

Causation and Conditionals in the Cognitive Science of Human Reasoning

Mike Oaksford^{1,*} and Nick Chater²

¹Department of Psychological Sciences, Birkbeck College, University of London

²Department of Cognition, Perceptual and Brain Sciences, University College London

Abstract: This article traces the philosophical and psychological connections between causation and the conditional, *if...then*, across the two main paradigms used in conditional reasoning, the selection task and the conditional inference paradigm. It is argued that hypothesis testing in the selection task reflects the philosophical problems identified by Quine and Goodman for the material conditional interpretation of causal laws. Alternative formal theories to the material conditional only became available with the advent of possible worlds semantics (Lewis, 1973; Stalnaker, 1968). The relationship proposed by this semantics between counterfactual and indicative conditionals is outlined and it is concluded that moving away from the abstractions of possible worlds proposes a central role for prior knowledge in conditional inference. This conclusion is consistent with probabilistic approaches to conditional inference which provide measures of the strength of a dependency between the antecedent and the consequent of a conditional similar to those proposed in causal learning. Findings in conditional inference suggest that people are influenced not only by the strength of a dependency but also by the existence of the structural relationship, the broader causal framework in which a dependency is embedded, and the inhibitory and excitatory processes like those required to implement Causal Bayes nets or neural networks. That these findings may have a plausible explanation using the tools of current theories in causal learning suggests a potentially fruitful convergence of research in these two areas.

Keywords: Conditionals, Causation, Counterfactuals, Possible Worlds, Causal Bayesian Networks, Conditional Reasoning, Psychology of Reasoning, Causal Learning.

INTRODUCTION

The conditional construction, “if p then q ” in English, occurs in all human languages [1] and allows people to express their knowledge of the causal or law-like structure of the world and of others’ behaviour, e.g., *if you turn the key the car starts, if John walks the dog he stops for a pint of beer*; to make promises, e.g., *if you cook tonight, I’ll wash up all week*; to regulate behaviour, e.g., *if you are drinking beer, you must be over 18 years of age*; to suggest what would have happened had things been different, e.g., *if the match had been dry it would have lit*, to explore purely hypothetical possibilities, e.g., *if Kennedy had not been shot, then would the Berlin wall have fallen earlier*, among many other possible uses. The way in which the conditional is modelled also determines the core of most logical systems. Unsurprisingly, it is also the most researched expression in the cognitive science and psychology of human reasoning.

Over the last 20 years, some of the most important developments, both theoretically and empirically, have occurred in the area of *causal conditional reasoning*. That is, studies of how people reason with conditional statements in which the antecedent, p , refers to a cause of an effect described in the consequent, q , e.g., *if the switch is flicked, the light comes*

on. Philosophically, understanding the concept of causation has involved analysing the *c caused e* locution in terms of the logic of conditionals. For the logical positivists and Popper a causal law could be analysed in terms of the *material conditional* of standard logic (see below). Moreover, Hume’s second counterfactual definition of causation—*c caused e* if and only if *if c had not occurred then nor would e*—has been analysed using possible worlds semantics (see below) for the corresponding counterfactual conditional by Stalnaker [2] and by Lewis [3]. More recently, causal Bayes’ nets/ structural equation models developed in AI [4, 5] have been used in philosophy to model counterfactual conditionals [6-9]. This is in the spirit of earlier probabilistic analyses of causation [10, 11]. There have also been corresponding probabilistic treatments of the conditional [12-14] which have been applied in the psychology of reasoning [15-19].

Our argument in this paper is that by examining the conceptual relationship between causation and the conditional, the history of the psychology of conditional reasoning can be re-interpreted in a much more rational light. We begin by first discussing early data on hypothesis testing where the hypothesis under test is formulated as a conditional statement. The first experimental findings on the psychology of conditional reasoning used tasks like this [20] and produced results widely interpreted at the time as demonstrating human irrationality. We argue that by considering appropriate causal interpretations of these rules and the factors that affect them, such premature judgements may have been avoided. We then show how similar factors have subsequently been

*Address correspondence to this author at the Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London, WC1E 7HX; Tel: 020 7079 0879; Fax: +44 (0)20 7631 6312; E-mail: mike.oaksford@bbk.ac.uk

investigated in the conditional inference paradigm. We then consider the relevance of more recent possible worlds approaches to the semantics of the conditional and how they relate to these data. Finally, we consider more recent probabilistic approaches to the conditional and their relationship to recent Bayes' net approaches to causation. Here we argue that considering this relationship allows a re-interpretation of the data using the causal conditional inference paradigm in a way that may resolve the apparent tension between logic based and probability based approaches.

1. Conditionals, Causation, and Hypothesis Testing

Some of the earliest work on the conditional did not directly address the inferences one could draw on accepting a conditional but on the evidence one needed to select to determine whether to accept a conditional in the first place, i.e., it addressed *causal induction* rather than *causal inference*. In Wason's selection task [20], which seemed to explicitly require participants to employ their knowledge of the logic of the conditional to test a hypothesis, participants are shown four cards which have a number printed on one side and a letter printed on the other (and participants know this). They are told that these cards obey the rule that *if there is an A on one side then there is 2 on the other side*. They are then shown four cards of which they can only see the uppermost side. One card shows an A, one shows a K, one shows a 2, and the final card shows a 7. They are then asked to select the cards that they must turn over to determine whether the rule is true or false. On the logical falsificationist strategy, participants should select the A and the 7 card as either could be a falsifying instance of the form p is true but q is false. This is because according to standard truth functional logic, a conditional *if p then q* , is only false if p is true and q is false otherwise it is true. This truth functional interpretation is called the "material conditional." However, participants tend to select the A and the 2 card or the A card alone, which seems to reveal a preference for confirmatory data. This result was interpreted as indicating that people possess a "confirmation bias," in contrast to the dictates of Popper's [21] falsificationist view of science based on the truth functional interpretation of the conditional.

The selection task was originally introduced as a problem in hypothesis testing [22], although it has been re-interpreted as an inferential problem (e.g., [23]). However, in its original guise there were already good reasons to question the task's core assumption that the material conditional provides an adequate interpretation of scientific laws. In the philosophy of science during the 1950s, the material conditional interpretation of scientific laws and falsificationism had been shown to be highly problematic.

