.g.* Griffiths, Kemp, and Tenenbaum (2008), then the non-commutativity of iterated applications of Jeffrey's rule may be seen as a good thing. Indeed, the human mind is sensitive to the order in which it receives evidence, that is, in which it is being primed. This 'order effect' of priming can be illustrated in simple examples. The author's favourite one is: what image arises in your mind from the following two sequences of sentences?

Alice is pregnant; Bob visits Alice
versus
Bob visits Alice; Alice is pregnant.

Maybe cognitive psychologists can provide more clarity about wether Jeffrey's or Pearl's rule works best in their field, see also the suggested connection to the work of Hohwy (2013) in Section 7.

6. If probabilistic updating is seen as a form of learning — in an informal sense, not as parameter/structure learning — then one can ask what is the best model for handling evidence: correction of existing knowledge, as in the Jeffrey's approach, or improvement of existing knowledge, as in Pearl's approach. The question also comes up by comparing Propositions 4.3 (2) and 5.3 (1). They can be read informally as follows.

   (a) Pearl's improvement-based rule says: if you update (improve) your belief with no information (a uniform likelihood), you learn nothing new.
   (b) Jeffrey's correction-based rule says: when you update (correct) your belief with what you already know, you learn nothing new.

   Both readings make sense and connect the informal reading (improvenment versus correction) to mathematical facts.

7. One might think that the distinction Jeffrey/Pearl is related to whether or not the base rate (prior distribution) is taken into account in probabilistic reasoning. As shown by Tversky and Kahneman (1982), people are not very good at doing so. But the outcomes of both rules do depend on the prior, see the (vertical scales of the) two plots in Figure 2.

8. In the end, one can imagine using a combination of Jeffrey's rule (JR) and a Pearl's rule (PR), via a convex sum

$$s \cdot \mathrm{JR} + (1 - s) \cdot \mathrm{PR}$$

   The number $s \in [0, 1]$ then captures the novelty of the evidence. Very speculatively, it may be related to the degree to which the evidence's effect is absorbed (in one's brain).

4. Channel-based Probabilistic Reasoning

This section lays the foundation for our mathematical description of soft updating, using either Pearl's or Jeffrey's rule. Traditionally in probabilistic logic *events* are used as evidence. Such events form subsets $E \subseteq X$ of the sample space X; they correspond to characteristic functions $\mathbf{1}_E \colon X \to \{0, 1\}$, defined by $\mathbf{1}_E(x) = 1$ iff $x \in E$. As is well-known, these events (subsets) form a Boolean algebra. In order to deal with softness we use more general 'fuzzy' predicates, of the form $p \colon X \to [0, 1]$, sending each element $x \in X$ to a probability

$p(x) \in [0,1]$, representing the strength of belief. Such a predicate is called 'likelihood evidence' by Mrad et al. (2015) or a 'fuzzy event' by Zadeh (1968) where $p(x)$ described the 'grade of membership'. These predicates do not form a Boolean algebra, but what is called an effect module, (see *e.g.* Jacobs, 2015, 2018b).

Below we sketch a reformulation of the basics of probabilistic reasoning, in order to systematically accomodate soft/uncertain/fuzzy evidence. This reformulation uses basic mathematical concepts like distribution (state), fuzzy predicate, channel, conditioning, state- and predicate-transformation. These concepts stem from the area of program semantics where a distinction between predicate-transformer and state-transformer semantics is common (see *e.g.* Dijkstra & Scholten, 1990; Kozen, 1985). For a more extensive introduction of these concepts in probabilistic reasoning we refer to Jacobs and Zanasi (2019), and to Jacobs (2018b) and Panangaden (2009) for more general probabilistic semantics. We shall use the Bayesian network from Section 2 to illustrate the abstract concepts that we introduce below.

4.1 Distributions / States

In this context, we use the words 'distribution' and 'state' interchangeably, for what is more precisely called a discrete probability distribution, or also a monomial. A distribution on a set, or sample space, X is a formal convex combination of elements of X, written as $r_1|x_1\rangle + \cdots + r_n|x_n\rangle$ with $x_i \in X$ and $r_i \in [0,1]$ satisfying $\sum_i r_i = 1$. For instance, the prior disease distribution in Section 2 can be written as $\frac{1}{100}|d\rangle + \frac{99}{100}|d^\perp\rangle$. This looks a bit heavy for a distribution over a two-element set $\{d, d^\perp\}$, but this works better for multiple elements, see *e.g.* Example 6.1. The ket notation $|-\rangle$ is syntactic sugar that separates probabilities r_i and elements x_i.

We shall write $\mathcal{D}(X)$ for the set of distributions (or states) on a set X. We do not require that X is finite itself, but $\mathcal{D}(X)$ contains only finite distributions. A distribution $\sum_i r_i|x_i\rangle \in \mathcal{D}(X)$ may equivalently be described via a probability mass function $\omega\colon X \to [0,1]$ with finite support $\{x \mid \omega(x) \neq 0\}$ and with $\sum_x \omega(x) = 1$. We shall freely switch back-and-forth between formal convex sums and probability mass functions.

4.2 Channels

A channel from a set X to a set Y is probabilistic computation taking an element $x \in X$ as input and producing a distribution on Y, indicating the probability of each output $y \in Y$. Thus, a channel is a function $c\colon X \to \mathcal{D}(Y)$. We often write it as $c\colon X \mathrel{-\!\!\circ\!\!\to} Y$, with a special arrow $\mathrel{-\!\!\circ\!\!\to}$. A channel formalises a conditional probability $\Pr(y \mid x)$ as an actual function $x \mapsto \Pr(y \mid x)$. A channel thus captures an arrow in a Bayesian network, namely as a stochastic matrix, or equivalently as a conditional probability table. For instance, the sensitivity table in Figure 1 can be described as a channel:

$$\{d, d^\perp\} \xrightarrow{s} \{t, t^\perp\} \qquad \text{with} \qquad \begin{cases} s(d) = \frac{9}{10}|t\rangle + \frac{1}{10}|t^\perp\rangle \\ s(d^\perp) = \frac{1}{20}|t\rangle + \frac{19}{20}|t^\perp\rangle. \end{cases} \tag{4}$$

This channel s represents the arrow $\boxed{\text{disease}} \to \boxed{\text{test}}$ in Figure 1 as a probabilistic function $\{d, d^\perp\} \mathrel{-\!\!\circ\!\!\to} \{t, t^\perp\}$. Channels provide a compositional semantics for Bayesian networks, see Jacobs and Zanasi (2019) for more details.

