A Formal Semantics of Influence in Bayesian Reasoning

Bart Jacobs[1] and Fabio Zanasi[2]

1 Radboud Universiteit, Nijmegen, The Netherlands
2 University College London, London, United Kingdom

Abstract

This paper proposes a formal definition of influence in Bayesian reasoning, based on the notions of state (as probability distribution), predicate, validity and conditioning. Our approach highlights how conditioning a joint entwined/entangled state with a predicate on one of its components has 'crossover' influence on the other components. We use the total variation metric on probability distributions to quantitatively measure such influence. These insights are applied to give a rigorous explanation of the fundamental concept of d-separation in Bayesian networks.

1998 ACM Subject Classification Probabilistic computation, F. 1.2

Digital Object Identifier 10.4230/LIPIcs.MFCS.2017.21

1 Introduction

A key feature of Bayesian (probabilistic) reasoning is that an observation leads to an update of knowledge. This is best seen in Bayesian networks: in these graph-like models, dependency relations between events are visually depicted as arcs between nodes. Information about a node-event A will update knowledge of all the nodes connected by an arc to A. However, influence may act also in more indirect ways, classified by Pearl [13] as the following "d-separation" scenarios:

(i) in a **serial** connection $\boxed{\mathrm{A}} \to \boxed{\mathrm{B}} \to \boxed{\mathrm{C}}$, event A influences C through B (and viceversa), but knowledge of B "blocks" this mutual influence — one also says that B *d-separates* A and C.

(ii) in a **fork** connection $\boxed{\mathrm{A}} \leftarrow \boxed{\mathrm{B}} \to \boxed{\mathrm{C}}$, information on A will influence C and viceversa, but this flow is blocked once B is known.

(iii) in a **collider** situation $\boxed{\mathrm{A}} \to \boxed{\mathrm{B}} \leftarrow \boxed{\mathrm{C}}$, any evidence about B (and its descendants) will make A and C depend on each other.

In these three scenarios one may observe many phenomena at work which are usually explained informally in terms of influence, dependence, blocking and evidence. But what is the formal semantics underpinning these concepts? The basic language of conditional probability, based on the reading of $\Pr(A|B)$ as "the probability of A given B", appears to be unsuitable for such an account. For instance, it cannot express that, in the collider situation, *any* evidence on the occurrence of B will make A and C dependent, whereas the blocking of the first two scenarios only occurs when B is known with certainty (probability 1).

This paper proposes a rigorous formal treatment of influence in Bayesian reasoning, yielding an expressive and firmly established language for describing the above scenarios. Our methodology draws inspiration from the area of programming language semantics, and in particular from *Effectus theory* [4, 2], a comprehensive logical framework for probabilistic and quantum computation. At the foundation of our approach there is a conceptual distinction between the knowledge of an event, called a *state*, and an observation/evidence of such event, called a *predicate*. Concretely, a state on a 'sample' space X will be a (finite) discrete probability distributions ω on X, whereas a (fuzzy) predicate p on X is a function

$X \to [0,1]$. The 'knowledge update' is then given by a conditioned distribution, which we write as $\omega|_p$, pronounced as: ω given p. Moreover, our approach includes predicate and state transformers, adding expressive power to the language.

Our first contribution (§ 3) is a semantic description of d-separation in the serial (i) and fork connection (ii): we reduce these scenarios into formal statements, whose proofs are made straightforward by our formalism. Here the phenomenon at stake is influence *blocking*, for which the basic language of states and predicates suffices. However, the collider scenario (iii), in which influence is not blocked but rather *enabled*, demands a deeper analysis.

This leads to our second contribution (§ 4), namely the concept of *state entwinedness*. Intuitively, for a joint state/distribution being entwined means, by analogy with the quantum world, that its components are entangled or, in the Bayesian jargon, they model dependent events. In order to capture the collider situation, the key observation is that the join (tensor product) of non-entwined states (say, in (iii), the join of A and C) may become entwined after conditioning (information about B); from that moment on, any new information on one component of the joint state will have influence also on the other component.

As a third contribution (§ 5), we introduce a formal, quantitative definition of such influence: we call it *crossover influence*, as it measures the non-local action between components of a joint states. We also define a notion of *direct influence*, which measures the local action of a predicate (an information update) on a certain state. Both definitions take as a parameter a notion of 'distance' between states: for our scenarios we pick the *total variation* metric on probability distributions, which coincides with the Kantorovich metric [17] on discrete metric spaces (sets). We make no claim on total variation being 'canonical' in the sense of [10]. Our emphasis is rather on the abstract definition of influence: this is independent of the choice of the underlying metric, which is not itself an essential part of our analysis, see also § 7. As far as we know, probabilistic influence has not been formalised and investigated in this quantitative form before.

We conclude our developments with a reprise of the collider scenario (§ 6), which we are now able to adequately describe using the toolkit introduced in § 4 and § 5. Our analysis clarifies that the commonly used description in the literature (see *e.g.* [14, 8, 15]) for describing the serial (i) and fork (ii) scenarios only works for very special 'singleton' predicates — which we call Dirac predicates, whereas in the collider scenario (iii) *any* predicate on B creates dependence (entwinedness) between A and C.

2 Background: states, predicates, and conditional probability

In this background section we introduce the notation, terminology and basic definitions for several constructions in (finite) discrete probability. There is a categorical formalisation using monads behind this, see e.g. [5], but we prefer to keep constructions more concrete.

States, predicates and validity. A (finite, discrete) distribution over a 'sample' set A is a weighted combination of elements of A, where weights are probabilities from the unit interval $[0,1]$ that add up to 1. We call such a distribution a *state*, as it expresses knowledge the occurrence of elements of A. As mentioned in §1, we pursue an analogy with quantum states, emphasised by the use of the 'ket' notation: a state ω is written as $\omega = r_1 |a_1\rangle + \cdots + r_n |a_n\rangle$, where $a_i \in A$, $r_i \in [0,1]$ and $\sum_i r_i = 1$. Also, $\mathcal{D}(A)$ is the set of states/distributions on A. We will sometimes treat $\omega \in \mathcal{D}(A)$ equivalently as a function $\omega\colon A \to [0,1]$ with finite support $\mathrm{supp}(\omega) = \{a \in A \mid \omega(a) \neq 0\}$ and with $\sum_{a\in A} \omega(a) = 1$.

An *event* is a subset $E \subseteq A$ of the sample space. We prefer to use a more general 'fuzzy' kind of predicate, namely functions $p\colon A \to [0,1]$. In this discrete case, states (distributions) are predicates, but not the other way around. Events can be identified with 'sharp' predicates taking values in the subset of booleans $\{0,1\} \subseteq [0,1]$. For $x \in A$, we write ∂_x for the (sharp) *Dirac* predicate over x, defined as $\partial_x(a) = 1$ if $x = a$ and $\partial_x(a) = 0$ otherwise.

