Integrating Causal Bayes Nets and Inferentialism in Conditional Inference

Mike Oaksford

Birkbeck College, University of London

Nick Chater

Warwick Business School, Warwick University

Author Note

We are very grateful to David Over and Igor Douven for very valuable critiques of the position taken in this chapter. They have forced us to rethink many issues, and to think more deeply about them than we initially intended.

Abstract

This paper argues that recent developments in inferentialism in the psychology of reasoning that challenge the suppositional approach advocated by David Over can be implemented in Causal Bayes Nets (CBNs). Inferentialism proposes that conditionals, *if p then q*, imply (either as a matter of their meaning or a conventional implicature) that there is an inferential dependency between *p and q*. These dependencies can be captured in the directional links of a CBN ($p \rightarrow q$), which can, therefore, provide a theory of mental representation and inference that inferentialism currently lacks. This approach has already been demonstrated for causal conditionals. We conclude that this proposal, while losing some inferences valid in the suppositional view, gains others that we know people make while also retaining consistency with the general Bayesian framework for human reasoning.

Keywords: Causal Bayes Nets, Reasoning, Conditional Inference, Inferentialism, Theory Integration.

Integrating Causal Bayes Nets and Inferentialism

in Conditional Inference

Before discussing some connections between the application of Causal Bayes Nets to the psychology of conditional reasoning and their relation to recent discussions of inferentialism, we make some brief comments on David, in this Festschrift for his 70th birthday. Mike has known, or known of, David for most of his academic life having first attended a seminar David gave at Department of Philosophy at Durham University when Mike was a mature undergraduate student in the early 1980s. It was David's work with Ken Manktelow (Mike's external PhD examiner at Edinburgh, Cognitive Science), proposing closer links between reasoning and decision making that provided part of the inspiration for our subsequent joint work on the Bayesian view of human reasoning (1994, 2007, 2020). David's early collaborator on probabilities and conditionals, Rosemary Stevenson (a PhD student of Phil Johnson-Laird's), was also one of Mike's lecturers as an undergraduate at Durham. Over the years, it feels like David and we have been on parallel tracks, occasionally sniping at each other, but going in the same direction. So much so that we all happily adopted the term he coined for these developments "The New Paradigm" in a critical review (Over, 2009) of Oaksford and Chater (2007).

David has taken on the mantle of the main proponent of the suppositional view of the conditional in which the probability of the conditional is the conditional probability (*The Equation,* Edgington, 1995) and the Ramsey test determines the conditional probability: suppose the antecedent is true, adjust you other beliefs to accommodate this supposition, and then read off your new probability of the consequent. More recently David and Mike have collaborated in jointly supervising Nicole Cruz's PhD with Jean Baratgin (Cruz, 2018; Cruz, Baratgin, Oaksford,

& Over, 2015; Cruz, Over, & Oaksford, 2017; Cruz, Over, Oaksford, & Baratgin, 2016) and in critiquing recent developments in mental models theory (Oaksford, Over, & Cruz, 2019).

However, this does not mean that we all see eye to eye on everything. Indeed within the broad church that is now the new paradigm, there are several seemingly conflicting views. This is a healthy development, as it can only be by rigorous argument and debate that ideas can evolve. However, there are two things from which we should not be distracted by these internal disagreements. First, we should continue to rigorously test the new paradigm against the old and not be overly distracted arguing amongst ourselves. Although, it is important to note that the old paradigm cannot even frame many of these internal disagreements and so they are in effect discriminatory. Second, we should be attempting as far as possible to develop integrative theories in the new paradigm. In this paper, we suggest, in very broad terms, that we can integrate recent developments in inferentialism with previous theory based on Causal Bayes Nets (CBNs; Pearl, 1988). We argue that CBNs can implement inferentialism, thereby providing a theory of mental representation and inference that it currently lacks.

We first look at CBNs and how they have been used to explain conditional reasoning (Ali, Chater, & Oaksford, 2011; Ali, Schlottmann, Shaw, Chater, & Oaksford, 2010; Chater & Oaksford, 2006; Fernbach & Erb, 2013; Hall, Ali, Chater, & Oaksford, 2016; Oaksford & Chater, 2010, 2013, 2014, 2016, 2017). We then introduce the two versions of inferentialism currently in the literature, the probabilistic view (Krzyżanowska, Collins, & Hahn, 2017; Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, & Klauer, 2019; Skovgaard-Olsen, Kellen, Hahn, & Klauer, 2019; Skovgaard-Olsen, Kellen, Krahl, & Klauer, 2017; Skovgaard-Olsen, Singmann, & Klauer, 2016, 2017) and the semantic/explanatory view (Douven, Elqayam, Singmann, & van Wijnbergen-

Huitink, 2018; Douven & Mirabile, 2018; Mirabile & Douven, 2019). All three accounts provide challenges for the suppositional view of the conditional.

Causal Bayes Nets

CBNs treat the causal dependencies that people believe to be operative in the world as basic (Pearl, 1988; 2000). CBNs represent causes as edges in a directed acyclic graph. The nodes represent Bayesian random variables and the directed links run from cause to effect (i.e., the arrows represent causal direction). Nodes that are not connected represent variables that are conditionally independent of each other. The *parents* of a node are those that connect to it further back down the causal chain. These networks have probability distributions defined over them that partly rely on the dependency structure.