Given a channel $c\colon X \nrightarrow Y$ one can transform a state $\omega \in \mathcal{D}(X)$ on X into a state $c \gg \omega \in \mathcal{D}(Y)$ on Y. This corresponds to prediction. Concretely, we can describe state transformation as below, first in mass function form, and then as convex formal sum:

$$\big(c \gg \omega\big)(y) = \sum_x \omega(x) \cdot c(x)(y) \quad \text{that is} \quad c \gg \omega = \sum_y \big(\sum_x \omega(x) \cdot c(x)(y)\big)\big|\, y\big\rangle.$$

For instance, the predicted test probability $\Pr(t) = \frac{117}{2000}$ in Figure 1 can be obtained via state transformation as:

$$\begin{aligned} & s \gg \Big(\tfrac{1}{100}|\, d\big\rangle + \tfrac{99}{100}|\, d^\perp\big\rangle\Big) \\ &= \big(\tfrac{1}{100} \cdot s(d)(t) + \tfrac{99}{100} \cdot s(d^\perp)(t)\big)|\, t\big\rangle + \big(\tfrac{1}{100} \cdot s(d)(t^\perp) + \tfrac{99}{100} \cdot s(d^\perp)(t^\perp)\big)|\, t^\perp\big\rangle \\ &= \big(\tfrac{1}{100} \cdot \tfrac{9}{10} + \tfrac{99}{100} \cdot \tfrac{1}{20}\big)|\, t\big\rangle + \big(\tfrac{1}{100} \cdot \tfrac{99}{100} + \tfrac{99}{100} \cdot \tfrac{19}{20}\big)|\, t^\perp\big\rangle \\ &= \tfrac{117}{2000}|\, t\big\rangle + \tfrac{1883}{2000}|\, t^\perp\big\rangle. \end{aligned}$$

Given two channels $c\colon X \nrightarrow Y$ and $d\colon Y \nrightarrow Z$ we can define their composite $d \bullet c\colon X \nrightarrow Z$ as:

$$\big(d \bullet c\big)(x) = d \gg c(x) = \sum_z \big(\sum_y c(x)(y) \cdot d(y)(z)\big)\big|\, z\big\rangle.$$

There is a 'Dirac' identity channel $\mathrm{id}\colon X \nrightarrow X$ for this composition $\bullet$, with $\mathrm{id}(x) = 1|\, x\big\rangle$. Moreover, $\bullet$ is associative and behaves well wrt. state transformation: $(d \bullet c) \gg \omega = d \gg (c \gg \omega)$. This gives an algebraic, compositional way for computing probabilities — especially in Bayesian networks.

Later on we shall use that each *function* $f\colon X \to Y$ can be turned into a 'deterministic' channel $\hat{f}\colon X \nrightarrow Y$ via $\hat{f}(x) = 1|\, f(x)\big\rangle$. Then it is easy to see that $\hat{g} \bullet \hat{f} = \widehat{g \circ f}$.

For instance, the Bayesian network on the right in Figure 1 involves an additional channel e that captures the certainty of the test evidence (for $r = \frac{8}{10}$):

$$\{t, t^\perp\} \overset{e}{\multimap} \{c, c^\perp\} \qquad \text{with} \qquad \begin{cases} e(t) = \frac{8}{10}|\, c\big\rangle + \frac{2}{10}|\, c^\perp\big\rangle \\ e(t^\perp) = \frac{2}{10}|\, c\big\rangle + \frac{8}{10}|\, c^\perp\big\rangle. \end{cases} \tag{5}$$

We can now compute the predicted certainty $\Pr(c) = \frac{4702}{20000}$, either via multiple state transformations, or via a single state transformation of the composed channel $e \bullet s\colon \{d, d^\perp\} \nrightarrow \{c, c^\perp\}$, in:

$$\begin{aligned} & (e \bullet s) \gg \Big(\tfrac{1}{100}|\, d\big\rangle + \tfrac{99}{100}|\, d^\perp\big\rangle\Big) \;=\; e \gg \Big(s \gg \Big(\tfrac{1}{100}|\, d\big\rangle + \tfrac{99}{100}|\, d^\perp\big\rangle\Big)\Big) \\ &= e \gg \Big(\tfrac{117}{2000}|\, t\big\rangle + \tfrac{1883}{2000}|\, t^\perp\big\rangle\Big) \\ &= \big(\tfrac{117}{2000} \cdot \tfrac{8}{10} + \tfrac{1883}{2000} \cdot \tfrac{2}{10}\big)|\, c\big\rangle + \big(\tfrac{117}{2000} \cdot \tfrac{2}{10} + \tfrac{1883}{2000} \cdot \tfrac{8}{10}\big)|\, c^\perp\big\rangle \\ &= \tfrac{4702}{20000}|\, c\big\rangle + \tfrac{15298}{20000}|\, c^\perp\big\rangle. \end{aligned}$$

4.3 Predicates, Validity and Updating

For a distribution $\sigma \in \mathcal{D}(X)$ on X and a predicate $p\colon X \to [0, 1]$ on X we write $\sigma \models p$ for the *validity* of p in σ. It can also be called the expected value, since the definition is:

$$\sigma \models p \;=\; \sum\nolimits_x \sigma(x) \cdot p(x). \tag{6}$$

For an event $E \subseteq X$ the validity $\sigma \models \mathbf{1}_E = \sum_{x \in E} \omega(x)$ is usually written as $\Pr(E)$, with the state σ left implicit. We need a new notation with the state σ explicit, since the state is not fixed: it changes through state transformation $\gg$.

If this validity $\sigma \models p$ is non-zero, we can define the updated, conditioned distribution $\sigma|_p$ on X as:

$$\sigma|_p(x) = \frac{\sigma(x) \cdot p(x)}{\sigma \models p} \qquad \text{that is} \qquad \sigma|_p = \sum_x \frac{\sigma(x) \cdot p(x)}{\sigma \models p} \big| x \big\rangle. \tag{7}$$

This updated/revised distribution $\sigma|_p$ is defined quite generally, for fuzzy predicates p. It allows us to express the usual form of conditional $\Pr(E \mid D)$ for events $E, D \subseteq X$ as $\sigma|_{\mathbf{1}_D} \models \mathbf{1}_E$.

The result below summarises some basic properties of updating with fuzzy predicates, including Bayes' rule (in fuzzy form), see Jacobs (2018b), Jacobs and Zanasi (2019). It uses conjunction $p \,\&\, q$ of two fuzzy predicates, defined as pointwise multiplication: $(p \,\&\, q)(x) = p(x) \cdot q(x)$. There is an associated truth predicate $\mathbf{1} \colon X \to [0,1]$ sending each element to 1, that is, $\mathbf{1}(x) = 1$. Then $p \,\&\, \mathbf{1} = p = \mathbf{1} \,\&\, p$. Moreover, it uses that a fuzzy predicate p can be multiplied with a scalar $s \in [0,1]$ to $s \cdot p \colon X \to [0,1]$, namely via $(s \cdot p)(x) = s \cdot p(x)$.

Lemma 4.1. *Let σ be distribution on a set X, and let p, q be predicates on X.*

1. *Bayes' rule holds for (fuzzy) predicates:*

$$\sigma|_p \models q \;=\; \frac{\sigma \models p \,\&\, q}{\sigma \models p} \;=\; \frac{(\sigma|_q \models p) \cdot (\sigma \models q)}{\sigma \models p}.$$

2. *Iterated conditionings commute:*

$$\big(\sigma|_p\big)|_q = \sigma|_{p \& q} = \big(\sigma|_q\big)|_p.$$

   Moreover, conditioning with truth has no effect: $\sigma|_{\mathbf{1}} = \sigma$.

3. *Conditioning is does not change when the predicate involved is multiplied with a non-zero scalar: $\omega|_p = \omega|_{s \cdot p}$.* □

A basic property of updating $\omega|_p$ is that the validity $\omega|_p \models p$ is greater than $\omega \models p$. Thus by changing ω into $\omega|_p$ the predicate p becomes 'more true' (see Jacobs, 2019, for a proof and more details). That's why we associate the phrase 'improvement' with this form of updating $\omega|_p$, which will be used below for Pearl's rule.