One particularly important source of problems stemmed from the Quine-Duhem thesis [24, 25]: that a scientific hypothesis cannot be decisively falsified by data, however apparently damning, because the data can always be explained away by adjusting so-called auxiliary hypotheses (e.g., about other forces that may be acting; about the operation of the data collecting equipment, and so on), and leaving the theory under test intact. For example, early predictions concerning the orbit of Uranus relied on the auxiliary assumption that there were only seven planets. The failure of Newton's celestial mechanics to successfully predict Uranus' orbit, should

according to the falsificationist logic, have led to the rejection of Newtonian theory. However, the discovery of an eighth planet, Neptune, which turned out to be influencing Uranus' orbit, showed that it was the auxiliary assumption about the number of planets that should be rejected, not Newtonian mechanics.

The fact that prediction from scientific laws requires a body of unstated further assumptions also figured in Goodman's [26] critique. Goodman's work focussed on two issues. First, he noted that scientific laws must also satisfy Hume's second definition of causation, which is counterfactual, i.e., *if the cause had occurred, then the effect would have occurred*. For example, if the match had been struck, it would have lit. Goodman noted that the truth of such a statement cannot be captured by the logic of the material conditional, since the antecedent of a counterfactual is always false and so, on this analysis, counterfactual conditionals are always true. But clearly *if the match had been struck, the moon would have been made of cheese* is simply not true. Moreover, statements like, *if the match had been struck, it would have lit*, also depend on a range of auxiliary assumptions, e.g., that there is oxygen present, the match is struck sufficiently hard, the match is not wet, and so on. These assumptions have to be what Goodman described as "cotenable" with the antecedent of the counterfactual before its truth can be affirmed. "Co-tenable" conditions are those which taken together with the antecedent would logically entail the consequent of a conditional (see also [27], on the "enthymematic basis" of a conditional).

However, there is a problem with cotenability theories as a definition of the counterfactual conditional. For a condition to be co-tenable seems to depend circularly on the truth of another counterfactual, e.g., *if the match had been dry, it would have lit* (but see [28]). These problems aside, such a cotenability account can also be applied to the standard "indicative" conditional. When used to describe a law, the antecedent of a conditional is perhaps better characterised as the conjunction of the proximate cause with the cotenable conditions: *if the match is struck and it is dry and there is oxygen...etc., then it lights*. As with our example concerning planetary orbits and Newtonian mechanics above, the observation of a struck match not lighting on some specific occasion does not falsify the hypothesis that it should light, because this evidence most likely bears on the cotenable conditions and not on the law-like relation itself.

The second issue Goodman also noted was that even if cotenability accounts of the counterfactual could not be made to work, it is still the case that, intuitively, good scientific laws should in any case be counterfactual supporting. So, we believe that if the match *had* been struck, it *would* have lit, when, all other things being equal (i.e., all cotenable conditions are in place), you believe that "if a match is struck it lights" describes a real, i.e., causal, relation in the world. However, the material conditional analysis cannot guarantee this as it says nothing about such relationships. Consequently, such an analysis cannot distinguish between counterfactual supporting conditionals that describe causal laws, like "if a match is struck it lights," from non-counterfactual supporting accidental generalisations, like "if this is a coin in my pocket today, then it is silver." Via his famous "grue" problem, Goodman argued that this difference could not de-

pend on any superficial characteristic, like temporal specificity (indexing to “today”). Rather, the difference was an issue of *content*, i.e., the nature of the predicates used in a law-like statement, not of the form of the conditional. Predicates like “being a struck match” have a history of being used to make predictions about future events, it is “projectable.” However, predicates like “being a coin in my pocket today” have no such history and consequently are not projectable. A non-circular account of projectability, which does not merely define projectable predicates as those that reliably support generalizations is, though, difficult to come by.

Goodman also pointed out that adopting the material conditional creates problems for the nature of the confirmation relation between evidence and a hypothesis about a causal law. The logical positivist position on confirmation could be viewed as, in a sense, the reverse of deduction. Thus, according to this viewpoint, predictions follow from a scientific theory by processes of deduction (or, rather, these predictions follow from the theory, conjoined with auxiliary hypotheses concerning initial conditions of the system under study, the operation of the measuring instruments with which data is gathered, and so on). To the extent that these predictions, derived by deduction, are correct, then the theory is viewed as receiving “inductive confirmation.” One set of difficulties for this perspective concerns the troublesome auxiliary hypotheses. Just as auxiliary hypotheses can be modified to save a theory, when its predictions appear to go astray, similarly it is not clear how much credit should be assigned to the auxiliary hypotheses, rather than the theory itself, when predictions are observed (the “credit apportionment” problem [29]).

Another potential source of concern is Hempel’s [30, 31] well-known “ravens paradox.” If a theory or hypothesis is confirmed when its deductive consequences are observed to be correct, then the statement *All ravens are black* (or, if you wish to frame this sentence in terms of the conditional, *if something is a raven, then it is black*, which is equivalent, according to the material conditional) must presumably be confirmed by the observation of a raven that is black (the deductive consequence is: suppose we observe a raven; then the theory predicts that it should be black; and so it is). But *All ravens are black* is, of course, according to the material conditional at least, equivalent to *All non-black things are non-ravens* (or, to again give the equivalent sentence in terms of conditionals, *if something is not black, then it is not a raven*), which is surely, by the same logic, confirmed by the observation of one of *its* instances—e.g., a white sock (once the object has been observed to be white, the hypothesis implies that it is not a raven; and indeed it is not). But according to some minimal assumptions at least, it seems difficult to escape the strange conclusion that observing a white sock confirms the hypothesis that *all ravens are black*, rather than being entirely irrelevant to it.

And indeed there may be an even more direct challenge to the idea that confirmation can be viewed as the inverse of deduction—cases where an instance of a generalization can actually disconfirm it! Howson and Urbach [32], for example, consider the following type of case. Consider the generalization *All beetles of species X are found in Ecuador*. Observing such a beetle, in Ecuador, but a matter of a few me-

ters from the border with Peru, would surely throw the generalization into severe doubt. If the beetles have spread this far, we would naturally argue, then surely they must have spread across the border too. But if this is right, then an observation which is in line with the deductive consequences of a theory (we observe the beetle; predict, using the hypothesis, that it should be in Ecuador; and duly observe that, yes, indeed it is), can *disconfirm*, rather than confirm the hypothesis. The more general problem is that questions of confirmation or disconfirmation of the conditional is not a matter purely of their truth functional properties. That is, it is not important merely that a particular piece of data is *consistent* with a conditional; the nature of the data, the nature of the claim made by the conditional, and our relevant background knowledge, all seem to contribute to determining whether or not an apparent instance of the conditional generalisation serves to make that generalisation more or less plausible. But this viewpoint only make sense on the assumption that conditionals do much more than make truth functional claims over a domain of possible instances—rather, they seem to make claims about the laws by which the world is governed (in the terminology of the philosophy of language, their claims are intensional, rather than purely extensional).