For predicates $p, q \in [0,1]^A$ and scalar $r \in [0,1]$ we define $p \,\&\, q$ as $a \mapsto p(a) \cdot q(a)$ and $r \cdot p$ as $a \mapsto r \cdot p(a)$. States and predicates are most effectively reasoned about using the language of *Kleisli categories*. We call a function of shape $f\colon A \to \mathcal{D}(B)$ a 'Kleisli' map from A to B and write its type as $A \mathrel{\circ\!\!\!\rightarrow} B$. Kleisli maps can be understood as *channels*, or as *stochastic matrices*, especially when A, B are finite sets. The (Kleisli) composition of maps $f\colon A \mathrel{\circ\!\!\!\rightarrow} B$ and $g\colon B \mathrel{\circ\!\!\!\rightarrow} C$ is written as $g \bullet f\colon A \mathrel{\circ\!\!\!\rightarrow} C$. It is essentially matrix multiplication:

$$\big(g \bullet f\big)(a) \;=\; \sum_{c \in C} \big(\textstyle\sum_{b \in B} f(a)(b) \cdot g(b)(c)\big)\big|\, c\,\big\rangle. \tag{1}$$

We write $\mathcal{K}\ell(\mathcal{D})$ for the Kleisli category whose objects are sets, and whose arrows from A to B are the Kleisli maps $A \mathrel{\circ\!\!\!\rightarrow} B$. The identity map $A \mathrel{\circ\!\!\!\rightarrow} A$ in $\mathcal{K}\ell(\mathcal{D})$ is the function $a \mapsto 1|a\rangle$. Note that arrows $1 \mathrel{\circ\!\!\!\rightarrow} B$ in $\mathcal{K}\ell(\mathcal{D})$ identify elements of $\mathcal{D}(B)$, *i.e.* the states on B, and arrows $B \mathrel{\circ\!\!\!\rightarrow} 2$ are elements of $[0,1]^B$, *i.e.* the predicates on B.

Each (ordinary) function $g\colon A \to B$ gives a trivial (diagonal) matrix map $\langle g \rangle\colon A \mathrel{\circ\!\!\!\rightarrow} B$ via $\langle g\rangle(a) = 1|g(a)\rangle$. Then: $\langle h \rangle \bullet \langle g \rangle = \langle h \circ g \rangle$.

We will see later, in Example 3, how Bayesian networks can be seen as graphs of Kleisli maps in $\mathcal{K}\ell(\mathcal{D})$. For this interpretation, it is of importance that $\mathcal{K}\ell(\mathcal{D})$ forms a monoidal category. The monoidal product $\otimes$ is defined on objects as the cartesian product $\times$ of sets, with tensor unit the one-element set 1. On Kleisli maps $f\colon A \mathrel{\circ\!\!\!\rightarrow} X$ and $g\colon B \mathrel{\circ\!\!\!\rightarrow} Y$ the map $f \otimes g\colon A \otimes B \mathrel{\circ\!\!\!\rightarrow} X \otimes Y$ is defined as $(f \otimes g)(a,b)(x,y) = f(a)(x) \cdot g(b)(y)$.

▶ **Definition 1.** Let $\omega \in \mathcal{D}(A)$ be a state and $p \in [0,1]^A$ be a predicate, both on the same set A. We write $\omega \models p$ for the *validity* or *expected value* of p in state ω. This validity is a number in the unit interval $[0,1]$ defined as:

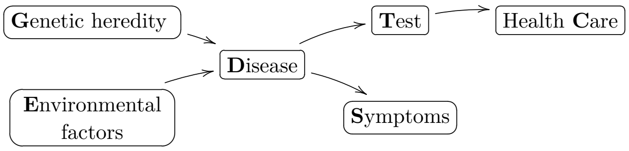
$$\omega \models p \;:=\; \textstyle\sum_{a \in A} \omega(a) \cdot p(a) \;=\; \big(A \overset{p}{\mathrel{\circ\!\!\!\rightarrow}} 2\big) \bullet \big(1 \overset{\omega}{\mathrel{\circ\!\!\!\rightarrow}} A\big). \tag{2}$$

If this validity is non-zero, it yields a *conditioning* operation on ω. We write $\omega|_p$ or for the conditional state "ω given p", defined as formal convex sum:

$$\omega|_p \;:=\; \sum_{a \in A} \frac{\omega(a) \cdot p(a)}{\omega \models p}\big|\, a\,\big\rangle. \tag{3}$$

▶ **Lemma 2** (From [5]). *(a)* $p \,\&\, \partial_x = p(x) \cdot \partial_x$ *and* $\omega \models \partial_x = \omega(x)$ *and* $\omega|_{\partial_x} = 1|x\rangle$*;*
(b) $\omega|_{r \cdot p} = \omega|_p$ *for* $r \neq 0$ *and* $\omega|_{p \& \partial_x} = 1|x\rangle$ *when* $p(x) \neq 0$ *and* $\omega(x) \neq 0$*;*
(c) Bayes' rule holds for fuzzy predicates: $\omega|_p \models q \;=\; \dfrac{\omega \models p \,\&\, q}{\omega \models p}$.

▶ **Example 3.** As a running example we will use the situation of a disease that can be caused by environmental factors or by genetic heredity. The presence of the disease in a patient will determine whether she manifests symptoms and also whether she tests positively. The test outcome will also influence whether she receives health care. We express these data with a Bayesian network, consisting of a graph together with conditional probability tables.



$\Pr(G) = 1/50$
$\Pr(E) = 1/10$

D	$\Pr(T)$
t	$9/10$
f	$1/20$

G	E	$\Pr(D)$
t	t	$9/10$
t	f	$8/10$
f	t	$4/10$
f	f	0

D	$\Pr(S)$
t	$9/10$
f	$1/15$

T	$\Pr(C)$
t	$4/5$
f	$1/10$

(4)

As illustrated in [7] (*cf.* also [3]), there is a canonical way to interpret our Bayesian network (4) as an arrow in the Kleisli category $\mathcal{K}\ell(\mathcal{D})$. Each node N of the graph, say with k incoming edges from nodes $N_1, N_2, \ldots, N_k$, is associated with an arrow $N\colon 2^k \mathrel{\circ\!\!\rightarrow} 2$ in $\mathcal{K}\ell(\mathcal{D})$; as a stochastic matrix, N is defined by the probability table of the node N. It will be convenient to write $2_N := \{n, n^\perp\}$ for the two-element target set of the node-arrow N, where n represents occurrence and $n^\perp$ non-occurrence of the event N. For instance, the arrow $D\colon 2_G \otimes 2_E \mathrel{\circ\!\!\rightarrow} 2_D$ for the disease node is defined by the channel $2_G \times 2_E \to \mathcal{D}(2_D)$

$$\begin{aligned} (g,e) &\mapsto \tfrac{9}{10}\,|d\rangle + \tfrac{1}{10}\,|d^\perp\rangle & (g,e^\perp) &\mapsto \tfrac{8}{10}\,|d\rangle + \tfrac{2}{10}\,|d^\perp\rangle \\ (g^\perp,e) &\mapsto \tfrac{4}{10}\,|d\rangle + \tfrac{6}{10}\,|d^\perp\rangle & (g^\perp,e^\perp) &\mapsto 1\,|d^\perp\rangle\,. \end{aligned}$$

Another example is the initial map $G\colon 1 \mathrel{\circ\!\!\rightarrow} 2_G$ for the genetic heredity node, which amounts to the distribution $1/50\,|g\rangle + 49/50\,|g^\perp\rangle$ in $\mathcal{D}(2_G) \cong [0,1]$. In order to recover the whole network (4), one pastes node-arrows together using the monoidal structure of $\mathcal{K}\ell(\mathcal{D})$. Nodes in (4) that have multiple outgoing edges are modeled by composing the corresponding arrow $2^k \mathrel{\circ\!\!\rightarrow} 2$ with the pairing map $\Delta\colon 2 \mathrel{\circ\!\!\rightarrow} 2 \otimes 2$ defined by $x \mapsto 1\,|(x,x)\rangle$. The Bayesian network (4) in its entirety is then expressed as the following arrow in $\mathcal{K}\ell(\mathcal{D})$.