4.4 Predicate Transformation

We have seen how a state $\omega \in \mathcal{D}(X)$ can be transformed in a forward manner along a channel $c \colon X \to\!\!\!\!\bullet\; Y$, to a state $c \gg \omega$ on the codomain Y of the channel. One can also transform predicates along a channel, but in opposite direction: given a predicate $q \colon Y \to [0,1]$, one obtains a predicate $c \ll q \colon X \to [0,1]$ on the domain X of the channel via:

$$\big(c \ll q\big)(x) = \sum_y c(x)(y) \cdot q(y).$$

One then easily checks that the validities $c \gg \omega \models q$ and $\omega \models c \ll q$ are the same. Further there is a compositionality result $(d \bullet c) \ll q = c \ll (d \ll q)$ so that predicate transformation can be done by following the arrow / channel structure of a Bayesian network in a step-by-step manner.

4.5 Forward and Backward Inference

We are now combining state transformation, predicate transformation, and conditioning in order to identify two basic inference patterns, namely forward inference and backward inference, see Jacobs and Zanasi (2016, 2019). We start from:

$$\text{a state } \omega \in \mathcal{D}(X) \text{ on } X, \text{ and a channel } X \mathrel{-\!\!\circ\!\!\to} Y.$$

1. **Forward inference** with a predicate $p\colon X \to [0,1]$ is done by updating-and-state-transformation:
$$c \gg \big(\omega|_p\big).$$
This yields a new distribution on Y.

2. **Backward inference** with a predicate $q\colon Y \to [0,1]$ is done by predicate-transformation-and-updating:
$$\omega|_{c \ll q}.$$
This gives a new distribution on X.

In the literature, see *e.g.* the textbook of Koller and Friedman (2009), forward inference is also called *prediction* or *causal reasoning*, and backward inference is called *evidential reasoning* or *explanation*.

In this context backward inference plays the more important role. We illustrate it for the disease-test example from Section 2. Recall the characteristic function $\mathbf{1}_E\colon X \to [0,1]$ associated with an event/subset $E \subseteq X$. For an element $x \in X$ we simply write $\mathbf{1}_x\colon X \to [0,1]$ instead of $\mathbf{1}_{\{x\}}$.

Let's write $\omega = \frac{1}{100}|d\rangle + \frac{99}{100}|d^\perp\rangle$ for the prior disease probability from Section 2. Updating it with positive test evidence $\mathbf{1}_t$ happens via backward inference as $\omega|_{s \ll \mathbf{1}_t}$, using the sensitivity channel s from (4). As illustration, we compute it explicitly in several steps:

$$\begin{aligned}
(s \ll \mathbf{1}_t)(d) &= \textstyle\sum_{x \in \{t,t^\perp\}} s(d)(x) \cdot \mathbf{1}_t(x) \;=\; s(d)(t) \;=\; \frac{9}{10} \\
(s \ll \mathbf{1}_t)(d^\perp) &= s(d^\perp)(t) \;=\; \tfrac{1}{20} \\
\omega \models s \ll \mathbf{1}_t &= \textstyle\sum_{x \in \{d,d^\perp\}} \omega(x) \cdot (s \ll \mathbf{1}_t)(x) \;=\; \frac{1}{100} \cdot \frac{9}{10} + \frac{99}{100} \cdot \frac{1}{20} \;=\; \frac{117}{2000} \\
\omega|_{s \ll \mathbf{1}_t} &= \textstyle\sum_{x \in \{d,d^\perp\}} \dfrac{\omega(x) \cdot (s \ll \mathbf{1}_t)(x)}{\omega \models s \ll \mathbf{1}_t} \big|\, x \big\rangle \\
&= \frac{1/100 \cdot 9/10}{117/2000}|d\rangle + \frac{99/100 \cdot 1/20}{117/2000}|d^\perp\rangle \;=\; \frac{117}{2000}|d\rangle + \frac{1883}{2000}|d^\perp\rangle.
\end{aligned}$$

We see how the probability $\Pr(d \mid t) = \frac{117}{2000}$ from Section 2 re-emerges, via a channel-based computation. In a similar way one can compute $\Pr(d \mid c) = \frac{148}{4702}$ via backward inference as:

$$\omega|_{(e \bullet s) \ll \mathbf{1}_c} = \omega|_{s \ll (e \ll \mathbf{1}_c)} \;=\; \tfrac{148}{4702}|d\rangle + \tfrac{4554}{4702}|d^\perp\rangle.$$

We see that backward inference can be done in a compositional manner, following the graph structure $\boxed{\text{disease}} \to \boxed{\text{test}} \to \boxed{\text{certainty}}$ on the right in Figure 1.

4.6 Pearl's Update Rule

We are now finally in a position to describe Pearl's rule of virtual evidence, that is. Let's write $2 = \{0, 1\}$ for a generic two-element set. The crucial observation is that extending a Bayesian network at node X with virtual evidence of the form $X \mathrel{-\!\!\circ\!\!\to} 2$ corresponds to introducing a fuzzy predicate for updating. This works since $\mathcal{D}(2) \cong [0, 1]$, so that a table/channel $X \mathrel{-\!\!\circ\!\!\to} 2$ to a binary node 2 corresponds to a fuzzy predicate $X \to [0, 1]$. Indeed, the soft evidence described in Section 2 can be captured by a fuzzy predicate $p\colon \{t, t^\perp\} \to [0, 1]$ with $p(t) = \frac{8}{10}$ and $p(t^\perp) = \frac{2}{10}$. Pearl's rule then amounts to backward reasoning of the form $\omega|_{s \ll p} = \frac{148}{4702}|d\rangle + \frac{4554}{4702}|d^\perp\rangle$, as computed above. This works since $p = e \ll \mathbf{1}_c$.

We now formalise Pearl's rule in a channel-based setting.

Definition 4.2. Let $c\colon X \mathrel{-\!\!\circ\!\!\to} Y$ be a channel with prior $\sigma \in \mathcal{D}(X)$. Given a predicate $q\colon Y \to [0, 1]$ on the channel's codomain Y, Pearl's rule uses backward inference to update the prior σ to the posterior:

$$\sigma|_{c \ll q} \in \mathcal{D}(X).$$

This formulation of Pearl's rule does not refer to any extension of a Bayesian network with a binary node. Still, one may consider the channel $c\colon X \mathrel{-\!\!\circ\!\!\to} Y$ as a mini-network that is extended with predicate q, as in: $\boxed{X} \xrightarrow{c} \boxed{Y} \xrightarrow{q} \boxed{2}$.

We finish this section with some basic properties. They follow easily from Lemma 4.1.