Integration Rules and Causal Power

Integration rules determine how the multiple parents of a node combine, for example, a common effect structure ($p \rightarrow q \leftarrow r$) combines two causes using the noisy-OR rule. Suppose, *p* and *r* represent *turning the key* and *hotwiring*, respectively. These are independent causes of *the car starting* (q). Assume there are no other causes of the car starting. On this assumption, the probability of the car starting is $1 - (1 - W_r)^{ind(r)}(1 - W_p)^{ind(p)}$, where $ind(p) = 1$ if the key is turned and 0 if it is not and W_i is the causal power of i to cause q. If this was a deterministic system (i.e., $W_r = W_p = 1$), then this formula is equivalent to logical inclusive *or* (i.e., it gives probability 1 unless both causes are absent, when it gives probability 0, this is also called Boolean addition ($\oplus$), i.e., if the key is not turned and the car is not hotwired, then the car does not start).

Causal power is determined by Cheng's (1997) formula:

$$W = \frac{\Pr(q|p) - \Pr(q|\neg p)}{1 - \Pr(q|\neg p)} \qquad (Eq. 1)$$

Alternatively, it can be derived from the aggregate power of disablers to prevent the cause from producing the effect (see, Oaksford & Chater 2017: Equations 8 & 9; Pearl. 1988). The numerator of causal power in Cheng's (1997) formula is delta-P (ΔP), which is a common measure of deviation from independence. It is one of many confirmation indices proposed as measures of the degree to which the evidence supports the existence of the dependency between *p* and *q* (see Hattori & Oaksford [2007] for 40 more).

CBNs and Conditionals

Why do we believe that CBN's can provide a psychologically acceptable account of conditional inference? In many papers, we have made the following argument, which we condense here for the sake of brevity (Oaksford & Chater, 2010, 2013, 2017). The inadequacy of the material conditional came into stark relief when the logical positivists attempted to use it as a characterisation of law-like relations (Chisholm, 1946; Goodman, 1947, 1955). Following Hume's second definition of *c* causes *e*, such relationships should support counterfactuals: *if c had not occurred than nor would e*. However, on the material conditional, counterfactuals are always true because their antecedents are false, which violates strong intuitions that counterfactual conditionals can be false. The subsequent development of possible worlds semantics provided an account of counterfactuals and hence causation, which, as has been pointed out many times (Stalnaker, 1984), is more obscure than the notion of causation itself.

Moreover, defining causation in terms of counterfactuals would appear to involve a circular appeal to causation (Stalnaker, 1984). Causes seem to be part of our basic building blocks for describing the world and not analyzable in terms of possible worlds or probabilities (Cartwright, 1983). Stalnaker (1984) argued that open indicative conditionals describe our methodological policies for changing our beliefs; that is, they describe our general habits of inference. Moreover, as such, they are generally regarded as supporting counterfactuals, which requires people to believe that they describe some real dependency in the world that they treat like a real causal relation. The idea that causal dependencies provide the core semantics for conditionals was also proposed in situation semantics (Barwise & Perry, 1983). [1]

These considerations suggest that we should view conditionals as describing the inferential dependencies that lead people to change their beliefs. Barwise and Perry 1983) interpreted the concept of a dependency, what they referred to as "constraints," to be very general, subsuming, for example, social conventions and people's dispositions and habits, In the latter case, the dependency in the world is an actual methodological policy in somebody else's or your own mind/brain. Moreover, social conventions and regulations often constrain our behaviour as much as causal laws, a view that underpins the intuitive appeal of social realism in sociology (e.g., Pharo, 2007). For our psychological purposes, of course, all we need to do is conjecture that people treat social conventions and cultural norms as realistically as they do physical causes. That is, they project their methodological policies to change their beliefs based on social conventions and norms onto the world. We do not need to sign up to social realism. Inferential dependencies are basic, but we can infer them from data (Griffiths & Tenenbaum, 2005). People suspect they exist, precisely when ΔP is non-zero. Moreover, if represented as the

[1] All these arguments, and there relevance to human conditional reasoning, were made in Oaksford (1989), that is, the PhD for which Ken Manhtelow was the external examiner.

arrows in an acyclic graph in a CBN, they can be parameterised with probabilities to allow further inferences to be drawn.

Conditionals describing inferential dependencies can be viewed as structure building operators in CBNs (Oaksford & Chater, 2010, 2011). Moreover, even if a conditional sentence inverts causal direction, as in abductive or diagnostic conditionals (e.g., *if there are shadows, then the sun is out*), children and adults set up the causally correct (*sun out* → *shadows*) mental representation and draw the appropriate discounting and augmentation inferences (Ali et al, 2010, 2011). These findings extend the range of inferences for conditionals and show that inferentially people treat them like causes. CBNs have also been shown to provide an account of suppression effects in conditional reasoning (Fernbach & Erb, 2013; Oaksford & Chater, 2010, 2013, 2017).