$$\begin{tikzcd} 1 \arrow[r, "G\otimes E", "\circ" marking] & 2_G \otimes 2_E \arrow[r, "D", "\circ" marking] & 2_D \arrow[r, "\Delta", "\circ" marking] & 2_D \otimes 2_D \arrow[r, "T\otimes S", "\circ" marking] & 2_T \otimes 2_S \arrow[r, "C\otimes \mathrm{id}", "\circ" marking] & 2_C \otimes 2_S \end{tikzcd} \tag{5}$$

Inference via predicate/state transformers. Associated with a Kleisli map $f\colon A \mathrel{\circ\!\!\rightarrow} B$ there are *state transformer* and *predicate transformer* maps f_* and f^*. For a state $\omega \in \mathcal{D}(A)$ and a predicate $p \in [0,1]^B$ we define $f_*(\omega) \in \mathcal{D}(B)$ and $f^*(p) \in [0,1]^A$ as:

$$f_*(\omega) \;=\; \sum_{b\in B} \big(\textstyle\sum_{a\in A} f(a)(b)\cdot\omega(a)\big)\big|\,b\big\rangle \qquad\qquad f^*(p)(a) \;=\; \sum_{b\in B} f(a)(b)\cdot p(b). \tag{6}$$

Notice that f_* works forwardly, transforming a state on A into a state on B, whereas f^* works backwardly, transforming a predicate on B into a predicate on A. One can understand these definitions in terms of Kleisli composition: $f_*(\omega) = f \bullet \omega$ and $f^*(p) = p \bullet f$. We collect a few basic results from [5].

▶ **Lemma 4.** *(a) For a Kleisli map $f\colon A \mathrel{\circ\!\!\rightarrow} B$, a state $\omega \in \mathcal{D}(A)$ and a predicate $p \in [0,1]^B$,*

$$f_*(\omega) \models p \;=\; p \bullet f \bullet \omega \;=\; \omega \models f^*(p).$$

(b) Predicate transformers f^* preserve $\mathbf{1}$, $\mathbf{0}$, *negation $(-)^\perp$ and scalar multiplication $r\cdot(-)$.*

(c) For an ordinary function $g\colon A \to B$ we have $\langle g\rangle_*(\omega)|_p = \langle g\rangle_*(\omega|_{\langle g\rangle^*(p)})$. □

Using transformers and conditioning one can formulate Bayesian inference (learning). We illustrate the relevant constructions with an example and refer to [7] for more details.

▶ **Example 5. Backward inference.** A typical learning task wrt. a Bayesian network is backward inference: how the occurrence of a certain event changes the likelihood of its causes. A formalisation of backward inference is proposed in [7] as "predicate transformation followed by conditioning". We illustrate this for Example 3, focusing on the part of the graph that describes the influence of having the disease on receiving health care. First, we compute our *a priori* knowledge on the likelihood of a disease. In the formalisation (5), this is the Kleisli arrow $D \bullet (G \otimes E)\colon 1 \mathrel{\circ\!\!\rightarrow} 2_D$, *i.e.* a state on 2_D.

$$\begin{aligned} G \otimes E \;&=\; 0.002\,|g,e\rangle + 0.018\,|g,e^\perp\rangle + 0.098\,|g^\perp,e\rangle + 0.882\,|g^\perp,e^\perp\rangle \\ D \bullet (G \otimes E) \;&=\; 0.055\,|d\rangle + 0.945\,|d^\perp\rangle \end{aligned} \tag{7}$$

The event of a positive test is interpreted as the Dirac predicate $\partial_t \in [0,1]^{2_T}$ on 2_T, *i.e.* it maps t to 1 and $t^\perp$ to 0. We can now ask a backward inference question: if the patient tested positive, what is the likelihood that she had the disease? The answer is enclosed in the state $(D \bullet (G\otimes E))|_{T^*(\partial_t)} : 1 \mathrel{\circ\!\!\rightarrow} 2_D$, obtained by first using $T : 2_D \mathrel{\circ\!\!\rightarrow} 2_T$ to transform the predicate ∂_t on 2_T into a predicate $T^*(\partial_t)$ on 2_D, and then conditioning the state $D \bullet (G \otimes E)$ over $T^*(\partial_t)$. The latter predicate maps d to $^9\!/\!_{10}$ and $d^\perp$ to $^1\!/\!_{20}$. Next,

$$\begin{aligned} D \bullet (G \otimes E) \models T^*(\partial_t) &= 0.097\,|d\rangle + 0.903\,|d^\perp\rangle \\ (D \bullet (G \otimes E))|_{T^*(\partial_t)} &= \textstyle\sum_{x\in 2_D} \frac{D\bullet(G\otimes E)(x)\cdot T^*(\partial_t)(x)}{D\bullet(G\otimes E)\models T^*(\partial_t)}\,|x\rangle \\ &= \tfrac{0.055\cdot {}^9\!/\!_{10}}{0.097}\,|d\rangle + \tfrac{0.945\cdot {}^1\!/\!_{20}}{0.097}\,|d^\perp\rangle = 0.51\,|d\rangle + 0.49\,|d^\perp\rangle\,. \end{aligned}$$

Thus evidence of a positive test raises the chances of a disease from 0.055 to 0.51.

Forward inference. A second kind of learning task is *forward inference*: how the occurrence of an event changes the likelihood of its effects. Again following [7], forward inference is formalised as "conditioning and then state transformation". To illustrate this in our leading example, consider a predicate p on 2_G given by $g \mapsto 88\%$ and $g^\perp \mapsto 0.1\%$: it expresses that medical records of a patient show high likelihood of a genetic transmission of the disease. Our forward inference question is: "how does the knowledge update given by predicate p influence the positivity of the test?" For the answer, one first extends p to a (weakened) predicate p' on $2_G \otimes 2_E$, then conditions $G \otimes E$ over p'. Finally, one applies $T \bullet D$ as state transformer to $(G \otimes E)|_{p'}$. Conditioning over p' makes a positive test much more likely:

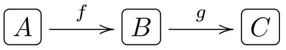
$$(T \bullet D)_*(G \otimes E) = 0.1\,|t\rangle + 0.9\,|t^\perp\rangle \qquad (T \bullet D)_*((G \otimes E)|_{p'}) = 0.505\,|t\rangle + 0.495\,|t^\perp\rangle\,.$$

3 Influence in d-separation

This section applies the language introduced in § 2 to give a precise explanation of the fundamental concept of 'd-separation' in Bayesian networks, which is used as a criterion for independence, via connections between nodes. These connections can be of three forms, namely 'serial', 'fork', and 'collider'. As we shall see, the language introduced so far is only adapted to describe the first two scenarios. The third scenario needs a richer formalism, which justifies the developments in the next sections.

3.1 Serial connections

Consider a 'serial connection' Bayesian network as on the right. Clearly, what we know about A influences our knowledge about C, and vice-versa. In the context of d-separation one considers the special cases when there is evidence about the state of B, so that the mutual influencing between A and B is blocked. We first quote how this is formulated in standard references (names of the nodes in the second quote are adapted to make them consistent with diagram (8)).