Proposition 4.3. *Let $\omega \in \mathcal{D}(X)$ be a distribution and $c\colon X \mathrel{-\!\!\circ\!\!\to} Y$ be a channel.*

1. *Backward inference is invariant under pointwise/scalar multiplication of the evidence predicate with a non-zero probability $s \in (0, 1]$,*

$$\omega|_{c \ll (s \cdot p)} = \omega|_{c \ll p}.$$

2. *Backward inference with a non-zero constant (uniform) predicate $s \cdot \mathbf{1}$ as evidence has no effect:*

$$\omega|_{c \ll (s \cdot \mathbf{1})} = \omega.$$

3. *Iterated applications of backward inference commute, and satisfy:*

$$\omega\big|_{c \ll p}\big|_{d \ll q} = \omega\big|_{(c \ll p) \& (d \ll q)} = \omega\big|_{d \ll q}\big|_{c \ll p}.$$ □

5. Bayesian Inversion and Jeffrey's Update Rule

One way to read Bayes' rule is as an 'inversion' property, turning a conditional probability $\Pr(y \mid x)$ into $\Pr(x \mid y)$. Since channels correspond to conditional probabilities, such inversion can be formulated for channels as well, see Clerc et al. (2017) and also Cho and Jacobs (2019). This inversion is relevant because it allows us to give a precise description of Jeffrey's rule.

5.1 Bayesian Inversion via Updating

Let $c\colon X \nrightarrow Y$ be a channel, with a prior distribution/state $\sigma \in \mathcal{D}(X)$ on its domain. In this situation, with a certain side-condition fulfilled, we can define an inverted channel $c^\dagger_\sigma\colon Y \nrightarrow X$ in the opposite direction. This function $c^\dagger_\sigma\colon Y \to \mathcal{D}(X)$ is defined via backward inference with point predicates $\mathbf{1}_y\colon Y \to [0,1]$, for $y \in Y$.

$$c^\dagger_\sigma(y) = \sigma|_{c \ll \mathbf{1}_y} = \sum_x \frac{\sigma(x) \cdot c(x)(y)}{(c \gg \sigma)(y)} |\, x \,\rangle. \tag{8}$$

The distribution $c^\dagger_\sigma(y) \in \mathcal{D}(X)$ is the posterior, obtained after observing $y \in Y$, that is, after updating with point evidence $\mathbf{1}_y$. This definition only makes sense if the transformed state $c \gg \sigma$ has full support, that is, if $(c \gg \sigma)(y) \neq 0$ for each $y \in Y$.

The dagger notation $c^\dagger_\sigma$ for probabilistic computations is used by Clerc et al. (2017); the subscript σ may be omitted if it is clear from the context. This dagger satisfies some basic algebraic properties: inverting twice yields the original channel: $(c^\dagger)^\dagger = c$. Moreover, inversion interacts appropriately with channel composition: $(d \bullet c)^\dagger = c^\dagger \bullet d^\dagger$. The dagger notation is more common in quantum theory, where unitary computations are reversible, and has been formalised in terms of dagger categories, see *e.g.* the textbook of Coecke and Kissinger (2016).

Given the above definition (8) we see that we have implicitly already computed the Bayesian inversion $s^\dagger$ of the sensitivity channel $s\colon \{d, d^\perp\} \nrightarrow \{t, t^\perp\}$ from Section 2, namely via the conditional probabilities $\Pr(d \mid t) = \frac{18}{117}$ and $\Pr(d \mid t^\perp) = \frac{2}{1883}$. Thus we have:

$$\{t, t^\perp\} \xrightarrow{s^\dagger} \{d, d^\perp\} \qquad \text{with} \qquad \begin{cases} s^\dagger(t) = \frac{18}{117}|\, d \,\rangle + \frac{99}{117}|\, d^\perp \,\rangle \\ s^\dagger(t^\perp) = \frac{2}{1883}|\, d \,\rangle + \frac{1881}{1883}|\, d^\perp \,\rangle. \end{cases} \tag{9}$$

5.2 Bayesian Inversion and Inference

In Subsection 4.5 we have described forward and backward inference along a channel. It turns out that forward becomes backward — and vice-versa — when we use an inverted channel. This illustrates that the basic notions of inference, transformation and inversion are mathematically closely related. The proof is obtained by unwrapping the relevant definitions and is left to the interested reader.

Theorem 5.1. *Let $c\colon X \nrightarrow Y$ be a channel with a state $\sigma \in \mathcal{D}(X)$ on its domain, such that $\tau = c \gg \sigma$ has full support.*

1. *Given a predicate q on Y, we can express backward inference along c as forward inference along $c^\dagger$ via:*
$$\sigma|_{c \ll q} = c^\dagger \gg \big(\tau|_q\big).$$
2. *Given a predicate p on X, we can express forward inference along c as backward inference along $c^\dagger$:*
$$c \gg \big(\sigma|_p\big) = \tau|_{c^\dagger \ll p}.$$ □

5.3 Jeffrey's Update Rule

At this stage we have prepared the grounds to give a channel-based formulation of Jeffrey's rule. It uses the inversion of a channel for backtracking.

Definition 5.2. Let $c\colon X \nrightarrow Y$ be a channel with prior $\sigma \in \mathcal{D}(X)$. Given a state $\rho \in \mathcal{D}(Y)$ on the channel's codomain Y, Jeffrey's rule involves using state transformation along the inverted channel $c^\dagger_\sigma$ to update the prior σ to the posterior:

$$c^\dagger_\sigma \gg \rho \in \mathcal{D}(X).$$

Indeed, this state transformation is what we have used to compute Jeffrey's update in Section 2 as convex combination (1). More explicitly, translating 80% certainty into a state, and using $s^\dagger$ from (9) we get approximately 12% disease likelihood via:

$$\begin{aligned} s^\dagger \gg \left(\tfrac{8}{10}|t\rangle + \tfrac{2}{10}|t^\perp\rangle\right) &= \left(\tfrac{8}{10}\cdot\tfrac{18}{117} + \tfrac{2}{10}\cdot\tfrac{2}{1883}\right)|d\rangle + \left(\tfrac{8}{10}\cdot\tfrac{99}{117} + \tfrac{2}{10}\cdot\tfrac{1881}{1883}\right)|d^\perp\rangle \\ &= \tfrac{27162}{220311}|d\rangle + \tfrac{193149}{220311}|d^\perp\rangle. \end{aligned}$$

We continue with some properties of Jeffrey's updating. The translations back-and-forth between the Pearl's and Jeffrey's rules are due to Chan and Darwiche (2005); they are translated here to the current setting

Proposition 5.3. *Let $c\colon X \nrightarrow Y$ be a channel with a state $\sigma \in \mathcal{D}(X)$ on its domain, such that $\tau = c \gg \sigma$ has full support.*

1. *Jeffrey's updating with the predicted state $\tau = c \gg \sigma$ does not have any effect:*

$$c^\dagger \gg \tau = \sigma.$$

2. *Successive Jeffrey updates do not commute: given evidence $\rho_1, \rho_2 \in \mathcal{D}(Y)$, giving $\sigma_i = c^\dagger_\sigma \gg \rho_i \in \mathcal{D}(X)$, then, in general,*

$$c^\dagger_{\sigma_1} \gg \rho_2 \neq c^\dagger_{\sigma_2} \gg \rho_1.$$

3. *Jeffrey's and Pearl's updating coincide on point evidence:*

$$c^\dagger_\sigma \gg 1|y\rangle = c^\dagger_\sigma(y) = \sigma|_{c \ll \mathbf{1}_y}.$$

4. *Pearl's updating can be expressed as Jeffrey's updating, by turning predicate evidence $q\colon Y \to [0,1]$ into state evidence $\tau|_q$, see again Theorem 5.1 (1).*