Inferentialism

Inferentialism is the new kid on the block in reasoning research, and it comes in two versions. First, the semantic version in which indicative conditionals express inferential or reason relations between the antecedent and consequent which are part of the truth conditions of the conditional (Douven et al. 2018; Douven & Mirabile, 2018; Mirabile & Douven, 2019). Second, the probabilistic implementation that treats the reason relation as probabilistic and part of the acceptability conditions of indicative conditionals (Skovgaard-Olsen, Collins et al., 2019; Skovgaard-Olsen, Kellen et al., 2019; Skovgaard-Olsen, Kellen et al., 2017; Skovgaard-Olsen, Singmann et al., 2017; Skovgaard-Olsen et al., 2016). The intuitions and evidence for these views derive from missing link conditionals.

Missing Link Conditionals

Missing link conditionals are used to point out problems for the material conditional. The truth conditions for the material conditional means that the falsity of *the moon is made of cheese* allows the inference from *the moon is not made of cheese* to *if the moon is made of cheese, Trump will win the next election*. Such a conditional makes no sense precisely because there is no inferential dependency between the antecedent and consequent. However, the conditionals do not have to have absurd antecedents. This example from Veltman (1986), *if it is raining in Ipanema, then I have toast for breakfast* (assuming I live in New York) seems to be just such a missing-link conditional (although we will query this below). But of course, any non-causal spurious correlation will do *if we spend more on science, more people will hang themselves* (Vigen, 2015). The question is what differentiates these missing link conditionals from conditionals like, *if I turn the key the car starts*. Both inferentialist accounts offer the same diagnosis, that what is missing is an inferential dependency between the two.

Probabilistic inferentialism. In the probabilistic account, assessing whether such a dependency exists, exploits the ΔP index of whether a causal dependency exists. Experiments asked people to evaluate $\Pr(q|p)$ and $\Pr(q|\neg p)$ for a variety of conditionals, allowing ΔP to be calculated, subsequently allowing the conditional to be classified as positively inferentially related ($\Pr(q|p) > \Pr(q|\neg p)$), not inferentially related ($\Pr(q|p) = \Pr(q|\neg p)$), and negatively inferentially related ($\Pr(q|p) < \Pr(q|\neg p)$). ΔP was found to moderate whether the Equation (Pr(*if p then q*) = $\Pr(q|p)$) holds. Only when $\Delta P > 0$, that is positive inferential relevance, does the Equation adequately predict whether a conditional is acceptable.[2] Results such as this seem to set

[2] There is a general problem with this idea, which is that the Equation *implies* the probabilistic independence of antecedent and consequent. This is one of Lewis' triviality results (Lewis, 1976). To

limitations on the applicability of the suppositional account of the conditional. Work in this area has concentrated on attempts to determine whether these effects of inferential relevance are semantic or pragmatic. The most recent research (Skovgaard et al., 2019) has concluded that inferential relevance is a conventional implicature, that is, part of a terms agreed meaning but not contributing to the truth conditions of the statement, in this case, the conditional.

There are some problems with this approach. First, some linguists deny that conventional implicatures exist (Bach, 1999). Second, the semantics/pragmatics distinction is obscure with different authors drawing different lines between the semantic and the pragmatic (see, for example, Levinson, 2000, p. 195, Table 3.1). Third, some authors, like Recanati (1988), maintain that pragmatic elements (elements that arise out of the use of a sentence in a given context) enter into the determination of truth conditions, so that a conventional implicature *can* contribute to the truth conditions of a conditional. Fourth, Gricean pragmatics assumes that conditionals have truth conditions. Therefore, it may be preferable on conceptual grounds, regardless of how many tests of being a pragmatic phenomenon it passes (Skovgaard et al., 2019), for inferential relevance to figure in the truth conditions of a conditional.

Semantic inferentialism. In the semantic account, experiments show that the strength of inferential connection predicts judgements of truth (Douven et al., 2018). These experiments involved a soritical series of objects, for example, colour patches, moving in graded steps from definitely blue to definitely green. Participants then judged the truth of conditional statements like *if patch 2 is blue, then patch 3 is blue* and *if patch 3 is blue, then patch 8 is blue.* The

---

circumvent it, one could deny that conditionals express truth conditions, but that comes at a considerable cost, as we would then be unable to explain how conditionals interact with the rest of the language. We thanl Igor Douven for this point.

connection between adjacent patches (2 and 3) is far stronger than that between patches with a greater separation (3 and 8). Douven et al. (2018) found that as separation increased judgements of truth fell. In further experiments, they have shown that judgements of explanatory goodness outperform conditional probability in predicting the acceptability of explanations (Douven & Mirabile, 2018). Explanatory quality also seems to be a better predictor of judging a conditional to be true than conditional probability (Mirabile & Douven, in submission; Experiment 3). In the latter experiment, the conditional probability was assessed using a probabilistic truth table task (e.g., Evans, Handley, & Over, 2003). This experiment used abductive or diagnostic conditionals, e.g., *if the car starts, the key was turned*. Assessing explanatory quality involved participants being told that the antecedent was a fact, they were then asked to judge how good the consequent was as an explanation.

These findings on inferentialism question the suppositional account of the conditional. However, as we now argue, they may be consistent with the CBN approach.