(8)

(I) [8, §1.2]: Obviously, evidence on A will influence the certainty of B, which then influences the certainty of C. Similarly, evidence on C will influence the certainty on A through B. On the other hand, if the state of B is known, then the channel is blocked, and A and C become independent.

(II) [14, §1.2.3]: Figuratively, conditioning on B appears to "block" the flow of information along the path, since learning about A has no effect on the probability of C, given B.

These descriptions are rather informal. (I) speaks about (mutual) independence, and (II) only about having no effect in the forward direction. We will make precise what is going on. Consider the same diagram (8), but now with f, g interpreted as maps in the Kleisli category $\mathcal{K}\ell(\mathcal{D})$ and with predicates as on the right. The three predicates are inhabitants $p \in [0,1]^A$, $\partial_x \in [0,1]^B$, $q \in [0,1]^C$.

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, dotted, "p"'] & B \arrow[r, "g"] \arrow[d, dotted, "\partial_x"'] & C \arrow[d, dotted, "q"] \\ 2 & 2 & 2 \end{tikzcd} \tag{9}$$

▶ **Proposition 6 (Blocking I).** *Consider the serial connection* (9)*, with Dirac evidence* ∂_x *on the middle node* B*, for some fixed* $x \in B$*. Then there is no influence from* A *to* C*, nor from* C *to* A*, in the sense that for each distribution/state* $\omega \in \mathcal{D}(A)$*,*

(a) for any predicate p *on* A *with* $\omega \models p \neq 0$*, there is an equality of states on* C*:*

$$g_*\big(f_*(\omega)|_{\partial_x}\big) \;=\; g_*\big(f_*(\omega|_p)|_{\partial_x}\big).$$

(b) for any predicate q *on* C *there is an equality of states on* A*:*

$$\omega|_{f^*(\partial_x)} \;=\; \omega|_{f^*(\partial_x \& g^*(q))}.$$

We recall how to read the equation in point (a): given a state ω on A, we can transform it to a state $f_*(\omega)$ on B. We can also first condition ω to $\omega|_p$ and then push forward to $f_*(\omega|_p)$ on B. These different states $f_*(\omega)$ and $f_*(\omega|_p)$ become equal when we condition with the Dirac predicate ∂_x, and then push them forward to C via g_*. Thus, the influence of p is 'annihilated' or 'blocked' via the knowledge $x \in B$ used in conditioning with ∂_x.

Proof For the first point it suffices to prove $f_*(\omega)|_{\partial_x} = f_*(\omega|_p)|_{\partial_x}$. But this equation follows directly from Lemma 2 (a) since both sides are equal to $1\,|x\rangle$.

For the second point we have $f^*(\partial_x \;\&\; g^*(q)) = f^*(g^*(q)(x) \cdot \partial_x)$ by Lemma 2(a), which is then equal to $g^*(q)(x) \cdot f^*(\partial_x)$ by Lemma 4(b). Finally, by Lemma 2(b), $\omega|_{f^*(\partial_x \& g^*(q))} = \omega|_{g^*(q)(x) \cdot f^*(\partial_x)} = \omega|_{f^*(\partial_x)}$.

▶ **Example 7.** Nodes 'Disease', 'Test' and 'Health Care' in the network of Example 3 form a serial connection, with Kleisli interpretation given by solid arrows as in (10) below. Clearly, new information about the Disease will impact the likelihood of receiving Health Care, and viceversa, via the intermediate Test node. We examined these phenomena as forward and backward inference in Example 5, following [7]. We now show that, as prescribed by d-separation, mutual influence may be blocked: a positive test will determine the availability of health care, disregarding whether the patient actually has the disease or not. Viceversa, a positive test will nullify any influence of receiving health care on having the disease, as health care is entirely determined by the test outcome. The dotted arrows in (10) describe a state $\omega = \frac{1}{100}\,|d\rangle + \frac{99}{100}\,|d^\perp\rangle$ on 2_D, giving a 1% disease probability, and the Dirac predicate $\partial_t \in [0,1]^{2_T}$, asserting the positivity of the test. For the transformed predicate $T^*(\partial_x)$ on 2_D we have:

$$\begin{tikzcd} 2_D \arrow[r, "T"] & 2_T \arrow[r, "C"] \arrow[d, dotted, "\partial_t"] & 2_C \\ 1 \arrow[u, dotted, "\omega"] & 2 & \end{tikzcd} \tag{10}$$

$$\begin{cases} T^*(\partial_t)(d) \;=\; \frac{9}{10} \\ T^*(\partial_t)(d^\perp) \;=\; \frac{1}{20} \end{cases} \qquad \omega \models T^*(\partial_t) \;=\; \frac{117}{2000} \qquad \omega|_{T^*(\partial_t)} \;=\; \frac{18}{117}\,|d\rangle + \frac{99}{117}\,|d^\perp\rangle\,.$$

The latter distribution $\omega|_{T^*(\partial_t)}$ equals $\omega|_{T^*(\partial_t \& C^*(q))}$ for each predicate $q \in [0,1]^{2_C}$ on 2_C, by Proposition 6 ((b)).

▶ Remark. We emphasise that, if we replace the predicate ∂_t on 2_D by a non-Dirac predicate $p \in [0,1]^{2_D}$, then there is *no blocking*, in general. For instance, take: $p(t) = \frac{1}{3}$, $p(t^\perp) = \frac{1}{4}$, $q(c) = \frac{1}{5}$ and $q(c^\perp) = 1$. Then we compute a difference between the following states on 2_D.

$$\omega|_{T^*(p)} = 0.013\,|d\rangle + 0.987\,|d^\perp\rangle \qquad \omega|_{T^*(p\,\&\,c^*(q))} = 0.006\,|d\rangle + 0.994\,|d^\perp\rangle$$

Hence, influence from right to left in (10) does exist for non-sharp predicates.

3.2 Fork connections

Next we consider a "fork" Bayesian network with predicates p, ∂_x, q, for a given element $x \in A$, as on the left. The informal description of this situation is: influence can pass between the children B and C via A, unless the state of A is known, as formulated *e.g.* in [8].

$$\begin{tikzcd}
& & A \arrow[dll, "f"'] \arrow[d, dashed, "\partial_x"] \arrow[drr, "g"] & & \\
& & 2 & & \\
2 & B \arrow[l, dotted, "p"'] & & C \arrow[r, dotted, "q"] & 2
\end{tikzcd} \tag{11}$$

▶ **Example 8.** The Bayesian network of Example 3 contains a fork, given by 'Disease', 'Test' and 'Symptoms'. If a patient tests positively, it gets more likely that she has the disease, and thus shows symptoms. However, if one gets to know with certainty that she has the disease, then any evidence about the test will not change the likelihood of showing symptoms.

▶ **Proposition 9 (Blocking II).** *In the fork network* (11)*, with Dirac evidence on the middle node A, there is no influence from B to C, nor from C to B. This lack of influence from B to C is expressed via the equation:*

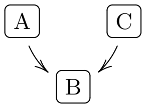
$$g_*\big(\omega|_{\partial_x}\big) = g_*\big(\omega|_{f^*(p)\&\partial_x}\big) \tag{12}$$

for each state ω on A and predicate p on B, and $x \in A$. The other direction is analogous.