In the light of these problems, the assumption that experimental participants should seek falsifying evidence in the selection task could have been questioned on sound philosophical grounds at the time of Wason’s original studies. There were some voices of discontent [33] with the interpretation that people were behaving irrationally but they were in the minority. The general consensus was that this behaviour on the selection task was a manifestation of a general “confirmation bias,” a bias that was to be found again and again in experiments on human cognition (see [34] for a review).

The bulk of the early research on the selection task was done using abstract alphanumeric stimuli. The reason for this was straightforward: the idea was to examine the phenomena unpolluted by world-knowledge. Similar strategies had been pursued in a variety of cognitive domains, such as memory (paired associate learning, [35]) and problem solving (tower of Hanoi, [36]), because of the concern that exponent results could be driven almost entirely by general world knowledge, rather than experimental manipulations of interest [37]. However, a result that, with the benefit of hindsight, one might argue is consistent with Goodman’s analysis, was that content did seem to matter. For example, Wason and Shapiro ([38]; see also, [39]), used rules like, *If I travel to Manchester, then I take the train*, which led to considerably greater falsification rates in the selection task. The interpretation put on this behaviour was that it is due to familiarity with the content [40]. Such an interpretation suggests that different hypothesis testing strategies are adopted, depending on how familiar people are with the content used in the putative law-like relation. Again, with hindsight, this is consistent with Goodman’s observation that law-like relations, even everyday ones, rely on their content being “projectable,” i.e., having a history of being projected in one’s linguistic community. *If I travel to Manchester, then I take the train* seems to describe a sensible counterfactual supporting disposition, i.e., it supports the claim *if I had travelled to Manchester, I would have taken the train*. This contrasts with the situation in the standard, abstract selection task. The rule *if there is a an A on one side of the card, then there is a 2 on the other*

side of the card, which does not seem to support the counterfactual claim that if there *had* been an A on one side, then there *would* have been a 2 on the other. In the later case, there would appear to be no mechanism or process that can guarantee this, which contrasts with the behavioural disposition in the train example.

What we hope to have demonstrated in this section is that if a causal interpretation of a conditional hypothesis is adopted then we gain a far better understanding of the early results on hypothesis testing behaviour using the Wason selection task. The problems Goodman, Quine and others identified in the philosophy of science stemmed from interpreting these conditional hypotheses as material conditionals of standard logic. Our understanding may be enhanced further by adopting a probabilistic interpretation of causality. Lindley [41] provided a Bayesian analysis that showed that if the properties described in the antecedent and consequent were rare, then confirmation “bias” was an optimal hypothesis testing strategy. Such a Bayesian analysis resolves most of the problems identified by Goodman, Quine and others for an account of hypothesis testing [32]. For example, on the assumption that properties are rare, things that are not black and not ravens do indeed confirm the hypothesis that ravens are black but because they are so common—almost everything is not black and not a raven—the amount of confirmation is tiny compared to a black raven. We proposed a similar Bayesian analysis to explain the experimental data on Wason’s selection task [42-44]. As for the philosophical problems, such an analysis also provided a far better account of the empirical results.

This probabilistic account addresses two further issues. First, the probabilities that attach to the antecedent and consequent are matters of *content*. They do not depend on the logical form of the conditional statement, *if p then q*, but on the properties described by the predicates in the antecedent, *p*, and consequent, *q*. Second, this account provides a probabilistic analysis of a causal dependency, i.e., when *p* is a cause, *c*, and *q* is an effect, *e*, then a dependency exists between them if and only if $P(e|c) \neq P(e)$, i.e., as long *e* is not statistically independent of *c*. Oaksford and Chater’s [42] analysis involved determining which data provided the maximal discrimination between the hypothesis that there is a dependency between *c* and *e*, and the hypothesis that they are independent. This probabilistic account of “*c* causes *e*” goes back at least to Salmon [10] and has formed the basis of many recent accounts of causal learning in cognitive science.

2. Causation and Conditional Inference

In the last section, we concentrated on the evaluation of conditionals and on the classes of evidence relevant to their confirmation rather than on their use in conditional *inference*. The truth conditions for the material conditional prescribe two inference rules: modus ponens (MP) and modus tollens (MT), and proscribe two fallacies: denying the antecedent (DA) and affirming the consequent (AC) (“ \neg ” = not).

$$\begin{array}{ll} \text{(MP)} & \frac{p \Rightarrow q, p}{\therefore q} \\ \text{(MT)} & \frac{p \Rightarrow q, \neg q}{\therefore \neg p} \end{array} \quad (3.1)$$

$$\begin{array}{ll} \text{(DA)} & \frac{p \Rightarrow q, \neg p}{\therefore \neg q} \\ \text{(AC)} & \frac{p \Rightarrow q, q}{\therefore p} \end{array}$$

These inference schemata read that if the premises above the line are true, then so is the conclusion below the line. Actually, the truth conditions for the material conditional prescribe many more inference rules than these, as we shall see, but these are the four most commonly investigated in the psychology of reasoning. So, in experiments where people are presented with these four inferences, the logical expectation is that people will endorse MP and MT but not DA and AC.

2.1. Inferential Asymmetries and Cotenability

The principal findings from early studies was that people endorse MT a lot less than MP and that they also endorse the fallacies AC and DA, although subsequent results showed that AC was endorsed more than DA [45, 46]. This “MP-MT inferential asymmetry” seems to have an immediate explanation in terms of Goodman’s cotenability account on the understanding that we adopt a causal interpretation of the conditionals used in these experiments. People know that an instrumental causal relation such as, if I strike the match, it lights, relies on the other unstated cotenable conditions, as we discussed earlier. However, we rarely take these into account when striking the match in the expectation that it will light, i.e., we are likely to *predict* that the match will light (MP), even in ignorance of the status of the cotenable conditions. But if one were told that the match does not light, it would be perverse to conclude that the match was not struck—that is, it would typically be inappropriate to draw an inference using MT. This is because, in everyday life, we expect matches to light only if they are struck and hence when one comments that the match did not light this is only surprising and worthy of comment if an attempt to strike it has been made. Indeed, this observation leads to the interesting and almost paradoxical conclusion that, on being told that the match did not light one can conclude that it was struck, and one can conclude this only on the basis of the rule to which the present case is itself a counterexample. We have argued that such an account of the MP-MT inferential asymmetry falls out of our probabilistic approach to conditional inference, see below [17, 19, 47].