CBNs and Inferentialism

As we have argued (Oaksford & Chater, 2010, 2017), the psychology of reasoning can benefit from exploiting connections with causal learning/reasoning and decision-making, as David and Ken Manktelow suggested some 30 years ago. In constructing a CBN, there are always two questions that we need to address, how do we infer the *structure* of the dependencies involved, and how do we infer their *strength* (Griffiths & Tenenbaum, 2005)? Answering these questions may require different information.

We propose to generalise the application of CBNs to the range of constraints proposed by Barwise and Perry (1983) to underpin the meaning of all conditionals, not just causal conditionals. Our mental representations of structure, that is, whether we include a directed link or not, depends on our belief that there is a counterfactual supporting relation in the world. It is this belief that the relation exists that determines whether we believe that the conditional is true, whether the relation is a physical cause, a disposition in someone's head (a psychological cause), or a social convention (a psychosocial cause). The links in the network are directional. Most often direction is given by causal direction, but it may also come from the order of condition-then-action pairs in instrumental behaviour based on our habits and dispositions.

Causal power determines the inferential strength of this relation, where we use "causal power" neutrally to refer to Equation 1. Causal power embodies the definition of inferential relevance according to the probabilistic view of inferentialism as ΔP is the numerator of causal power and hence causal power tracks inferential relevance. This much is obvious simply from the definition of causal power (Eq. 1). The information that underpins our belief in a dependency, and hence whether we include such a link in our CBN representation, is different from the information underlying the strength of the dependency. Moreover, the surface form of the conditional does not determine the direction of inferential links in a CBN (Ali et al., 2011). However, structure and strength can be intimately related. A ΔP value greater than zero may lead us to suspect that a dependency exists, but since it is only a measure of covariation and it could not guarantee that one exists. We demonstrate the independence of the sources of information underlying structure and strength judgements using an example of a spurious cause.

Spurious Causes

Take, for example, a spurious cause expressed as a conditional

If the cock crows, then the sun rises (1)

(1) can be construed as an inductive inferential conditional based solely on statistical observation (Krzyżanowska, Wenmackers, & Douven, 2013), that is, it is a learned association. Alternatively, we could construe it as an abductive conditional reversing the order of cause and effect in antecedent and consequent. One thing is sure it does not describe a counterfactual supporting dependency in the world because *if the cock had not crowed the sun would not have risen* is surely false. So, according to the current thesis, we should not add an inferential link to our stock of beliefs, at least not one in the direction *cock crow* → *sun rises*.

However, if (1) is judged solely based on statistical information, then we may end up concluding that such an inferential dependency exists. These events are constantly conjoined, and in people's experience, the cock crowing immediately precedes the sun rising, so the probability that the sun rises given that the cock crows, $\Pr(q|p) = 1$. According to probabilistic inferentialism, this is a spurious causal connection presumably because the probability that the sun rises given the cock doesn't crow, $\Pr(q|\neg p) = 1$ and so $\Delta P = 0$. However, this calculation is not without problems.

Calculating causal power. In the real world, it is difficult to know on what data people could calculate $\Pr(q|\neg p)$. In laboratory experiments on causal learning, participants are provided with the relevant contingency table with four cells *a*-*d*, where $a = \Pr(p, q)$, $b = \Pr(p, \neg q)$, $c = \Pr(\neg p, q)$, and $d = \Pr(\neg p, \neg q)$, and $\Pr(q|\neg p) = c/(c + d)$. But in people's experience, or at least in the experience of somebody contemplating (1), the sun has never risen without the cock crowing, $\Pr(\neg p, q) = 0$. Moreover, because these are rare events (Oaksford & Chater, 1994), most of the time the sun is not rising, and the cock is not crowing, the *d* cell it is presumably very large (Hattori & Oaksford, 2007). Consequently, people's experience seems to indicate that $\Pr(q|\neg p) \approx 0$, and hence $\Delta P > 0$. That is, like causes, it may be that acquiring inferential dependencies solely from experience using covariational measures could lead people to endorse spurious causes and possibly missing-link conditionals.

In the causal learning literature, participants significantly underweight the *d*-cell (e.g., Arkes & Harkness, 1983; Jenkins & Ward, 1965; Kao & Wasserman, 1993; Schustack & Sternberg, 1981) rendering it difficult to assess $\Pr(q|\neg p)$. Hattori and Oaksford (2007) proposed a dual-factor heuristic, *H*, to get around this problem. *H* is based on a limiting case of the normative Phi-coefficient as the probability of the *d*-cell tends to infinity. *H* is equivalent to the geometric mean of $\Pr(q|p)$ and $\Pr(p|q)$ ($H = \sqrt{\Pr(q|p).\Pr(p|q)}$). Hattori and Oaksford (2007) showed that this measure provided a better fit to judged causal strength than ΔP and forty other measures of covariation. Note, however, that because *p* and *q* are constantly conjoined in our example, $\Pr(q|p) = \Pr(p|q) = 1$, and therefore $H = 1$. So relying only on these statistical measures, we would conclude that cocks crowing causes the sun to rise. Consequently, we should predict that people will endorse this spurious causal conditional. There would appear to be more to establishing a causal/inferential dependency than covariation detection using ΔP or causal power.