Proof The state transformer g_* is irrelevant, as $\omega|_{\partial_x} = 1\,|x\rangle = \omega|_{f^*(p)\&\partial_x}$. The first equation is in point (a) in Lemma 2, and the second one in point (b). □

3.3 Collider connections

The last d-separation scenario is the one of a collider:



which becomes in $\mathcal{K}\ell(\mathcal{D})$, with the addition of a predicate q,

$$\begin{tikzcd}
A \otimes C \arrow[d, "f"'] & \\
B \arrow[r, "q"] & 2.
\end{tikzcd} \tag{13}$$

In [14] one can read about this situation: "if the two extreme variables are (marginally) independent, they will become dependent (*i.e.* connected through unblocked path) once we condition on the middle variable (*i.e.* the common effect) or any of its descendants."

In our formalisms, this explanation unravels as follows. We fix states $\sigma \in \mathcal{D}(A)$ and $\tau \in \mathcal{D}(C)$, giving rise to a product state $\sigma \otimes \tau \in \mathcal{D}(A \otimes C)$. If we have evidence $q\colon B \to 2$ on B, then we can pull it back to evidence $f^*(q)\colon A \otimes C \to 2$. Now, in order to complete our formalisation, we would like to express that σ and τ are initially independent of each other when joint in $\sigma \otimes \tau$, but they get correlated after conditioning $(\sigma \otimes \tau)|_{f^*(q)}$. This correlation should be witnessed by the fact that from now on any predicate on the A-component σ will also have influence on the C-component τ, and viceversa. However, our formalisms of § 2 still lacks the means of expressing such 'crossover' properties, which echo the entanglement phenomena commonly studied in quantum theory. We devote the next two sections to rigorously describe them within our approach, and return to the collider scenario in § 6.

4 Joint states and entwinedness

We now commence the formal investigation of correlation phenomena which will lead to the notion of *crossover* influence. We give an elementary illustration first.

▶ **Example 10.** Consider two diseases A_1 and A_2 which may occur together, as given by the prior joint probability distribution: $\omega = \frac{1}{6}|a_1a_2\rangle + \frac{1}{4}|a_1a_2^\perp\rangle + \frac{1}{3}|a_1^\perp a_2\rangle + \frac{1}{4}|a_1^\perp a_2^\perp\rangle$. Assume that there is a test for disease A_1 with sensitivity 90% positive when a patient has the disease A_1, and 5% positive when the patient does not. It turns out the prior probability of A_2 is $\frac{1}{2}$, but decreases to $\frac{40}{97}$ after a A_1-positive test. We shall see how this works in Example 15.

For two states/distributions $\sigma \in \mathcal{D}(A_1)$ and $\tau \in \mathcal{D}(A_2)$ we can form the joint 'product' distribution $\sigma \otimes \tau \in \mathcal{D}(A_1 \otimes A_2)$ as $(\sigma \otimes \tau)(a_1, a_2) = \sigma(a_1) \cdot \tau(a_2)$, as already used in (7). The two original states σ and τ can be recovered as marginals of this product state: $\mathsf{M}_1(\sigma \otimes \tau) = \sigma$ and $\mathsf{M}_2(\sigma \otimes \tau) = \tau$. Marginalisation (of states) and weakening (of predicates) are special cases of state and predicate transformation, namely for the (Kleisli) projection maps $\pi_i \colon A_1 \otimes A_2 \rightsquigarrow A_i$, given by $\pi_i(a_1, a_2) = 1|a_i\rangle$. Marginalisation moves a 'joint' state on a product to one of the components, and weakening moves a predicate on a component to the product. These two operations play a special role in the sequel, and therefore we introduce explicit notation M and W. First, for a joint state $\omega \in \mathcal{D}(A_1 \otimes A_2)$ we have first and second marginalisation $\mathsf{M}_i(\omega) = (\pi_i)_*(\omega) \in \mathcal{D}(A_i)$ determined by (6) as:

$$\mathsf{M}_1(\omega)(a_1) = \textstyle\sum_{a_2 \in A_2} \omega(a_1, a_2) \qquad \mathsf{M}_2(\omega)(a_2) = \textstyle\sum_{a_a \in A_1} \omega(a_1, a_2). \tag{14}$$

Similarly we have weakening operations $\mathsf{W}_i(p_i) = (\pi_i)^*(p_i) \in [0,1]^{A_1 \otimes A_2}$ for predicates $p_i \in [0,1]^{A_i}$ given by:

$$\mathsf{W}_1(p_1)(a_1, a_2) = p_1(a_1) \qquad \mathsf{W}_2(p_2)(a_1, a_2) = p_2(a_2). \tag{15}$$

Also, for two predicates $p_i \in [0,1]^{A_i}$, we introduce their *parallel conjunction* $p_1 \odot p_2 \in [0,1]^{A_1 \times A_2}$, mapping (a_1, a_2) to $p_1(a_1) \cdot p_2(a_2)$. The following definition describes the interaction — dependence, in Bayesian jargon — between the components of a joint state.

▶ **Definition 11.** A joint state $\omega \in \mathcal{D}(A_1 \otimes A_2)$ is called *non-entwined* if it is the product of its marginals: $\omega = \mathsf{M}_1(\omega) \otimes \mathsf{M}_2(\omega)$. It is called *entwined* otherwise.

▶ **Lemma 12.** *(a)* $\mathsf{M}_1(\omega) \models p \;=\; \omega \models \mathsf{W}_1(p)$ *and* $\mathsf{M}_2(\omega) \models p \;=\; \omega \models \mathsf{W}_2(p)$.
(b) $\mathsf{W}_1(p) = p \odot \mathbf{1}$ *and* $\mathsf{W}_2(q) = \mathbf{1} \odot q$ *and* $p \odot q = \mathsf{W}_1(p) \,\&\, \mathsf{W}_2(q)$.
(c) $(\sigma \otimes \tau) \models (p \odot q) = (\sigma \models p) \cdot (\tau \models q)$ *and* $(\sigma \otimes \tau)|_{p \odot q} = (\sigma|_p) \otimes (\tau|_q)$. □

The next result plays an important role in the sequel. The first equation below says that if one takes the marginal of a joint state conditioned with a weakened predicate, then one may as well condition the marginal directly. This holds if the weakening and marginalisation use the same component. But it fails if the components are different, see the subsequent inequality $\neq$ below. The latter fact is remarkable, because it involves a form of influence between components. This is also called 'signalling' in the quantum world, but apparently already appears in the current probabilistic setting — only for entwinted states.

▶ **Proposition 13.** *Let $p \in [0,1]^A$ be a predicate on a set A.*

(a) For an arbitrary joint state $\omega \in \mathcal{D}(A \otimes B)$,

$$\mathsf{M}_1\big(\omega|_{\mathsf{W}_1(p)}\big) = \mathsf{M}_1(\omega)|_p \qquad \textit{but in general} \qquad \mathsf{M}_2\big(\omega|_{\mathsf{W}_1(p)}\big) \neq \mathsf{M}_2(\omega).$$

(b) For the special case of a (non-entwined) product state $\sigma \otimes \tau \in \mathcal{D}(A \otimes B)$,

$$\mathsf{M}_1\big((\sigma \otimes \tau)|_{\mathsf{W}_1(p)}\big) \;=\; \sigma|_p \qquad\qquad \mathsf{M}_2\big((\sigma \otimes \tau)|_{\mathsf{W}_1(p)}\big) \;=\; \tau.$$

Proof We only prove the equality in point (a), and refer to Example 14 (b) for the inequality in point (b), where a counterexample is given.

$$\begin{aligned}\mathsf{M}_1\big(\omega|_{\mathsf{W}_1(p)}\big)(a) \overset{(14)}{=} \textstyle\sum_b \omega|_{\mathsf{W}_1(p)}(a,b) \overset{(3)}{=} \textstyle\sum_b \dfrac{\omega(a,b)\cdot \mathsf{W}_1(p)(a,b)}{\omega \models \mathsf{W}_1(p)} &\overset{(15)}{=} \dfrac{\sum_b \omega(a,b)\cdot p(a)}{\omega \models \mathsf{W}_1(p)} \\ &\overset{\text{Lem.}12(a)}{=} \dfrac{\mathsf{M}_1(\omega)(a)\cdot p(a)}{\mathsf{M}_1(\omega) \models p} \\ &\overset{(3)}{=} \mathsf{M}_1(\omega)|_p(a). \qquad \square\end{aligned}$$

We illustrate two significant related phenomena via an example.