2.2. Empirical Investigations of Cotenability and Alternative Causes

The initial finding of an MP-MT asymmetry can be explained by the factors suggested by a cotenability account. These factors have also been experimentally investigated directly [48-50], although the initial motivation behind some of these studies was not to investigate causal interpretations of the conditional. The primary motivation behind Byrne’s [48] study was to show that even valid inferences like MP could be affected by these factors, so arguing against the existence of formal syntactic rules of inference as part of the cognitive system, as argued by, for example, Rips [23]. Byrne [48] showed that MP could be effectively “suppressed” by the provision of additional pragmatic world knowledge, as a cotenability account would lead one to suspect. She used rule pairs like 3.2a and 3.2b:

$$\text{If it is sunny tomorrow, John will play tennis} \quad (3.2a)$$

If the baby sitter is on time, John will play tennis (3.2b)

Participants would then be told that it is sunny tomorrow and asked whether they would endorse the MP inference to the conclusion that John plays tennis. Participants endorsed the MP inference far less when both 3.2a and 3.2b were presented than when 3.2a was presented on its own. What such a “suppression effect” seems to show, in Goodman’s terms, is the potential failure of a cotenable condition. The fundamental difference between these cases and the types of case that Goodman considered is that 3.2a is not a scientific law or indeed law-like at all. It is a specific conditional tied to a particular space-time location, describing a particular behavioural intention. It could be backed up by a more enduring behavioural disposition, i.e., whenever it is sunny, John plays tennis, which one may be inclined to regard as counterfactual supporting. That is, if John has this disposition, then one might be inclined to view the claim that if it had been sunny yesterday, John would have played tennis as true.

People’s sensitivity to manipulations like Byrne’s [48] and the above example suggests that the factors discussed by Goodman with respect to scientific or causal laws may be of more general applicability. The same factors that affect scientific laws—additional “cotenable” conditions and the relationship between indicative conditionals and counterfactual conditionals—may apply to conditionals more generally [51]. The principal philosophical issue concerns the status of scientific, i.e., causal laws, behavioural dispositions, intentions, and the like. Where, contra Hume, we may be inclined to view causal laws as ontologically respectable, i.e., as a part of the physical world, we may perhaps be less inclined to treat behavioural dispositions and intentions in the same way. However, psychologically, with respect to how people think about the world and interpret conditional statements, this may not be a relevant distinction. People may be as inclined to project their habits of inference onto the world whether they relate to causes (as Hume suggested) or to behavioural dispositions. That is, the way people think about the world may incline them to be as realist about behavioural dispositions as they are about causes.¹ Moreover, inferentially they may be expected to behave in similar ways based on a common set of mental representations and processes. In this sense, causal conditionals become central to a psychological analysis of conditionals by providing the core examples that drive our intuitions about the factors that affect conditional inference as opposed to standard truth functional logic.

Initial studies in the psychology of reasoning demonstrating the impact of alternative or additional causes on human reasoning [48], mentioned these factors explicitly to participants. The manipulations in Romain *et al.* [52] and Byrne [48] were of this kind, i.e., the information was provided in materials like 3.2a and 3.2b. Cummins *et al.* [50] importantly showed that, for causal conditionals, the factors such as alternative and additional information can affect conditional inference even when it is implicit. Cummins *et al.* [50] pre-tested causal conditionals for possible additional or cotenable conditions and for possible alternative causes of the effect. Cummins *et al.*’s [50] methodology, and the focus

on causal conditionals, has come to dominate much of the empirical work on conditional inference over the last two decades and for good reason. As we have noted, the insights derived from the study of causal conditionals may extend much more broadly to provide a general account of the psychology of conditional inference.

2.3. Ramsey, Possible Worlds and Counterfactuals

The phenomenon of explicit and implicit “suppression effects” seems to show how Goodman’s philosophical observations about cotenable conditions have direct parallels in the inferential behaviour of participants in causal conditional reasoning tasks. The main problem for a cotenable account was the lack of a formal normative theory of inference like standard logic to provide an evaluative theory by which to judge people’s inferential performance as errorful or not (although see [28]). In logic and formal semantics, it is generally agreed that serious attempts to account for the meaning of the conditional of natural language, as opposed to the conditional of mathematics, did not really start until about 40 years ago with the Lewis-Stalnaker possible world semantics [2, 3, 53]. Moreover, the analysis of the counterfactual conditional provided by these accounts was explicitly advanced as a semantic theory of causal statements.

The intuition behind the possible worlds approach and most contemporary accounts of the conditional come from attempting to formalise Ramsey’s [54] famous test:

“If two people are arguing ‘if A will C’ and are both in doubt as to A, they are adding A hypothetically to their stock of knowledge and arguing on that basis about C...We can say that they are fixing their degrees of belief in C given A.” [54]

For Stalnaker, the Ramsey test amounts to the claim that in assessing a conditional, people first add the antecedent hypothetically to their current set of beliefs. They then make minimal adjustments to accommodate the new belief. They then consider whether the consequent of the conditional is true in their revised set of beliefs. To idealise, before adding the consequent, they will have a belief about every matter of fact and after adding the antecedent they will have a revised set of beliefs about every matter of fact. These epistemically ideal states are what Stalnaker refers to as “possible worlds.” In the statement of the Ramsey test, it is explicit that the truth or falsity of the antecedent in the actual world is unknown.

There are some aspects of the formal theory that are central to understanding Stalnaker’s [55] interpretation of this semantic theory. In possible worlds semantics, the proposition expressed by a sentence is the subset of possible worlds in which the sentence is true. The core of the theory as it applies to the conditional is the *selection function*. If we take a conditional, *if p then q*, the selection function, *s*, takes the set of worlds in which the antecedent, *p*, is true, which is written [*p*], as one argument, and the actual world, α , as another argument, $s([p], \alpha)$, and it returns the subset of [*p*] which is most similar to α . If this subset is non-empty and is included in the set of possible worlds in which the consequent is true, [*q*], then the conditional is true. For example, suppose someone asserts that:

¹ Indeed this interpretation remains causal, in the sense that our Folk Psychology regards these behavioural dispositions to be the mental *causes* of overt behaviour.