5. *Jeffrey's updating can also be expressed as Pearl's updating: for a state $\rho \in \mathcal{D}(Y)$ write ρ/τ for the predicate $y \mapsto \frac{\rho(y)}{\tau(y)}$, suitably rescaled to $[0,1]$ if needed; then:*

$$c^\dagger_\sigma \gg \rho = \sigma|_{c \ll \rho/\tau}.$$

Proof. Only the last point is non-trivial. First we note that $\tau \models \rho/\tau = 1$, since ρ is a state:

$$\tau \models \rho/\tau = \sum_y \tau(y) \cdot \tfrac{\rho(y)}{\tau(y)} = \sum_y \rho(y) = 1.$$

But then, for $x \in X$,

$$\begin{aligned}
\big(c^\dagger_\sigma \gg \rho\big)(x) = \sum_y \rho(y) \cdot c^\dagger_\sigma(y)(x) &\overset{(8)}{=} \sum_y \rho(y) \cdot \frac{\sigma(x) \cdot c(x)(y)}{\tau(y)} \\
&= \sigma(x) \cdot \textstyle\sum_y c(x)(y) \cdot \rho/\tau(y) \\
&= \frac{\sigma(x) \cdot (c \ll \rho/\tau)(x)}{c \gg \sigma \models \rho/\tau} \qquad \text{as just shown} \\
&= \frac{\sigma(x) \cdot (c \ll \rho/\tau)(x)}{\sigma \models c \ll \rho/\tau} \\
&\overset{(7)}{=} \big(\sigma|_{c \ll \rho/\tau}\big)(x).
\end{aligned}$$

□

We conclude this section with a couple of remarks.

Remark 5.4. 1. Proposition 5.3 (1) shows that $\sigma = c^\dagger_\sigma \gg (c \gg \sigma)$. This means that a state σ can be reconstructed, via Jeffrey's updating, from what we can predict, namely from $c \gg \sigma$. At this same time it shows that in Jeffrey's updating $c^\dagger_\sigma \gg \rho$ the 'state of affairs' ρ that we encounter as evidence replaces the prediction $c \gg \sigma$, where the inversion $c^\dagger_\sigma$ is used for back-tracking. This replacement, of $c \gg \sigma$ by ρ, is where the 'shock' or 'surpise' of Jeffrey's rule can be located. We also use the terms 'correction', 'adjustment' and 'adaptation' for this process, see Table (3).

2. We briefly come back to the issue whether softness/uncertainty should be represented as a state or as a predicate. Pragmatically, one can go either way, since each state is a predicate, and in the other direction a predicate (on a finite) set can be normalised to a state, and the scaling factor involved does not affect the outcome in conditioning, see Lemma 4.1 (3).

   From a structural, algebraic perspective however, there are significant differences between states and predicates. For one, they form different mathematical structures: states are convex sets, whereas predicates are effect modules with a monoid structure (for conjunction), see *e.g.* Jacobs (2015, 2018b) for details. This means that they come with different algebraic operations. For instance, predicates are closed under scalar multiplication, but states are not. In addition, there are different transformation operations: states can be transformed forwardly $\gg$ along a channel, and predicates backwardly $\ll$. These operations are mathematically well-behaved: convex combinations of states are preserved by state transformation, whereas the effect module structure is preserved by predicate transformation. States and predicates are dual to each other, see *e.g.* Jacobs (2017, 2018b) for a wider perspective. These structural differences suffice to keep states and predicates apart in a mathematically precise manner.

   In addition, the mathematical distinction between states and predicates fits the terminological distinction of Table (3): a state of affairs in Jeffrey's updating corresponds to a state / probability distribution, whereas (soft) evidence corresponds to a (fuzzy) predicate. This means that the terminology has a mathematical basis.

6. Literature Review

This section compares the channel-based explanation of Jeffrey's/Pearl's updating of the previous two sections, together with its informal interpretation of Section 3, to some relevant material in the literature. It first reviews some examples and then looks at earlier approaches to soft evidence that focus mainly on how to formulate softness in the first place.

6.1 Examples from the Literature

Example 6.1. First we consider the following question from Halpern (2003, Example 3.10.1).

> Suppose that an object is either red (r), blue (b), green (g), or yellow (y). An agent initially ascribes probability 1/5 to each of red, blue, and green, and probability 2/5 to yellow. Then the agent gets a quick glimpse of the object in a dimly lit room. As a result of this glimpse, he believes that the object is probably a darker color, although he is not sure. He thus ascribes probability .7 to it being green or blue and probability .3 to it being red or yellow. How should he update his initial probability measure based on this observation?

The prior probability distribution is in this case $\sigma = \frac{1}{5}|r\rangle + \frac{1}{5}|b\rangle + \frac{1}{5}|g\rangle + \frac{2}{5}|y\rangle$. We see that the colors in this example are partitioned in two combinations, namely 'green or blue' and 'red or yellow'. We capture this via a two-element set $\{gb, ry\}$. There is then an obvious (deterministic) channel:

$$\{r, b, g, y\} \xrightarrow{c} \{gb, ry\} \qquad \text{with} \qquad \begin{cases} c(r) = 1|ry\rangle \\ c(b) = 1|gb\rangle \\ c(g) = 1|gb\rangle \\ c(y) = 1|ry\rangle \end{cases}$$

The above quote does not suggest whether the new information should be used for correction, or for improvement. Halpern (2003) chooses the first approach. Here we elaborate both.

The posterior (updated) probability distribution, computed via Jeffrey's rule, is obtained by doing state transformation with the inverted channel and the 'glimpse' as state of affairs:

$$c^\dagger_\sigma \gg \left(\tfrac{7}{10}|gb\rangle + \tfrac{3}{10}|ry\rangle\right) = \tfrac{1}{10}|r\rangle + \tfrac{7}{20}|b\rangle + \tfrac{7}{20}|g\rangle + \tfrac{1}{5}|y\rangle.$$

However, one can also translate the 'glimpse' into a fuzzy predicate $p\colon \{gb, ry\} \to [0, 1]$ with $p(gb) = \frac{7}{10}$ and $p(ry) = \frac{3}{10}$. Pearl's update rule then gives a different outcome:

$$\sigma|_{c \ll p} = \tfrac{3}{23}|r\rangle + \tfrac{7}{23}|b\rangle + \tfrac{7}{23}|g\rangle + \tfrac{6}{23}|y\rangle.$$

This example is an instance of a frequently occurring setting in which Jeffrey's rule is formulated — notably by Halpern (2003), to which we refer for details — namely when the channel involved is deterministic. Consider a function $f\colon X \to I$, giving a partition of the set X via subsets $U_i = f^{-1}(i) = \{x \in X \mid f(x) = i\}$. This function can be turned into a 'deterministic' channel $\hat{f}\colon X \to\!\!\!\!\circ\; I$, via $\hat{f}(x) = 1|f(x)\rangle$.