▶ **Example 14.** Given sets $X = \{x, y\}$ and $A = \{a, b\}$, one can prove that a state $\omega = r_1\,|x, a\rangle + r_2\,|x, b\rangle + r_3\,|y, a\rangle + r_4\,|y, b\rangle \in \mathcal{D}(X \otimes A)$, where $r_1 + r_2 + r_3 + r_4 = 1$, is non-entwined if and only if $r_1 \cdot r_4 = r_2 \cdot r_3$. This fact also holds in the quantum case, see *e.g.* [12, §1.5]. It plays a role in the next two illustrations.

(a) **Conditioning creates entwinedness**. Recall from Example 3 the joint state $G \otimes E$ on $2_G \otimes 2_E$, defined as in (7). Consider now a predicate $p \in [0,1]^{2_D}$ defined by $d \mapsto 85\%$ and $d^\perp \mapsto 20\%$. It models, for instance, the information that pallor appears as a symptom in 85% of patients with the disease, but also healthy patients may be pale for other reasons, 20% of the times. Using D as a predicate transformer, we can form the conditioned state $\omega = (G \otimes E)|_{D^*(p)} = 0.007\,|g, e\rangle + 0.055\,|g, e^\perp\rangle + 0.191\,|g^\perp, e\rangle + 0.747\,|g^\perp, e^\perp\rangle$. This state is now entwined, see the above characterisation of non-entwinedness.

(b) **Influence between marginals of entwined states**. Let's now start with an entwined state $\sigma = \frac{1}{3}\,|g, e\rangle + \frac{1}{4}\,|g, e^\perp\rangle + \frac{1}{6}\,|g^\perp, e\rangle + \frac{1}{4}\,|g^\perp, e^\perp\rangle \in \mathcal{D}(2_G \otimes 2_E)$ and a predicate $q = \partial_g \in [0,1]^{2_G}$. By weakening we get $\mathsf{W}_1(q) = q \bullet \pi_1 \in [0,1]^{2_G \otimes 2_E}$. Then: $\sigma \models \mathsf{W}_1(q) = \frac{1}{3}\cdot 1 + \frac{1}{4}\cdot 1 = \frac{7}{12}$, so that:

$$\sigma|_{\mathsf{W}_1(q)} \;=\; \tfrac{1/3}{7/12}\,|g, e\rangle + \tfrac{1/4}{7/12}\,|g, e^\perp\rangle \;=\; \tfrac{4}{7}\,|g, e\rangle + \tfrac{3}{7}\,|g, e^\perp\rangle\,.$$

Below, the second marginal of the original state σ differs from the second marginal of this conditioned state, illustrating the inequality $\neq$ in Proposition 13 (a).

$$\mathsf{M}_2(\sigma) \;=\; \tfrac{1}{2}\,|e\rangle + \tfrac{1}{2}\,|e^\perp\rangle \qquad \text{whereas} \qquad \mathsf{M}_2\big(\sigma|_{\mathsf{W}_1(q)}\big) \;=\; \tfrac{4}{7}\,|e\rangle + \tfrac{3}{7}\,|e^\perp\rangle\,.$$

▶ **Example 15.** We conclude with the formal description of the two-disease scenario with which we started this section (Example 10). The test is a map $T\colon 2_{A_1} \to 2_T$ given by $T(a_1) = \frac{9}{10}\,|t\rangle + \frac{1}{10}\,|t^\perp\rangle$ and $T(a_1^\perp) = \frac{1}{20}\,|t\rangle + \frac{19}{20}\,|t^\perp\rangle$. The impact of a positive test on the disease A_2 is given by the marginal of the conditional: $\mathsf{M}_2(\omega|_{\mathsf{W}_1(T^*(\partial_t))}) = \frac{40}{97}\,|a_2\rangle + \frac{57}{97}\,|a_2^\perp\rangle$.

5 A quantitative definition of influence

Last section showed how evidence on one component of an entwined state may influence the other component. But *how much* did it change the latter component with respect to our prior belief? This section addresses such aspect by introducing a quantitative semantics for our influence vocabulary. We begin by recalling the total variation metric on distributions.

▶ **Definition 16.** Let $\sigma, \tau \in \mathcal{D}(X)$ be two distributions on a set X. Their *total variation distance* $\mathsf{d}(\sigma,\tau)$ is defined as the following number in the unit interval $[0,1]$.

$$\mathsf{d}(\sigma,\tau) \;=\; \frac{1}{2}\sum_{x\in X}\big|\sigma(x) - \tau(x)\big|.$$

▶ **Lemma 17.** *Let $f\colon X \rightarrow\!\!\!\!\circ\; Y$ be a Kleisli map. The associated state transformer $f_*\colon \mathcal{D}(X) \to \mathcal{D}(Y)$ from* (6) *is non-expansive: $\mathsf{d}(f_*(\sigma), f_*(\tau)) \leq \mathsf{d}(\sigma,\tau)$. This yields a functor $\mathcal{K}\ell(\mathcal{D}) \to \mathbf{Met}_1$, where $\mathbf{Met}_1$ is the category of 1-bounded metric spaces and non-expansive maps.*

Proof

$$\begin{array}{rcl}
\mathsf{d}(f_*(\sigma), f_*(\tau)) & = & \frac{1}{2}\sum_{y\in Y}\big|\, f_*(\sigma)(y) - f_*(\tau)(y)\,\big| \\
& \overset{(6)}{=} & \frac{1}{2}\sum_{y\in Y}\big|\, \sum_{x\in X} f(x)(y)\cdot\sigma(x) - \sum_{x\in X} f(x)(y)\cdot\tau(x)\,\big| \\
& \leq & \frac{1}{2}\sum_{x\in X}\sum_{y\in Y} f(x)(y)\cdot\big|\,\sigma(x) - \tau(x)\,\big| \\
& = & \frac{1}{2}\sum_{x\in X} 1\cdot\big|\,\sigma(x) - \tau(x)\,\big| \;=\; \mathsf{d}(\sigma,\tau).
\end{array}$$

□

We refer to [6] for more information about total variation (and Kantorovich) distance and the distribution monad $\mathcal{D}$, and turn to our formal definition of influence. First we define it in direct form, as a number indicating how much a predicate p influences a state σ via conditioning $\sigma|_p$, given by the (total variation) distance between σ and $\sigma|_p$. This seems fairly simple. But, as we have seen in Section 4, there may also be indirect, 'crossover' influence between the components of a joint entwined state: this is the content of our second definition.