If it's sunny at Wimbledon today, John will be playing tennis. (3.3)

Being, say, in Bloomsbury, we don't know whether it's sunny in Wimbledon or not. Nonetheless, our interpretation of the antecedent is the set of all possible worlds in which it is sunny in Wimbledon. The selection function then picks the subset of these worlds that are most similar to the actual world. If in all of these worlds John is also playing tennis, then the original conditional will be regarded as true.

The factors that affect the selection function depend on the actual world. So, for example, a possible world in which it is sunny at Bloomsbury but there is a blizzard at Wimbledon is less like the actual world than one where it is sunny at both locations due to their physical proximity. This could be expressed as another conditional,

If it's sunny in Bloomsbury, then it's sunny in Wimbledon. (3.4)

In particular, the *specific* conditional (3.3) about what John does today may be underwritten by the general claim that:

If it is sunny at Wimbledon, John plays tennis (3.5)

3.5 describes one of John's enduring behavioural dispositions. Stalnaker refers to 3.4 and 3.5 as "open" indicative conditionals (where the truth or falsity of the antecedent may be unknown), which describe people's *methodological policies* to change their beliefs. The selection function is an abstract characterisation of people's dispositions to alter their beliefs in response to new information. It is these dispositions or methodological policies that determine how we alter our beliefs when the antecedent is hypothetically added as in the Ramsey test.

Central to Stalnaker's account is the relation between open indicative conditionals like 3.4 and 3.5 and counterfactual conditionals, like,

If it had been sunny at Wimbledon today, John would have played tennis. (3.6)

Our methodological policy in 3.5 gives us good grounds to believe 3.6 and if we do so it must be because we believe that 3.5 is a real behavioural disposition of John's. That is, it is causally responsible for John's behaviour. The relationship between 3.5 and 3.3 is simply that of a general rule to an instance respectively, so if 3.5 is one of our methodological policies to change our beliefs then we are bound to believe 3.3. As Stalnaker points out, our inclinations to believe specific claims like 3.3 and 3.6 depend not only on our methodological policies to change our beliefs (3.4 and 3.5) but also on our other factual knowledge. For example, if we know that *John has a broken leg* then we would not infer 3.6 based on 3.5. This behaviour is directly related to fact that the logical rule of *strengthening the antecedent* is not valid in Stalnaker's semantic theory.² Before the addition of the

extra information that John had a broken leg, the subset of worlds closest to the actual world in which it is sunny at Wimbledon, $s([\text{sunny at Wimbledon}], \alpha)$, will be included in the set of worlds where John plays tennis, [John plays tennis]. However, after this information about John's injury becomes known, the subset of worlds closest to the actual world in which it is sunny at Wimbledon *and* John has a broken leg, $s([\text{sunny at Wimbledon \& John has a broken leg}], \alpha)$, will not be included in the set of worlds where John plays tennis, [John plays tennis]. Of course, this depends on our further use of the information, that if your leg is broken, then you can't play tennis. The fact that strengthening is not valid in this semantic theory reflects the empirical data we have reviewed on suppression effects.

The possible worlds semantics for the conditional has never had a great influence on the psychology of reasoning (although, Oaksford [56] discussed this theory at length, and more recently Evans and Over [15] have discussed its consequences). Given that in the areas of logic and formal semantics, the emergence of these theories marked the first real attempts to account for the meaning of natural language conditionals (see, for example, [53]), this is perhaps surprising. One possible explanation is provided by the relatively swift dismissal of the psychological relevance of possible worlds semantics in Johnson-Laird's ([57], pp. 58-59) groundbreaking book, *Mental Models*:

"Possible worlds'...are highly abstract, and since any proposition is either true or false in a given possible world, each possible world goes far beyond what any individual can comprehend...The real problem, however, is that universes of possible worlds are metaphysical luxuries that have nothing to do with the way in which people ordinarily understand conditionals."

So given that each possible world is a specification of the truth or falsity of every possible matter of fact, it might seem that such astonishingly rich representations could not possibly fit "inside the head."

Stalnaker's [55] conceptualist interpretation may not be completely devoid of psychological relevance, however. Such an interpretation proposes that the selection function is an abstraction from the collection of open conditional sentences that describe an individual's world knowledge and which constrain the way in which they alter their beliefs in response to new information. Thus, these methodological policies or habits of inference constrain the possibilities we can consider. For example, consider the methodological policies that support 3.2a and 3.2b, i.e., 3.7a and 3.7b respectively:

If it is sunny, John plays tennis (3.7a)

If the baby sitter is on time, John plays tennis (3.7b)

Suppose John's friend A does not know that John's playing tennis depends on his baby sitting arrangements (3.7b is not one of A's methodological policies). She is likely to evaluate the counterfactual *if it had been sunny yesterday, John would have played tennis* as true. Friend B, on the other hand, knows of 3.7a and of John's reliance on the baby sitter (3.7b) and moreover saw the baby sitter at the beach yesterday when John usually plays tennis. Consequently, B

² Strengthening is the logically valid inference rule that from if p then q one can infer if p and r then q. Notice that suppression of MP by additional information r seems inconsistent with this logical rule. Consequently, it would appear that standard logic is incompatible with suppression effects. One way out is to assume the general form of the conditional is if (p or r) and s, then q (where r denotes alternative causes, and s additional conditions), see for example, Byrne, [48]. However, if all a reasoner is given is the categorical premise p, this general form permits no inferences at all.

won't evaluate the counterfactual as true. Moreover, Friend C, who knows about the baby sitter, also knows that he is unreliable, often failing to turn up without warning. Consequently, C will only endorse the counterfactual as *probable* depending on their estimate of how likely the baby sitter is to turn up.

So once we move away from epistemically ideal states, different epistemic agents, with different sets of methodological policies and other factual knowledge will evaluate the same counterfactual differently. That is, *prior knowledge matters* in the evaluation of conditionals. Moreover, people's prior knowledge may not incline them to endorse a counterfactual as true but only as probable. Such considerations also apply equally to indicative conditionals like 3.2a and 3.7a. We may endorse 3.2a because we possess the methodological policy in 3.7a. So when we add the antecedent, *it is sunny tomorrow*, to our set of beliefs, 3.7a will generate a new belief state in which the consequent of 3.2a is true. Alternatively, of course, we may believe it to be true because John made a conditional promise to this effect and we know John is trustworthy. Or if we believe that John is untrustworthy or just generally unreliable, we may only assign a greater than chance probability to John playing tennis in our revised set of beliefs and hence to the conditional. Of course, we would evaluate an assertion of 3.7a as true if we also possess it as one of our methodological policies. If we did not then it can still be assessed by adding its antecedent to our stock of beliefs and making adjustments based on our other methodological policies and factual knowledge: perhaps we know John is a keen tennis player and that his tennis club has no indoor courts and so this sounds like a perfectly reasonable methodological policy for John to possess. Alternatively, we may have accumulated enumerative evidence for this proposition, i.e., recollections of many instances of John playing tennis on sunny days.