Lemma 6.2. *Let $f\colon X \to I$ be a function/partition, to be used as deterministic channel, as just described, together with a prior $\omega \in \mathcal{D}(X)$. Applying Jeffrey's rule to a new state of affairs $\rho \in \mathcal{D}(I)$ gives as posterior:*

$$\hat{f}^{\dagger}_{\omega} \gg \rho = \textstyle\sum_i \rho(i) \cdot \omega|_{\mathbf{1}_{U_i}} \qquad \textit{satisfying} \qquad \hat{f} \gg \big(\hat{f}^{\dagger}_{\omega} \gg \rho\big) = \rho. \tag{10}$$

Moreover, wrt. the total variation distance function d one has:

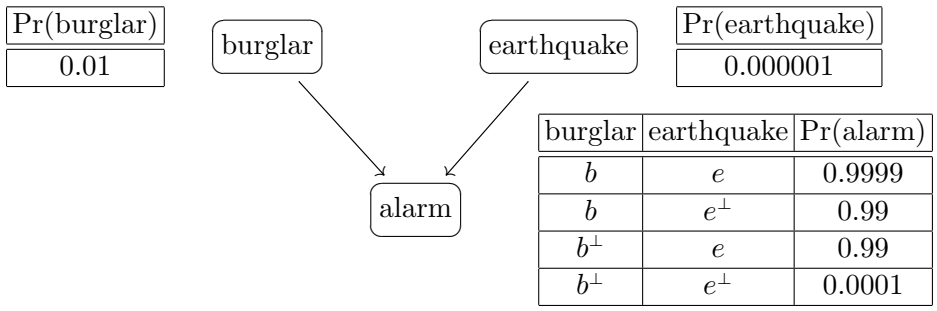
$$d\big(\hat{f}^{\dagger}_{\omega} \gg \rho, \omega\big) = \bigwedge \{d(\omega, \omega') \mid \omega' \in \mathcal{D}(X) \textit{ with } \hat{f} \gg \omega' = \rho\}. \qquad \square$$

The equation on the left in (10) describes Jeffrey's update as a convex combination of updated states $\omega|_{\mathbf{1}_{U_i}}$, conditioned to the partitions U_i, with probabilities $\rho(i)$. The equation on the right in (10) illustrates the 'destructive' character of Jeffrey's rule: the prediction after the update is equal to new situation: the original prediction $\hat{f} \gg \omega$ is simply overridden by ρ. The equations in this lemma hold because $\hat{f}$ is a deterministic channel and do not hold for arbitrary channels. Since many early examples of Jeffrey's rule involve such partitions via deterministic channels, where the effect / predicted state $\hat{f} \gg \omega'$ of the updated state $\omega' = \hat{f}^{\dagger}_{\omega} \gg \rho$ is equal to the uncertain evidence ρ, *i.e.* $\hat{f} \gg \omega' = \rho$, the idea emerged that for Jeffrey's rule the specification of the evidence ρ must happen in terms of the effect $\hat{f} \gg \omega'$. But, as said, this only works for deterministic channels, not in general. We return to this point below, in point 1 in Subsection 6.2.

For a general, not-deterministic channel $c\colon X \nrightarrow Y$ with prior state $\omega \in \mathcal{D}(X)$ and evidence state $\rho \in \mathcal{D}(Y)$ one can prove:

$$d\big(c^{\dagger}_{\omega} \gg \rho, \omega\big) \leq \bigwedge_{\omega' \in \mathcal{D}(X)} d(\omega, \omega') + d(c \gg \omega', \rho).$$

Example 6.3. We turn to the following Bayesian network.



burglar	earthquake	Pr(alarm)
b	e	0.9999
b	$e^{\perp}$	0.99
$b^{\perp}$	e	0.99
$b^{\perp}$	$e^{\perp}$	0.0001

The a prori probabilities of a burglary and an earthquake, given in the upper tables, can be written as probability distributions:

$$\omega = 0.01|b\rangle + 0.99|b^{\perp}\rangle \qquad \text{and} \qquad \sigma = 0.000001|e\rangle + 0.999999|e^{\perp}\rangle.$$

These two states can be combined to a 'joint' product state on the product space $\{b, b^{\perp}\} \times \{e, e^{\perp}\}$, written as:

$$\begin{aligned}\sigma \otimes \omega = {} & 0.00000001|b, e\rangle + 0.00999999|b, e^{\perp}\rangle \\ & + 0.00000099|b^{\perp}, e\rangle + 0.98999901|b^{\perp}, e^{\perp}\rangle.\end{aligned}$$

The conditional probability table for alarm translates in a straightforward manner into a channel c from $\{b, b^\perp\} \times \{e, e^\perp\}$ to $\{a, a^\perp\}$, namely as:

$$\begin{aligned} c(b,e) &= 0.9999|a\rangle + 0.0001|a^\perp\rangle & c(b,e^\perp) &= 0.99|a\rangle + 0.01|a^\perp\rangle \\ c(b^\perp,e) &= 0.99|a\rangle + 0.01|a^\perp\rangle & c(b^\perp,e^\perp) &= 0.0001|a\rangle + 0.9999|a^\perp\rangle. \end{aligned}$$

The following question is asked by Barber (2012, Example 3.1 and 3.2):

> Imagine that we are 70% sure we heard the alarm sounding. What is the probability of a burglary?

Again it is not clear if we should interpret this situation in terms of improvement (Pearl) or correction (Jeffrey). The latter seems more natural since there is no 'surprise' that needs correction. Nevertheless, Barber (2012) uses the former.

For Jeffrey's approach we translate the 70% certainty into a state $\rho = 0.7|a\rangle + 0.3|a^\perp\rangle$. We can take the Bayesian inversion of the channel c wrt. the product state $\sigma \otimes \omega$, giving $c^\dagger_{\sigma\otimes\omega} \colon \{a, a^\perp\} \nrightarrow \{b, b^\perp\} \times \{e, e^\perp\}$. Jeffreys' rule $c^\dagger_{\sigma\otimes\omega} \gg \rho$ thus gives a distribution on $\{b, b^\perp\} \times \{e, e^\perp\}$. Taking its first marginal yields the outcome that is computed by Barber (2012), namely:

$$0.693|b\rangle + 0.307|b^\perp\rangle.$$

For Pearl's approach we translate the 70% certainty into a predicate $p \colon \{a, a^\perp\} \to [0, 1]$ with $p(a) = 0.7$ and $p(a^\perp) = 0.3$. Pearl's rule $(\sigma \otimes \omega)|_{c \ll p}$ also yields a a distribution on $\{b, b^\perp\} \times \{e, e^\perp\}$, whose first marginal is:

$$0.0229|b\rangle + 0.9771|b^\perp\rangle.$$

This outcome is obtained by Jacobs and Zanasi (2019). It differs considerably from the previous one — 69% versus 2% — and demonstrates that it is highly relevant which interpretation — Jeffreys' or Pearl's — is chosen.

Example 6.4. We look at one more illustration, introduced by Dietrich et al. (2016), where we see an interesting combination of Jeffrey's and Pearl's rule. The setting is: Ann must decide about hiring Bob, whose characteristics are described in terms of competence (c or $c^\perp$) and experience (e or $e^\perp$). The prior is a joint distribution on the product space $\{c, c^\perp\} \times \{e, e^\perp\}$ given as:

$$\omega = \tfrac{4}{10}|c, e\rangle + \tfrac{1}{10}|c, e^\perp\rangle + \tfrac{1}{10}|c^\perp, e\rangle + \tfrac{4}{10}|c^\perp, e^\perp\rangle.$$

The first marginal of ω is the uniform distribution $\frac{1}{2}|c\rangle + \frac{1}{2}|c^\perp\rangle$. It is the base rate for Bob's competence.