▶ **Definition 18.** Let $p \in [0,1]^A$ be a predicate on a set A with discrete metric.

1. For a state $\sigma \in \mathcal{D}(A)$ on A the *direct influence* of p on σ is defined as:

$$\mathcal{I}_d(p,\sigma) \;:=\; \mathsf{d}\big(\sigma, \sigma|_p\big) \qquad \text{provided } \;\sigma \models p \neq 0.$$

2. For a joint state $\omega \in \mathcal{D}(A \otimes B)$ the *crossover influence* of p on ω is:

$$\mathcal{I}_c(p,\omega) \;:=\; \mathsf{d}\big(\mathsf{M}_2(\omega), \mathsf{M}_2(\omega|_{\mathsf{W}_1(p)})\big) \qquad \text{provided } \;\omega \models \mathsf{W}_1(p) \neq 0.$$

In general we say that a predicate has *no* (direct or crossover) influence on a state if the corresponding influence function ($\mathcal{I}_d$ or $\mathcal{I}_c$) has outcome zero.

▶ **Example 19.** We give an example of *direct* influence, postponing a detailed illustration of *crossover* influence to the collider scenario in Section 6. Recall the Kleisli map (5) modeling the Bayesian network of Example 3. We fix three different states on $2_D = \{d, d^\perp\}$:

$$\omega \;=\; \tfrac{4}{5}\,|d\rangle + \tfrac{1}{5}\,|d^\perp\rangle \qquad \rho \;=\; \tfrac{1}{2}\,|d\rangle + \tfrac{1}{2}\,|d^\perp\rangle \qquad \sigma \;=\; \tfrac{1}{5}\,|d\rangle + \tfrac{4}{5}\,|d^\perp\rangle\,.$$

Intuitively, in state ω it is likely that the patient has the disease, in state σ it is rather unlikely, and ρ sits in the middle. Consider the Dirac predicate $\partial_t \in [0,1]^{2_T}$ expressing positivity of the test: we first use the predicate transformer T^* associated with the Kleilsi map $T\colon 2_D \rightarrow\!\!\!\!\circ\; 2_T$ to obtain a predicate $T^*(\partial_t) \in [0,1]^{2_D}$; subsequently, we compute the influence of $T^*(\partial_t)$ on the above three states. This is done via a script.

$$\mathcal{I}_d\big(T^*(\partial_t),\omega\big) \;=\; 0.19 \qquad \mathcal{I}_d\big(T^*(\partial_t),\rho\big) \;=\; 0.45 \qquad \mathcal{I}_d\big(T^*(\partial_t),\sigma\big) \;=\; 0.62$$

Influence measures how radically the positivity of the test challenges our belief on the disease: a positive test does not come at surprise in state ω, but it is more unexpected in state σ.

▶ **Example 20.** Clearly, $\mathcal{I}_d(\mathbf{1}, \omega) = 0$, for the truth predicate $\mathbf{1}$, since $\omega|_{\mathbf{1}} = \omega$. Is there also an example where the (direct and crossover) influence reaches the maximal distance 1? We show how to approximate it. Take $A = \{a, b\}$ with predicate $p(a) = 1, p(b) = 0$ and state $\sigma = \epsilon\,|a\rangle + (1-\epsilon)\,|b\rangle$. The direct influence $\mathcal{I}_d(p, \sigma)$ goes to 1 as $\epsilon \to 0$. Similarly, by taking $\omega = \epsilon\,|aa\rangle + (1-\epsilon)\,|bb\rangle \in \mathcal{D}(A \times A)$ we get $\mathcal{I}_c(p, \omega) \to 1$ as $\epsilon \to 0$ for crossover influence.

We now establish some facts on crossover influence: (1) it only makes sense if the state is entwined, since for a product state the crossover influence is zero; (2) weakening and marginalisation must work in different components, since otherwise we have direct influence; (3) crossover influences is always less than direct influence. In the context of Definition 18:

▶ **Lemma 21.** *1.* $\mathcal{I}_c(p, \sigma \otimes \tau) = 0$*;*
2. $\mathsf{d}\big(\mathsf{M}_1(\omega), \mathsf{M}_1(\omega|_{\mathsf{W}_1(p)})\big) = \mathcal{I}_d(p, \mathsf{M}_1(\omega))$*;*
3. $\mathcal{I}_c(p, \omega) \leq \mathcal{I}_d(\mathsf{W}_1(p), \omega)$
4. For each function $g\colon X \to Y$*, considered as a Kleisli map* $\langle g\rangle\colon X \to \mathcal{D}(Y)$*, we have:* $\mathcal{I}_d(p, \langle g\rangle_*(\sigma)) \leq \mathcal{I}_d(\langle g\rangle^*(p), \sigma)$*, where* $\sigma \in \mathcal{D}(X)$*.*

Proof The first two points follow directly from Proposition 13 (b) and (a). The inequality in point (3) from the fact that marginalisation is a special form of state transformation, which, as we know from Lemma 17, is non-expansive:

$$\begin{aligned}\mathcal{I}_c(p, \omega) \;&=\; \mathsf{d}\big(\mathsf{M}_2(\omega), \mathsf{M}_2(\omega|_{\mathsf{W}_1(p)})\big) \;=\; \mathsf{d}\big((\pi_2)_*(\omega), (\pi_2)_*(\omega|_{\mathsf{W}_1(p)})\big) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\;\leq\; \mathsf{d}\big(\omega, \omega|_{\mathsf{W}_1(p)}\big) \;=\; \mathcal{I}_d(\mathsf{W}_1(p), \omega).\end{aligned}$$

Finally, for point (4) we use both Lemma 4 (c) and Lemma 17 in:

$$\begin{aligned}\mathcal{I}_d(p, \langle g\rangle_*(\omega)) \;&=\; \mathsf{d}\big(\langle g\rangle_*(\omega), \langle g\rangle_*(\omega)|_p\big) \\ &=\; \mathsf{d}\big(\langle g\rangle_*(\omega), \langle g\rangle_*(\omega|_{\langle g\rangle^*(p)})\big) \;\leq\; \mathsf{d}\big(\omega, \omega|_{\langle g\rangle^*(p)}\big) \;=\; \mathcal{I}_d(\langle g\rangle^*(p), \omega).\end{aligned}$$ □

▶ Remark. Crossover and direct influence are instances of a more general definition of influence of a predicate $p \in [0,1]^{A_j}$ on the i-th marginal A_i of a joint state $\omega \in \mathcal{D}(A_1 \otimes \ldots \otimes A_n)$. For $n = 2$ and $i \neq j$, this corresponds to crossover influence, whereas for $n = i = j = 1$ it would be direct influence. We chose not to work within this uniform approach as we believe that it is more insightful to think of the two notions of Definition 18 as conceptually distinct.

As observed in §3, the blocking action of Dirac predicates plays a key role in d-separation. We can use Definition 18 to express that no predicate p has any influence on a Dirac-conditioned state $\omega|_{\partial_x}$— by Lemma 2, $(\omega|_{\partial_x})|_p = (\omega|_p)|_{\partial_x} = 1\,|x\rangle = \omega|_{\partial_x}$, so $\mathcal{I}_d(p, \omega|_{\partial_x}) = 0$.