Consequently, the intuitive motivation for possible worlds semantics (about truth and falsity) makes direct appeal to people's knowledge states. From a psychological point of view, this is probably where we should be looking to find an account of how people, who possess different knowledge, evaluate and draw inferences from conditionals. This is a point recently emphasised by Rescher ([27], Preface) who proposes a "treatment of conditionals based on epistemological principles rather than upon semantical principles...[which] makes it easier to understand how conditionals actually function in our thought and discourse." Rescher attributes this approach to Ramsey's earlier proposal (see above on the Ramsey Test) and makes a second suggestion, that to understand conditionals requires moving, "into the realm claimed by theorists in artificial intelligence as they try to simulate our actual information-processing practices" ([27], Cover).

Despite the rich connections between indicative conditionals describing our methodological policies to revise our beliefs and counterfactual conditionals, perhaps the dominant psychological theory of counterfactuals, mental models, does not address this connection. This theory involves the mental representation of the truth table cases associated with a binary logical connective between two propositions p and q

[58, 59]. So for example, in the real world, α , p and q may occur, which means that the other three possibilities— $\neg p$ and $\neg q$, $\neg p$ and q , $\neg p$ and $\neg q$ —do not occur. In a mental model, p and q is labelled the "factual" possibility and the other three possibilities are labelled "counterfactual." In representing the assertion that "if p had not occurred..." people represent both the factual possibility, p , and the supposed possibility, $\neg p$. This way of looking at counterfactuals suggests that counterfactual thought "might be far more logical than previously supposed" ([58], p. 430). And indeed it seems to account for the fact that the logics—the inferences they licence—of indicatives and counterfactuals are very similar [60]. However, it fails to address the rich connections between indicatives and counterfactuals, insofar as it is the causal dependencies in the world, represented linguistically as indicative conditionals and mentally as methodological policies, which underpin claims about the truth or falsity of counterfactuals. As we will see, in probabilistic Bayes net approaches, counterfactuality is captured by the ability to intervene on dynamic representations of methodological policies, to directly test what would happen if p had not happened or p had happened [61-63].

2.4. Ramsey, Probabilities and Causal Conditional Reasoning

Moving away from the idealisations of possible worlds semantics leads to an epistemic approach based on people's stored world knowledge. As Rescher [27] acknowledges, such an approach also has its origins in the Ramsey test which is regarded as the definition of conditional probability in the subjective Bayesian approach to the interpretation of probability statements [60]. As Bennett ([60], p. 53) says:

"The best definition we have [of conditional probability] is the one provided by the Ramsey test: your conditional probability for q given p is the probability for q that results from adding $P(p) = 1$ to your belief system and conservatively adjusting to make room for it."

So the assessment of the conditional probability involves one's prior beliefs, B , held in long term memory. The probabilistic approach to conditional reasoning developed by Adams [12,14] begins from the assumption that the probability of the conditional $P(\text{if } p \text{ then } q)$ can be identified with the conditional probability, subjectively interpreted, i.e., as $P(q|p, B)$. There are several variants on this idea (see [64, 65]) but following some proposals in Adams [14], we have proposed a computational level account of conditional inference as dynamic belief update by Bayesian conditionalisation [19, 47].

On this view, if a high probability is assigned to *if x is a bird, x flies*, then on acquiring the new information that *Tweety is a bird*, one's degree of belief in *Tweety flies* should be revised to one's degree of belief in *Tweety flies given Tweety is a bird*, i.e., one's degree of belief in the conditional. So using P_0 to indicate *prior* degree of belief and P_1 to indicate *posterior* degree of belief, then:

$$P_1(q) = P_0(q|p), \text{ when } P_1(p) = 1. \quad (3.8)$$

Thus according to this account, the probability with which someone should endorse the MP inference is the con-

ditional probability (see [19]). However, as Oaksford and Chater [47] point out there is a problem with extending this account to MT, DA and AC [66]. The appropriate conditional probabilities for the categorical premise of these inferences to conditionalize on are $P(\neg p|\neg q)$, $P(\neg q|\neg p)$, and $P(p|q)$ respectively. However, the premises of MT and the fallacies do not entail values for these conditional probabilities [66-68]. Oaksford *et al.* [19] suggested that people had prior knowledge of the marginals, $P(p)$ and $P(q)$, which together with $P(q|p)$ do entail appropriate values (see, Wagner [2004] for a similar approach) ($P_0(q|p) = a$, $P_0(p) = b$, $P_0(q) = c$):

$$\text{MP } P_1(q) = P_0(q|p) = a \quad (3.9)$$

$$\text{DA } P_1(\neg q) = P_0(\neg q|\neg p) = (1 - c - (1 - a)b)/(1 - b) \quad (3.10)$$

$$\text{AC } P_1(p) = P_0(p|q) = ab/c, \quad (3.11)$$

$$\text{MT } P_1(\neg p) = P_0(\neg p|\neg q) = (1 - c - (1 - a)b)/(1 - c) \quad (3.12)$$

Equations (3.9) to (3.12) show the posterior probabilities of the conclusion of each inference assuming the posterior probability of the categorical premise is 1. By using Jeffrey conditionalization [69], these cases can be generalized to when the probability of the categorical premise is less than 1 [47]. However, the well documented problems with Jeffrey conditionalization [70-72] mean that we do not necessarily endorse this approach (for alternatives see, [73, 74]).