Real or imagined interventions and mechanism. In causal induction, covariation detection may only be the first stage in inferring a causal dependency from data (Hattori & Oaksford, 2007). More generally, Pearl (2000, pp 252-253) has argued that,

> "Because of the circularity inherent in all definitions of causal relevance (see section *CBNs and Conditionals*), probabilistic causality cannot be regarded as a program for extracting causal relations from temporal/probabilistic information; rather, it should be viewed as a program for validating whether a proposed set of causal relationships is consistent with the available temporal-probabilistic information."

The proposed causal or inferential dependency may derive from covariational information, but to be a useful habit of inference, we need to go beyond this source of information. An important further source of information is *intervention*, whether in reality or imagination (Lagnado, Waldmann, Hagmayer, & Sloman, 2007). So, we intervene in the world by conducting an experiment, in which we prevent the cock from crowing and see whether the sun rises. This experiment should dispel the idea of a causal dependency. However, we may only be able to conduct the experiment with imagination.

Consider, the converse of (1), *if the sun rises, the cock crows*. We cannot conduct the experiment in which we prevent the sun from rising and see whether this leads to the cock not crowing. We can only carry out this intervention in imagination. The asymmetry with (1) depends on intuitions about possible mechanisms which connect the two events. For the converse, the changes in lighting could cause the cock to crow, which is why we believe that *if the sun had not risen, the cock would not have crowed* is true. But how the cock's crow could cause the sun to rise is, at least to modern thinkers, mysterious. Today we have more mechanistic, and less magical, ways of thinking about how the world works.

In the mental representation of this situation, people include the directional link from cause to effect, as in the converse of (1). (1) carries useful information because of our belief in the existence of the actual causal dependency which we represent as a direct inferential link (*sun rises* → *cock* crows) and which is not inferred from statistical evidence alone. One can interrogate the CBN representation in any order, which means that the inference to the sun rising from hearing the cock crow is an inference to the best explanation. The apparently non-causal *if it is raining in Ipanema, then John has toast for breakfast*, raises other issues.

Habits and methodological policies. A mechanistic causal linkage between the weather and what John has for breakfast seems unlikely. However, these events can be linked intentionally by John's habitual patterns of behaviour. John may use his favourite weather app to decide to eat toast for breakfast in New York only if it is raining in Ipanema. Such a conditional seems to support the corresponding counterfactual. We would tend to evaluate *if it had not been raining in Ipanema this morning, then John wouldn't have had toast for breakfast* as true. The law-like dependency in the world in this instance is John's methodological policy to change his beliefs about what he will have for breakfast, dependent on the current weather in Ipanema. We can also actually observe John's breakfast eating behaviour contingent on the weather in Ipanema, and check that if it is not raining, he does not eat toast.

More generally the inclusion of a causal link in a CBN often derives from other sources of information:

> "The second valuable source of causal knowledge is linguistic advice: explicit causal sentences about the workings of things which we obtain from parents, friends, teachers, and books and which encode the manipulative experience of past generations. As obvious and uninteresting as this source of causal information may appear, it probably accounts for the bulk of our causal knowledge, and understanding how this transference of knowledge works is far from trivial." (Pearl, 2000, 252-3).[3]

[3] These two quotations from Pearl have been cited as evidence against a CBN approach (Stenning & Vam Lambalgen, 2016). However, the alternative proposal seems to involve treating world knowledge as a globally "consistent" (in a three valued logic) default theory or data base. However, the psychological evidence seems to suggest the human knowledge is unlikely to be organisd in this way as it is highly fragmented, incomplete, and inconsistent (Rozenbilt & Keil, 2002; Oaksford & Chater, 1991; Sloman & Fernbach, 2017). People resolve inconsistencies by public argument and debate in the social domain, rarely does this happen, or have to happen, in isolated individuals (Oaksford & Chater, 2016, 2020).

A lot of causal knowledge is conveyed implicitly, by our use of words like "because," "in spite of'," "nonetheless," and so on. Of course, conditionals also convey this knowledge. Often the best explanation of the presence of an inferential connection between antecedent and consequent is the presence of a causal connection between them.

In building a CBN, we must take into account both structure and strength (Griffiths & Tenenbaum, 2005). We base the inclusion of specific links encoding putative dependencies on different sources of information than inferring causal power. In the case of John's breakfast habits, a reliable source may tell us that this is indeed one of his methodological policies for changing his beliefs. We then expect his overt behaviour to conform to this dependency such that we would be willing to evaluate the corresponding counterfactual as true. Only then would we be willing to include this as one of our own methodological policies for changing our beliefs about what John will eat for breakfast.

Summary. Our grounds for including a structural link ($p \rightarrow q$) between two variables in a CBN depend on a variety of factors. We may consider these factors before or after we consider statistical information about the strength of the link. We may (i) be told about a disposition or habit of inference of John's, or (ii) we may observe the statistical regularity relating cocks crowing to the sun rising. Both may suggest including a directed link encoding an inferential dependency between the two variables. For (i), including the link may depend on the reliability of the source of this information. To determine the strength of the link, we want to make some observations of John's behaviour. Certainly, we would want to see that it not raining in Ipanema leads to at least a lowering of the probability that John's eats toast for breakfast. For (ii), we would want to intervene and prevent the cock from crowing and see whether the sun rose. Or we could do this in imagination which would involve seeing if we could imagine an appropriate mechanistic link between these events.[4] For the converse, this is all we can do, but the appropriate link seems clear, cocks can detect early signs of a sunrise as they have a perceptual system. There is a mechanism by which the sun rising can cause the cock to crow.