▶ **Example 22.** For instance, we can reformulate the fork scenario as follows. Because conditioning is commutative, (12) is the same as: $\omega|_{\partial_x} = \big(\omega|_{\partial_x}\big)|_{f^*(p)}$. Thus Proposition 9 says that $\mathcal{I}_d(f^*(p), \omega|_{\partial_x}) = 0$, *i.e.* $f^*(p)$ has no influence on $\omega|_{\partial_x}$.

In the same vein, one may also revisit Example 8, an instance of the serial connection scenario: in short, from (5), use states D, $T \bullet D$ and $S \bullet T \bullet D$ to construct a joint state on $2_D \otimes 2_T \otimes 2_S$; check that a 'positive test' predicate $\partial_t \in [0,1]^{2_T}$ has crossover influence on the marginal 2_S, then prove that a 'disease' predicate $\partial_d \in [0,1]^{2_D}$ blocks such influence.

6 Influence in d-separation (reprise)

We conclude with a return on the collider scenario, left unfinished at the end of § 3. With the notation introduced therein, we now explain the collider situation in Diagram (13): the initial joint (product) state $\sigma \otimes \tau$ is non-entwined, but it becomes entwined after conditioning with evidence q on B, as in $(\sigma \otimes \tau)|_{f^*(q)}$. Now any new evidence $p \in [0,1]^A$ on A may have *crossover* influence on C— *cf.* Example 14 (b). It can be explicitly quantified by computing $\mathcal{I}_c\big(p, (\sigma \otimes \tau)|_{f^*(q)}\big)$.

A conceptual insight stemming from our analysis is the asymmetry between blocking and enabling influence: while in the serial and fork scenarios only Dirac predicates are able to block, in a collider *any* predicate may enable. We give a concrete example below.

▶ **Example 23.** The Bayesian network of Example 3 includes a collider, given by nodes 'Genetic Heredity' and 'Environmental Factors' both pointing to 'Disease'. The two possible causes for the disease are represented as a joint state $G \otimes E$ on $2_G \otimes 2_E$, see (7). *A priori*, they are unrelated. For instance, a Dirac predicate $\partial_{g^\perp} \in [0,1]^{2_G}$ that excludes any genetic disorder of the patient has no effect on the chances that she has been exposed to the environmental factors: formally, the crossover influence $\mathcal{I}_c(\partial_{g^\perp}, G \otimes E)$ is 0, as guaranteed by Lemma 21.1. However, let's include the information that pallor is a symptom of the disease, modeled as a predicate $p \in [0,1]^{2_D}$ as in Example 14(a): it turns $G \otimes E$ into an entwined state $(G \otimes E)|_{D^*(p)}$. In this changed scenario, d-separation tells that ruling out genetic heredity (predicate $\partial_{g^\perp}$) *does* influence the belief that environment was the cause. We can formally expressed it with crossover influence:

$$\mathcal{I}_c(\partial_{g^\perp}, G \otimes E) = 0 \qquad\qquad \mathcal{I}_c(\partial_{g^\perp}, (G \otimes E)|_{D^*(p)}) = 0.006.$$

Note that a Dirac predicate $\partial_d \in [0,1]^{2_D}$ expressing certainty of the disease entwines G and E much more: indeed $\mathcal{I}_c(\partial_{g^\perp}, (G \otimes E)|_{D^*(\partial_d)}) = 0.26 > \mathcal{I}_c(\partial_{g^\perp}, (G \otimes E)|_{D^*(p)})$.

7 Discussion

Our ambition in this paper was to develop a framework where grounding concepts of Bayesian reasoning (influence, dependence, blocking, evidence, ...) are given a clear, completely formal meaning, building on [7], and can be reasoned about in an abstract and flexible manner. As a proof of concept, we analysed d-separation: the intention was to show how event interactions with a subtle and potentially ambiguous natural language description can be reduced to elementary formulas of our language, with a simple and transparent proof.

We based our approach on Kleisli categories, in harmony with the increasing importance of algebraic methods from program semantics in the analysis of probabilistic systems [11, 16, 10]. The highlight of our developments is the notion of crossover influence, which we believe may foster research in two directions. First, it draws a parallelism with non-locality phenomena of quantum theory, see also [6]: we plan investigate the meaning of our definitions in that setting, exploiting the formal bridge offered by Effectus theory [4, 2]. Second, our definition is abstract enough to accommodate different choices for the underlying notion of distance between states. The total variation metric suits the applications of this paper, but other choices are also worth investigating: we think in particular of the Kantorovich metric [17], for when the sample set has a non-discrete metric, and quantitative analyses of information leakage [1]. Also connections with Kullback-Leibler divergence [9], focussing on loss of information, in Shannon style, and to mutual information, remain to be investigated.

A related point concerns the relationship between the total variation distance and Bayesian influence. In our choice, we simply aimed at the most basic additive distance which does not (unlike Kantorovich) builds on a pre-existing metric, as our sample sets have none. Admittedly, the suitability of total variation is only empirically justified by examples. In future work we aim at a more satisfactory investigation: recent advances on an axiomatic treatment of metrics [10] appear to be very suitable for the purpose.

References

1 M. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. Measuring information leakage using generalized gain functions. In *Computer Security Foundations Symposium (CSF 2012)*, pages 265–279, 2012.

2 K. Cho, B. Jacobs, A. Westerbaan, and B. Westerbaan. An introduction to effectus theory. see `arxiv.org/abs/1512.05813`, 2015.

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

4 B. Jacobs. New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, 11(3):1–76, 2015.

5 B. Jacobs. From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, 156, 2016. See `http://dx.doi.org/10.1016/j.jlamp.2016.11.006`.

6 B. Jacobs. A note on distances between probabilistic and quantum distributions. In A. Silva, editor, *Math. Found. of Programming Semantics*, Elect. Notes in Theor. Comp. Sci. Elsevier, Amsterdam, 2017.

7 B. Jacobs and F. Zanasi. A predicate/state transformer semantics for Bayesian learning. In L. Birkedal, editor, *Math. Found. of Programming Semantics*, number 325 in Elect. Notes in Theor. Comp. Sci., pages 185–200. Elsevier, Amsterdam, 2016.

8 F. Jensen. *Bayesian Networks and Decision Graphs.* Statistics for Engineering and Information Science. Springer, 2001.

9 S. Kullback. *Information Theory and Statistics.* John Wiley & Sons, 1959.

10 R. Mardare, P. Panangaden, and G. Plotkin. Quantitative algebraic reasoning. In *Logic in Computer Science, LICS '16*, pages 700–709. IEEE, Computer Science Press, 2016.

11 A. McIver, C. Morgan, and T. Rabehaja. Abstract hidden Markov models: A monadic account of quantitative information flow. In *Logic in Computer Science*, pages 597–608. IEEE, Computer Science Press, 2015.

12 N.D. Mermin. *Quantum Computer Science: An Introduction.* Cambridge Univ. Press, 2007.

13 J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

14 J. Pearl. *Causality. Models, Reasoning, and Inference.* Cambridge Univ. Press, 2$^{\text{nd}}$ ed. edition, 2009.

15 S. Russel and P. Norvig. *Artificial Intelligence. A Modern Approach.* Prentice Hall, Englewood Cliffs, NJ, 2003.

16 S. Staton, H. Yang, C. Heunen, O. Kammar, and F. Wood. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *Logic in Computer Science, LICS '16*, pages 525–534, 2016.

17 C. Villani. *Optimal Transport — Old and New.* Springer Berlin Heidelberg, 2009.