Recent evidence shows that people do regard the probability of a conditional to be the conditional probability as the probabilistic account presupposes [75-77]. For example, Evans *et al.* [75] assessed people's probabilistic interpretations of conditional rules and their contrapositives (*if $\neg q$ then $\neg p$*). They tested three possibilities. First, material implication predicts that the probability of a conditional should be $1 - P(p, \neg q)$, i.e., 1 minus the probability of finding a falsifying case. Second, the conditional probability account predicts that the probability of a conditional should be $P(q|p)$. Finally, in the "defective truth table" account [78], where false antecedent cases are irrelevant, the probability of a conditional should be the joint probability, $P(p, q)$. According to material implication, conditionals and their contrapositives should be endorsed equally because they are logically equivalent. Consequently, there should be a strong correlation between ratings of how likely the conditional and its contrapositive are to be true. However, according to the conditional probability account, $P(q|p)$ and $P(\neg p|\neg q)$ can differ considerably and would not be expected to reveal a perfect correlation.

Evans *et al.* [75] varied $P(q|p)$, $P(\neg q|p)$ and $P(\neg p)$ by describing the distribution of cards in packs of varying sizes. For example, given a conditional *if the card is yellow then it has a circle printed on it*, participants could be told that there are four yellow circles, one yellow diamond, sixteen red circles and sixteen red diamonds (Oaksford *et al.* [19] used similar manipulations). So $P(q|p) = .8$, $P(\neg q|p) = .2$ and $P(\neg p) = 32/3$. On material implication, increases in $P(\neg p)$ should increase ratings of $P(\text{if the card is yellow then it has a circle printed on it})$; according to conditional probability, they should be independent of $P(\neg p)$; and according to conjunction interpretation, they should decrease with increases in $P(\neg p)$. The evidence supported conditional probability

with some evidence for a joint probability interpretation. Over *et al.* [77] replicated these findings for causal conditionals pre-tested for $P(p)$ and $P(q)$ as in Oaksford, Chater and Grainger [79] and Oaksford *et al.* ([19], Experiment 3). They also found that for these conditionals the conjunctive interpretation was not adopted by a significant proportion of participants unlike in Evans *et al.* [75]. Consequently, the conjunctive interpretation is probably an artefact of unrealistic stimuli. Similar results have been found by Oberauer and Wilhelm [76].

Over *et al.* [77] also found that the delta-p rule, $P(q|p) - P(q|\neg p)$, was also in evidence for causal conditionals and, in their Experiment 3, for counterfactual conditionals (but see [80]). However, in correlational analyses the effect of $P(q|\neg p)$ on ratings of $P(\text{if } p \text{ then } q)$ were much smaller than the effect of $P(q|p)$. The delta-p rule has been used a measure of causal strength and remains, appropriately scaled, the basis of causal power theory [81]. Delta-p and causal power have been used predominantly to explain judgements of causal strength from 2×2 contingency tables, where it is often observed that the cells required to compute $P(q|p)$, i.e., those where p occur, are weighted more highly than those required to compute $P(q|\neg p)$, i.e., those where p does not occur. Normatively, it can be shown that this is to be expected from a Bayesian perspective when p and q are rare [82], e.g., in the category of birds most are not swans (p) and most are not white (q). The rarity assumption also figured in the resolution of the problem of why people appear to confirm in the selection task when selecting data to test a hypothesis (see Section 2).

The work of Evans *et al.* [75], Oberauer and Wilhelm [76], and Over *et al.* [77], concentrated on the probabilistic factors affecting the degree of belief in the conditional, $P(\text{if } p \text{ then } q)$. The main factor was the conditional probability although there was weaker evidence for the delta-p rule. One might argue that the failure of delta-p to act as a strong predictor indicates that the factors that determine people's degrees of belief in cause-effect relations differs from those that determine people's degrees of belief in conditionals. However, in their Experiment 2, Over *et al.* [77] asked participants to rate the causal strength of the relationship described in the causal conditionals they used. These ratings were almost perfectly correlated with their judgements of $P(\text{if } p \text{ then } q)$. This result may suggest that delta-p is not the measure used to judge causal strength and as Hattori and Oaksford [83] observed there are many other potential contenders that have yet to be investigated with respect to degree of belief in the conditional. But we doubt that this indicates that the conditional probability is the best measure of causal strength.

When we move away from the factors affecting the degree of belief in the conditional and turn to the factors that affect conditional inference, a more nuanced picture emerges. Even this weak relationship between delta-p and $P(\text{if } p \text{ then } q)$ suggests that people have access to more information than required to compute $P(q|p)$, i.e., they have access to the information to compute $P(q|\neg p)$. Our approach to conditional inference (3.9 -3.12) suggests that this has to be the case. Without access to the marginals, $P(p)$ and $P(q)$, there is insufficient information to compute the conditional probabilities required by the other inferences, DA, AC, and

MT (see, 3.10 to 3.12) that have been investigated in conditional reasoning. But once the conditional probability and the marginals are known a contingency table is given from which measures like delta- p can be computed. In formal semantics, knowing the meaning of a statement is often equated with the knowing the inferences it licenses [84]. Thus while $P(\text{if } p \text{ then } q)$ may well be equated with $P(q|p)$, understanding the full meaning of the conditional will involve information about $P(p)$ and $P(q)$. Moreover, we have recently argued that additional measures, like delta- p , are required in defining the “inductive strength” of an inference [85].

Measures like delta- p are used in causal learning to establish that there is a causal dependency between p and q and to assess its strength. This involves seeing how far the evidence supports the claim that the effect, p , is not independent of the cause, q , i.e., $P(q|p) \neq P(q)$. Oaksford and Hahn [85] have suggested that related measures may be important in conditional inference. In 3.9 -3.12, Bayesian conditionalisation on the conditional premise given the truth of the categorical premise provides a measure of argument strength, i.e., the posterior probability of the conclusion. However, as a measure of the strength of an argument it fails to account for the intuition that an argument might be regarded as weak if it does not lead to a sufficient *change* in one’s degree of belief. So for example, one might have the following set up for an MP inference: $P_0(q) = .7$, $P_0(q|p) = .7$, on learning that $P_1(p) = 1$, they therefore infer that $P_1(q) = P_0(q|p) = .7$. So my degree of belief in q is moderate but it has not actually changed. Hahn and Oaksford [86] argued that assessing the strength of arguments involves both strength—the probability of the conclusion—and *force*, i.e., the change in probability from prior to posterior.

Rips ([49], p. 128, footnote 1) considered a particular measure of argument force:

$$\text{Argument force} = \frac{P(\text{conclusion} | \text{premises}) - P(\text{conclusion})}{1 - P(\text{conclusion})} \quad (3.13)$$

This measure, which is called *conditional agreement* [87], has the advantage that it looks at the *change* in the probability of the conclusion brought about by the premises and scales this by how far someone needs to go to be convinced of the conclusion. As Rips [49] observes, it is important to have a measure like this because otherwise arguments where the premises are irrelevant to the conclusion would not be assigned 0 argument strength/force but would be as strong as the prior probability of the conclusion.