We use the two projection functions $\{c, c^\perp\} \xleftarrow{\pi_1} \{c, c^\perp\} \times \{e, e^\perp\} \xrightarrow{\pi_2} \{e, e^\perp\}$ as deterministic channels $\hat{\pi}_1$ and $\hat{\pi}_2$.

When Ann would learn that Bob has relevant work experience, given by point evidence $\mathbf{1}_e$, her strategy is to factor this in via Pearl's rule / backward inference: this gives $\omega|_{\hat{\pi}_2 \ll \mathbf{1}_e}$, whose first marginal is $\frac{4}{5}|c\rangle + \frac{1}{5}|c^\perp\rangle$. It is then more likely that Bob is competent.

Ann reads Bob's letter to find out if he actually has relevant experience. We quote Dietrich et al. (2016):

> Bob's answer reveals right from the beginning that his written English is poor. Ann notices this even before figuring out what Bob says about his work experience. In response to this unforeseen learnt input, Ann lowers her probability that Bob is competent from $\frac{1}{2}$ to $\frac{1}{8}$. It is natural to model this as an instance of Jeffrey revision.

Bob's poor English is a new state of affairs — a surprise — which translates to a competence state $\rho = \frac{1}{8}|c\rangle + \frac{7}{8}|c^\perp\rangle$. This is not something that Ann wants to *factor in*; no, she wants to *adjust* to this new situation, so she uses Jeffrey's rule, giving a new joint state:

$$\omega' = (\hat{\pi}_1)^\dagger_\omega \gg \rho = \tfrac{1}{10}|c, e\rangle + \tfrac{1}{40}|c, e^\perp\rangle + \tfrac{7}{40}|c^\perp, e\rangle + \tfrac{7}{10}|c^\perp, e^\perp\rangle.$$

If the letter now tells that Bob has work experience, Ann will factor this in, in this new situation ω', giving $\frac{4}{11}|c\rangle + \frac{7}{11}|c^\perp\rangle$ as first marginal of $\omega'|_{\hat{\pi}_1 \ll \mathbf{1}_e}$. The likelihood of Bob being competent is now lower than in the prior state. This example reconstructs the illustration of Dietrich et al. (2016) in channel-based form, with the associated formulations of Pearl's and Jeffrey's rules, and produces exactly the same outcomes as in *loc. cit.*

6.2 All Things, or Nothing Else, Considered

We now take a closer look at the work of Chan and Darwiche (2005) and Darwiche (2009) where the Jeffrey/Pearl distinction has been described in terms of the way that soft evidence is described.

1. In this context, Jeffrey's rule is called "all things considered" by Goldszmidt and Pearl (1996); briefly, quoting Darwiche (2009, §3.6.1): "One method for describing soft evidence on event β is by stating the new belief in β after the evidence has been accomodated."

   We make this more concrete in terms of a probability distribution $\omega \in \mathcal{D}(X)$ which is somehow updated to a distribution $\omega' \in \mathcal{D}(X)$. There is an event $E \subseteq X$ whose "strength" is given as its validity $q \in [0, 1]$ in the updated state, that is $q = \omega' \models \mathbf{1}_E$. This validity in the updated state is thus the way in which softness is specified. This makes sense from Jeffrey's perspective, since it involves adjustment/correction. It is a rather indirect, *post hoc* way of specifying, but it can be done like this.

   We elaborate this situation in the current framework, using the partition-based special case of Lemma 6.2. The event $E \subseteq X$ forms a two-element partition of X, consisting of E and its complement $\neg E$, so we take as index set $I = \{1, 2\}$ with function $f\colon X \to I$ given by $f(x) = 1$ if $x \in E$ and $f(x) = 2$ if $x \notin E$. The validity q can be understood as a 'state of affairs' distribution $\sigma = q|1\rangle + (1 - q)|2\rangle$ on the index set $I = \{1, 2\}$. Then, following the formula in (10) for Jeffrey's updating with a partition, we get convex combination:

$$\omega' = \hat{f}^\dagger_\omega \gg \sigma = q \cdot \omega|_{\mathbf{1}_E} + (1 - q) \cdot \omega|_{\mathbf{1}_{\neg E}}.$$

By elaborating the definition of conditioning (7) we get:

$$\omega'(x) = \begin{cases} \dfrac{q \cdot \omega(x)}{\omega \models \mathbf{1}_E} & \text{if } x \in E \\ \dfrac{(1-q) \cdot \omega(x)}{\omega \models \mathbf{1}_{\neg E}} & \text{if } x \in \neg E \end{cases}$$

This is precisely Eqn. (3.20) of Darwiche (2009).

2. Pearl's rule is called "nothing else considered". The strength is now given by a "Bayes factor" $k > 0$. Skipping many details, we can turn this factor k into a predicate $p\colon I \to [0,1]$ on the index set $I = \{1,2\}$, with $p(1) = r$ and $p(2) = r/k$. The number r is some scaling factor that ensures that p's values are in the unit interval $[0,1]$. It drops out in updating, see Lemma 4.1 (3).

   We elaborate the technicalities of Pearl's approach, using the above partition $E, \neg E$ of X over $I = \{1,2\}$. Then we compute Pearl's update $\omega|_{\hat{f} \ll p}$ step-by-step:

$$\begin{aligned} \big(\hat{f} \ll p\big)(x) = p(f(x)) &= \begin{cases} r & \text{if } x \in E \\ r/k & \text{if } x \notin E \end{cases} \\ \omega \models \hat{f} \ll p &= \textstyle\sum_{x \in E} r \cdot \omega(x) + \sum_{x \notin E} r/k \cdot \omega(x) \\ &= r \cdot (\omega \models \mathbf{1}_E) + r/k \cdot (\omega \models \mathbf{1}_{\neg E}) \\ \omega|_{\hat{f} \ll p} &= \sum_{x \in E} \frac{r \cdot (\omega \models \mathbf{1}_E)}{r \cdot (\omega \models \mathbf{1}_E) + r/k \cdot (\omega \models \mathbf{1}_{\neg E})} \big| x \big\rangle \\ &\qquad + \sum_{x \notin E} \frac{r/k \cdot (\omega \models \mathbf{1}_{\neg E})}{r \cdot (\omega \models \mathbf{1}_E) + r/k \cdot (\omega \models \mathbf{1}_{\neg E})} \big| x \big\rangle. \end{aligned}$$

   We can rewrite the latter formal convex sum as probability mass function:

$$\big(\omega|_{\hat{f} \ll p}\big)(x) = \begin{cases} \dfrac{k \cdot \omega(x)}{k \cdot (\omega \models \mathbf{1}_E) + (\omega \models \mathbf{1}_{\neg E})} & \text{if } x \in E \\ \dfrac{\omega(x)}{k \cdot (\omega \models \mathbf{1}_E) + (\omega \models \mathbf{1}_{\neg E})} & \text{if } x \in \neg E \end{cases}$$

   This is Eqn. (3.25) of Darwiche (2009).

We conclude that, even though the approaches "all things considered" and "nothing else considered" take a completely different route to specifying softness, they still fit in the current general setting.