Inferentialism and CBNs

As in inferentialism, CBNs embody the idea that conditionals describe inferential dependencies. In CBNs these are represented as directed links. Our mental representations are directional (Evans, 1993; Oberauer, Hornig, Weidenfeld, & Wilhelm, 2005; Oberauer & Wilhelm, 2000), but they follow causal/explanatory lines, not logical lines (Ali et al. 2011). Assessing the truth or acceptability of a conditional involves judging both structure and strength. But these are different judgements, based on different information, which may affect judgements of truth and probabilistic acceptability differently. Sometimes, as we have seen, calculating strength alone may be impossible due to problems in calculating $\Pr(q|\neg p)$ or with the inability to intervene in the world to prevent *p*, which could substitute for $\Pr(q|\neg p)$ in this assessment. Determining the causal/inferential link in imagination requires us to contemplate what the relevant mechanism could be, and in modern times, this probably rules out divine or magical accounts, but other people's habits and dispositions—their methodological policies for changing their beliefs (Stalnaker, 1984)—are fine. Only when we are happy that there are no missing links can we exploit, via assignments of causal/inferential strength, the representational and inferential power of CBNs. If, as in causal model theory (Sloman, 2005; Sloman, Barbey, & Hotaling, 2009), we

[4] If, of course, we allow God into the process, then, in a similar vein to Berkeley, we could reduce (ii) to (i): there may be an appropriate directional methodological policy linking these two events in the mind of God. When omnipotent or magical powers can be invoked, any link may be justified.

treat small-scale CBNs as the mental mechanisms (Chater & Oaksford, 2006) of conditional inference, then we have provided a formalism and theory of mental representation that can underpin inferentialism.

We have already observed that because we parameterise inferential links with causal power, they must track inferential relevance as defined in probabilistic inferentialism (e.g., Skovgaard et al., 2016). Moreover, the and-to-if inference, which is probabilistically valid because Pr(*if p, then q*) ≥ Pr(*p*,*q*), does not seem to hold other than in conditions of positive relevance, $\Delta P > 0$ (Skovgaard et al., 2016). This inference, also called "centering" (Cruz, Over, Oaksford, & Baratgin, 2016), holds in almost all logics. However, it seems to be at the root of the problem with missing-link conditionals. Centering means that from the conjunction of any two true assertions, for example, "the moon is not made of cheese" and "the earth's circumference is 24,901 miles", we can infer a conditional "if the moon is not made of cheese, then the earth's circumference is 24,901 miles." We could reject this inference on pragmatic grounds, that is, centering is valid, but in missing link cases, they are not assertable because they violate discourse coherence. However, empirically it would appear that people do not treat discourse coherence alone as yielding an assertable conditional in the absence of probabilistic inferential relevance (Krzyżanowska, Collins, & Hahn, 2017). The pragmatic response is also questionable from arguments that the semantics/pragmatics distinction is a difficult one to enforce (Bach, 1999; Levinson, 2000). Centering is also in doubt if we only represent a conditional as a directed link in a CBN assuming that a dependency exists.

However, there is a problem with denying centering while advocating the representation of conditionals as directed links in a CBN. It would appear, in general, that the logic induced over Pearl's (2000) structural equation models respect centering, because axiomatically they are equivalent to Lewis's (1973) logic for the counterfactual. That such an axiomatization is possible is important to show, in general, how the interpretation of the conditionals interacts with the rest of language. In Pearl's structural models, links are only causal if *q given do p* and *not-q given do not-p* must hold. Only then can we interpret *if p then q* causally. The *do* operator cuts any links between *p* and its parents, implementing Lewis' "small miracles." However, psychologically we are not concerned with the general case. To draw inferences, we have conjectured that people only construct small-scale local models of premises in which they interpret all the links as real dependencies or constraints (Oaksford & Chater, 2010, 2013, 2020). We are not concerned with general rationality constraints on our all our beliefs, just the local model, because, generally, our beliefs are shallow, fragmented and inconsistent (Oaksford & Chater, 1991; Rozenblit & Keil, 2002; Sloman & Fernbach, 2017). Ultimately this is a psychological question, to quote Sloman and Lagnado (2005) "do we do" when we assert a conditional and when we believe a conditional do we believe probabilistic relevance holds? Indeed, a programme of research is suggested, using conditionals expressing a range of relations, for example, habits, dispositions, norms, social conventions and so on, to see whether they also produce inferences similar to the causal case.[5]

A further finding on probabilistic relevance was that when told that the antecedent and consequent of a negative relevance conditional—*if you hit the brakes, the car will speed up*—are both true participants judge the conditional true (Skovgaard-Olsen, Kellen et al., 2017). These results seem to contradict the semantic account of inferentialism (Douven et al., 2018). However, there may be no contradiction. We may base the truth evaluation on different mechanistic

[5] For utility conditionals, where utilities would need to be added, we already know that causal connections are important (Bonnefon & Sloman, 2013; Elqayam, Thompson, Wilkinson, Evans, & Over, 2015).

information than the statistical information used to judge negative relevance. Inferential dependencies or reason relations are radically context-sensitive; for example, people's methodological policies for changing their beliefs are relative to a particular individual. Although in the normal course of events this is a negative relevance conditional, that this particular car speeds up when you hit the brakes, must mean that its internal mechanisms are set up differently to normal cars, such that, hitting its brakes provides a good reason to expect it to speed up.[6] In statistically judging this to be a negative relevance conditional, people must think of the general or usual case. In Douven et al.'s (2018) experiments, it is quite clear that participants must evaluate truth only with respect to the specific soritical series they investigated.