7. Concluding Remarks

This paper uses hard maths for soft evidence. It provides a systematic account of two different forms of probabilistic updating with soft evidence, namely Jeffrey's rule and Pearl's

method. These two approaches are provided with informal conceptualisations, like: adjusting/adapting to, correction (Jeffrey style), and: factoring in, improvement (Pearl style). The paper's technical contribution lies in providing a mathematically precise formulation of Jeffrey's and Pearl's updating, systematically using the concept of channel. This makes it possible to reformulate several results from the literature, notably of Chan and Darwiche (2005) and of Darwiche (2009), to add new results, and to describe various (confusing) examples from a uniform perspective.

In the end we briefly suggest a connection between the channel-based formalism and the cognitive explanation of perception by Hohwy (2013). We formalise it as follows: a consistent, relevant portion of the human mind may be represented by a probability distribution σ, forming the internal state at hand. We use a channel c to translate this internal state into predictions $c \gg \sigma$ about the outside world. The confrontation of this prediction with observation leads to an update of the internal state σ. In the setting of this paper, the update may happen using Jeffrey's approach, when σ is adjusted/corrected to $c^{\dagger}_{\sigma} \gg \rho$ for an observed external state ρ. It may also happen according to Pearl, so that σ is improved to $\sigma|_{c \ll p}$ for external evidence p that is factored in. It remains an intriguing open question, far beyond the scope of this paper, if this Jeffrey/Pearl distinction between correcting and improving makes cognitive sense.

Finally, a question that might arise is whether Jeffrey's/Pearl's updating can also be described (and distinguished) in continuous probability. The answer is yes. Pearl's updating is essentially conditioning and can be done with continuous probability, see *e.g.* Jacobs (2018b). Jeffrey's approach involves disintegration (or Bayesian inversion), which is a rather subtle topic in a continuous setting, as illustrated by Clerc et al. (2017) (and the references there) for more information: daggers of channels may not exist, or may not be determined uniquely (up to null-sets).

Acknowledgements

Thanks to the anonymous reviewers for their constructive feedback.

References

Barber, D. (2012). *Bayesian Reasoning and Machine Learning.* Cambridge Univ. Press. Publicly available via `http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage`.

Bernardo, J., & Smith, A. (2000). *Bayesian Theory.* John Wiley & Sons.

Chan, H., & Darwiche, A. (2005). On the revision of probabilistic beliefs using uncertain evidence. *Artif. Intelligence*, *163*, 67–90.

Cho, K., & Jacobs, B. (2019). Disintegration and Bayesian inversion via string diagrams. *Math. Struct. in Comp. Sci.*, *29(7)*, 938–971.

Clerc, F., Dahlqvist, F., Danos, V., & Garnier, I. (2017). Pointless learning. In Esparza, J., & Murawski, A. (Eds.), *Foundations of Software Science and Computation Structures*, No. 10203 in Lect. Notes Comp. Sci., pp. 355–369. Springer, Berlin.

Coecke, B., & Kissinger, A. (2016). *Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge Univ. Press.

Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Networks*. Cambridge Univ. Press.

Diaconis, P., & Zabell, S. (1982). Updating subjective probability. *Journ. American Statistical Assoc.*, *77*, 822–830.

Diaconis, P., & Zabell, S. (1983). Some alternatives to Bayes' rule. Tech. rep. 339, Stanford Univ., Dept. of Statistics.

Dietrich, F., List, C., & Bradley, R. (2016). Belief revision generalized: A joint characterization of Bayes' and Jeffrey's rules. *Journ. of Economic Theory*, *162*, 352–371.

Dijkstra, E., & Scholten, C. (1990). *Predicate Calculus and Program Semantics*. Springer, Berlin.

Fong, B. (2012). Causal theories: A categorical perspective on Bayesian networks. Master's thesis, Univ. of Oxford. see `arxiv.org/abs/1301.6201`.

Giry, M. (1982). A categorical approach to probability theory. In Banaschewski, B. (Ed.), *Categorical Aspects of Topology and Analysis*, No. 915 in Lect. Notes Math., pp. 68–85. Springer, Berlin.

Goldszmidt, M., & Pearl, J. (1996). Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artif. Intelligence*, *84(1-2)*, 57–112.

Griffiths, T., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In Sun, R. (Ed.), *Cambridge Handbook of Computational Cognitive Modeling*, pp. 59–100. Cambridge Univ. Press.

Halpern, J. (2003). *Reasoning about Uncertainty*. MIT Press, Cambridge, MA.

Hohwy, J. (2013). *The Predictive Mind*. Oxford Univ. Press.

Jacobs, B. (2015). New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, *11(3)*. See `https://lmcs.episciences.org/1600`.

Jacobs, B. (2017). A recipe for state and effect triangles. *Logical Methods in Comp. Sci.*, *13(2)*. See `https://lmcs.episciences.org/3660`.

Jacobs, B. (2018a). A channel-based exact inference algorithm for Bayesian networks. See `arxiv.org/abs/1804.08032`.

Jacobs, B. (2018b). From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, *94*, 200–237.

Jacobs, B. (2019). Learning along a channel: the Expectation part of Expectation-Maximisation. In König, B. (Ed.), *Math. Found. of Programming Semantics*, Elect. Notes in Theor. Comp. Sci. Elsevier, Amsterdam. To appear.

Jacobs, B., & Zanasi, F. (2016). A predicate/state transformer semantics for Bayesian learning. In Birkedal, L. (Ed.), *Math. Found. of Programming Semantics*, No. 325 in Elect. Notes in Theor. Comp. Sci., pp. 185–200. Elsevier, Amsterdam.

Jacobs, B., & Zanasi, F. (2019). The logical essentials of Bayesian reasoning. In Barthe, G., Katoen, J.-P., & Silva, A. (Eds.), *Probabilistic Programming*. Cambridge Univ. Press. See `arxiv.org/abs/1804.01193`, to appear.

Jeffrey, R. (1983). *The Logic of Decision* (2nd rev. edition). The Univ. of Chicago Press.

Jensen, F., & Nielsen, T. (2007). *Bayesian Networks and Decision Graphs* (2nd rev. edition). Statistics for Engineering and Information Science. Springer.

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models. Principles and Techniques*. MIT Press, Cambridge, MA.

Kozen, D. (1985). A probabilistic PDL. *Journ. Comp. Syst. Sci*, *30(2)*, 162–178.

Mrad, A., Delcroix, V., Piechowiak, S., Leicester, P., & Abid, M. (2015). An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, *23(4)*, 802–824.

Panangaden, P. (2009). *Labelled Markov Processes*. Imperial College Press, London.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Graduate Texts in Mathematics 118. Morgan Kaufmann.

Pearl, J. (1990). Jeffrey's rule, passage of experience, and neo-Bayesianism. In H. Kyburg, J. (Ed.), *Knowledge Representation and Defeasible Reasoning*, pp. 245–265. Kluwer Acad. Publishers.

Shafer, G. (1981). Jeffrey's rule of conditioning. *Philosophy of Science*, *48(3)*, 337–362.

Skyrms, B. (1996). The structure of radical probabilism. *Erkenntnis*, *35*, 439–60.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgement under uncertainty: Heuristics and biases*, pp. 153–160. Cambridge Univ. Press.

Valtorta, M., Kim, Y.-G., & Vomlel, J. (2002). Soft evidential update for probabilistic multiagent systems. *Int. Journ. of Approximate Reasoning*, *29(1)*, 71–106.

Zadeh, L. (1968). Probability measures of fuzzy events. *Journ. Math. Analysis and Appl.*, *23(2)*, 421–427.