The factors we have considered are consistent with the focus on explanatory quality in semantic inferentialism. In Douven et al. (2018), the soritical series provides a mechanism by which less separated patches are bound to be more similar in colour. Moreover, the vast majority of explanations are causal and require some mechanism. Although the sun rising explains why the cock crows, the cock crowing could not explain why the sun rises, which requires to completely different mechanistic explanation. Once we have sorted out issues of possible mechanism, we can incorporate the directional links into a CBN model, and then, of course, all good explanations are inferentially upstream. Only parent nodes explain their children not the other way round, and this is a matter of structure, not strength. More recent results show that explanatory quality better predicts truth evaluations than conditional probability (Mirabile & Douven, in submission). As we have seen, conditional probability cannot settle all issues of whether an adequate inferential dependency exists, which requires consideration of possible

[6] This example differs from our example of *if the cock crows the sun rises*, where a mechanistic explanation is unavailable.

mechanisms and the ability to intervene in reality or imagination. So in truth evaluation, judging explanatory quality, and, we conjecture, in drawing inferences, we would expect people to consider structure as well as strength. However, when judging probabilities in the probabilistic truth table task, only strength is at issue.

Conclusions

We have argued that causal Bayes nets can provide an implementation of inferentialism that captures the core idea that conditionals describe inferential dependencies that are not present for missing link conditionals. We have suggested that spurious causes provide a useful illustration o where the conditional probability of the effect given the putative cause is high, but we doubt that preventing the cause from occurring (intervening) would reduce the probability of the effect occurring because of the lack of an appropriate mechanism. Only if we think there is a possible mechanism do we introduce a directional link to the CBN capturing the dependency. However, dependencies form a broad category, including our own and others' methodological policies for changing our beliefs that we can model in CBN's in the same way. They also include social norms and cultural conventions, which differ in that we assess them not for truth but for whether they are obeyed. However, given the large range of potential dependencies, it will be an interesting future project to investigate the extent to which they behave as a CBN implementation would predict. For example, would we see discounting behaviour for John's breakfast habits, if he also ate toast when it was sunny in Bondi?

Where does this leave us concerning our view of the suppositional conditional? It seems consistent to view the suppositional account as an overarching framework for the new paradigm in reasoning. In this, our view has not changed since we articulated our account of quantified syllogistic reasoning (Chater & Oaksford, 1999). There we treated these two premise arguments as defining a dependency graph capturing the idea that the end terms were independent given the middle term. This assumption went beyond probability theory and simplified our calculations of probabilistic validity for syllogisms. Similar assumptions are made when treating two premise conditional arguments as defining a CBN to model suppression effects (Fernback & Erb, 2013; Oaksford & Chater, 2017) or discounting (Ali et al., 2010, 2011). Consequently, if, as it would appear, people regard conditionals as describing inferential dependencies (Skovgaard-Olsen, Kellen et al., 2019), then this can be captured in the formalism of CBNs, which¸ by making additional assumptions, may lose some inferences, e.g., centering (psychologically all links in a local model are treated like causes), but gains others, e.g., discounting, (the causes in the model are the sole causes of the common effect). These additional assumptions go beyond the pure suppositional account but the resulting inferences are consistent with a general Bayesian probabilistic approach.

References

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional inference: Causal models or mental models. *Cognition*, *119*, 403-418.

Ali, N, Schlottmann, A., Shaw, C., Chater, N., & Oaksford, M., (2010). Conditionals and causal discounting in children. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals*: *Probability and logic in human thinking* (pp. 117-134). Oxford: Oxford University Press.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*, 117–135.

Bach, K. (1999). The myth of conventional implicature, *Linguistic and Philosophy*, *22*, 327-366.

Barwise, J., & Perry, J, (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

Cartwright, N. (1983). *How the laws of physics lie*. Cambridge: Cambridge University Press.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*, 191-258.

Chater, N., & Oaksford, M. (2006). Mental mechanisms. In K. Fiedler, & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 210-236). Cambridge: Cambridge University Press.

Chisolm, R. (1946). The contrary-to-fact conditional. *Mind*, *55*, 289–307

Cruz, N. (2018). *On the Role of Deduction in Reasoning from Uncertain Premises*. PhD Thesis. Department of Psychological Science¸ Birkbeck College, University of London.

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, *6*, article 192. http://dx.doi.org/10.3389/fpsyg.2015.00192

Cruz, N., Over, D. E., & Oaksford, M. (2017). The elusive oddness of or-introduction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 259–264). Austin, TX: Cognitive Science Society.

Cruz, N., Over, D. E., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1104–1109). Austin, TX: Cognitive Science Society.

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–81. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.09.002

Douven, I., & Mirabile, P. (2018). Best, second-best, and good-enough explanations: How they matter to reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1792–1813. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000545

Edgington, D. (1995). On conditionals. *Mind, 104*, 235–329.

Elqayam, S., Thompson, V. A., Wilkinson, M. R., Evans, J. StB. T., & Over, D. E. (2015). Deontic introduction: A theory of inference from is to ought. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1516–1532. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0038686

Evans, J. St.B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, *48,* 1–20.

Evans, J. St.B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–355.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1327–1343. doi:10.1037/a0031851

Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, *44*, 113–128.

Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.

Hall, S., Ali, N., Chater, N., & Oaksford, M. (2016). Discounting and augmentation in causal conditional reasoning: Causal models or shallow encoding. *PLoS One*, *11*(*12*): e0167741. doi: 10.1371/journal.pone.0167741

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for co-variation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, *31*, 765-814.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*, 1–17.

Kao, S.-F.,&Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363–1386.

Krzyżanowska, K., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs Probabilistic relevance. *Cognition*, *164*, 199–205. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2017.03.009

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and evidentiality. *Journal of Logic, Language and Information*, *22*, 315–334. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/s10849-013-9178-4

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154-172). New York, NY, US: Oxford University Press.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297–315.

Mirabile, P., & Douven, I. (2019). Abductive conditionals as a test case for inferentialism. *Manuscript in submission.*

Oaksford, M. (1989). *Cognition and inquiry*. PhD Thesis. Centre for Cognitive Science¸ University of Edinburgh.

Oaksford, M. & Chater, N. (1991). Against logicist cognitive science. *Mind & Language*, *6,* 1-38.

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

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2010). Causation and conditionals in the cognitive science of human reasoning. *Open Psychology Journal*, *3,* 105-118. (Special issue on *Causal learning beyond causal judgment*, Edited by J. C. Perales, & D. R. Shanks).

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking and Reasoning, 19,* 346-379.

https://doi.org/10.1080/13546783.2013.808163

Oaksford, M., & Chater, N. (2014). Probabilistic single function dual-process theory and logic programming as approaches to non-monotonicity in human vs. artificial reasoning. *Thinking and Reasoning*, *20*, 269-295. doi: 10.1080/13546783.2013.877401

Oaksford, M., & Chater, N. (2016). Probabilities, causation, and logic programming in conditional reasoning: A reply to Stenning and van Lambalgen. *Thinking and Reasoning*, *22*, 336-354. https://doi.org/10.1080/13546783.2016.1139505

Oaksford, M., & Chater, N. (2017). Causal models and conditional reasoning. In M. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 327-346). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*, 12.1-12.26. https://doi.org/10.1146/annurev-psych-010419-051132

Oaksford, M., Over, D., & Cruz, N. (2019). Paradigms, possibilities, and probabilities: Comment on Hinterecker, Knauff, and Johnson-Laird (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 288–297. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000586

Oberauer, K., Hornig, R., Weidenfeld, A., & Wilhelm, O. (2005). Effects of directionality in deductive reasoning II: Premise integration and conclusion evaluation. *Quarterly Journal of Experimental Psychology*, *58A*, 1225–1247.

Oberauer, K., & Wilhelm, O. (2000). Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *26*, 1702–1712.

Over, D. E. (2009). New paradigm psychology of reasoning: Review of *Bayesian Rationailty*. *Thinking & Reasoning*, *15*, 431–438. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546780903266188

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

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

Pharo, P. (2007). How is sociological realism possible? Sociology after cognitive science. *European Journal of Social Theory, 10*, 481-496. https://doi.org/10.1177/1368431007080708

Recanati, F. (1988). The pragmatics of what is said. *Mind and Language, 4,* 295-329.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521-562.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110*, 101–120.

Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., & Klauer, K. C. (2019). Cancellation, negation, and rejection. *Cognitive Psychology*, *108*, 42–71. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2018.11.002

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, *126*, 611–633. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000150.supp (Supplemental)

Skovgaard-Olsen, N., Kellen, D., Krahl, H., & Klauer, K. C. (2017). Relevance differently affects the truth, acceptability, and probability evaluations of ’and’, ’but’, ’therefore’, and "if–then." *Thinking & Reasoning*, *23*, 449–482. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2017.1374306

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2015.12.017

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and reason relations. *Cognitive Science*, *41*, 1202–1215. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1111/cogs.12462

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY, US: Oxford University Press.

Sloman, S., Barbey, A., & Hotaling, J. (2009). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science*, *33*, 21–50.

Sloman, S. A., & Fernbach, P., (2017). *The knowledge illusion: Why we never think alone*. New York: Riverhead.

Sloman S. A., & Lagnado, D. A., (2005). Do we do? *Cognitive Science*, *29*, 5-39.

Stalnaker, R. C. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Stenning, K., & van Lambalgen. M. (2016). Logic programming, probability, and two-system accounts of reasoning: A rejoinder to Oaksford and Chater (2014). *Thinking & Reasoning*, *22*, 355-368

Veltman, F. (1985). *Logics for conditionals*. PhD Thesis, University of Amsterdam.

Vigen (2015). *Spurious correlations*. Paris, France: Hachette Books.