However, as Oaksford and Hahn observed, (3.13) has some counterintuitive properties. Consider an argument that moves someone from having a very low degree of belief in a conclusion, $P(\text{conclusion}) = .01$, to complete uncertainty, $P(\text{conclusion}) = .5$. Intuitively, this argument is stronger than one that moves someone from a high degree of belief in a conclusion, $P(\text{conclusion}) = .9$, to an even higher degree of belief in a conclusion, $P(\text{conclusion}) = .99$. However, according to (18), the argument strength of the former is .495, while that of the latter is .9. That is, counter-intuitively, the second argument is by far the stronger of the two.

There are many different measures that might be useful as an index of argument force, probably as many as there are

indices of deviations from independence, i.e., $P(\text{conclusion} | \text{premises}) = P(\text{conclusion})$, of which Hattori and Oaksford [83] recently counted at least forty. Hahn *et al.* [73] suggested that the *likelihood ratio* may be a better measure of argument force. The likelihood ratio maps the prior odds onto the posterior odds and so can also be calculated from $P(\text{conclusion} | \text{premises})$ and $P(\text{conclusion})$:

$$\text{Likelihood Ratio} = \frac{P(\text{conclusion} | \text{premises})(1 - P(\text{conclusion}))}{P(\text{conclusion})(1 - P(\text{conclusion} | \text{premises}))} \quad (3.14)$$

Using (3.14) with the previous example, the argument force of the first argument, taking someone from .01 to .5, is 93, whereas the argument force of the second argument, taking someone from .9 to .99, is 11. So the likelihood ratio matches intuition in judging the first argument to be stronger than the second. However, as the degree of conviction gets closer to 1 in the second argument, then this measure may reverse this pattern. Nonetheless, as long as non-extreme values are assumed, this measure seems to capture intuition.

An intriguing possibility is that if most conditional inferences were based on prior learned knowledge of causal and other dependencies in the world then they may come with an inbuilt warrant of argument force. This is because learning these dependencies precisely involves discovering that p and q are not independent, e.g., 3.14 deviates from 1. Such learned dependencies make up our stock of methodological policies to change our beliefs and only methodological policies that can do this are useful to us. Moreover, most human inference is based on such learned knowledge. That is, explicit verbal inference like that investigated in the reasoning lab is very rare. As Dennett ([88], p. 289) has observed:

“But it is obviously true that most people never engage in *explicit* non-enthymematic formal reasoning”

In “enthymematic” reasoning one or more of the premise is filled in by one of our learned methodological policies to change our beliefs, e.g., “Socrates is a man, therefore, Socrates is mortal” misses out the major premise, “All men are mortal” (in conditional form: if x is a man, then x is mortal). In real life, the true inferential nature of much of human thought is hidden by the fact that all the premises are rarely explicitly stated and yet we are filling in the details *via* inference from our learned methodological policies all the time.

There are some additional complexities with the idea that learned methodological policies come with their own warrant of argument force for the conditionals used to describe them. Not everyone will have the same priors and indeed an individual’s priors must vary from context to context such that in some contexts an argument is forceful but in others it is not. For example, in the UK, someone might have the methodological policy that if a bird is white, it is a swan but if they are in a swannery where they expect most birds to be swans, an argument by MP is not going to change their degree of belief about a bird being a swan by very much, if at all. Moreover, there is little point in trying to persuade someone that global warming is real, by telling them if the ice caps are melting, global warming is occurring and that

the ice caps are melting, if they already believe that global warming is real. We do not have the space to consider these issues at length here but note that we believe the general form of a conditional should always be regarded as implicitly bound to a context, i.e., *if p then q|C*, *q* follows from *p* but only given *C*.

Of course, as we have already discussed, the relevant context, *C*, is also given by the larger causal structures in which the dependencies described by a conditional are embedded, including alternative causes and cotenable conditions. So cars starting follows from their ignition keys being turned only given there is petrol in the tank, the battery is not flat and so on. Moreover, cars can also start when they are hot-wired. How such complex causal structures may be used in inference and even perhaps mentally represented has been discussed only recently in the psychology of reasoning by appeal to the formalism of Causal Bayesian networks ([61, 89, 90]. Bayesian networks are directed acyclic graphs. The nodes represent random variables in the Bayesian sense and the edges represent conditional dependencies. Nodes that are not connected represent variables which are conditionally independent of each other. We know that even pre-school children are capable of learning the more complex causal structures of the type assumed in the causal Bayes net framework (e.g., [91-93]). Moreover, Chater and Oaksford [90] suggested that analogous “mental mechanisms” may provide the kind of dynamic mental representation of the parts of the world that underpin many forms of inference.

The ability to intervene in such a representation underpins its ability to directly implement the Ramsey test and to account for counterfactual reasoning. Directly setting a variable *p* to 1 or 0—like clamping on or off a node in a neural network—represents supposing it to be true or false. This action can also be carried out by the world, i.e., when *p* actually is, rather than just supposed to be, 1. So someone may see the key being turned which sets the variable representing this event, *p*, to 1. They therefore predict the car will start because the conditional dependency between the variables representing these events in the network causes the “car starts” variable, *q*, to move towards 1. Of course, if after the car has started our reasoner also speculated that if the key had not been turned the car would not have started, all she need do is set, *p* to 0, and see what happens to *q*. In the absence of information about the presence of alternative causes, e.g., hot-wiring, *q* should tend to 0. Very simplistically, this is how such networks can also address the relationship between indicative and counterfactual conditionals [6-9].

If dynamic mental mechanisms analogous to causal Bayes nets constitute the mental representations that underpin conditional reasoning, it would be useful in distinguishing this hypothesis from alternatives if there were effects on conditional inference that are uniquely predicted by this hypothesis. This possibility has recently been explored by Ali, Schottmann, Shaw, Chater, & Oaksford [89]. Rather like Byrne [48], participants were presented with more than one conditional statement. Ali *et al.* [89] also used causal—*if c then e*—and diagnostic—*if e then c*—conditionals. We describe two conditions of their experiments which are logically identical but which make contrasting predictions under a causal interpretation:

Causal	Diagnostic
(1) If it rains (<i>c</i> 1), the streets are wet (<i>e</i>)	If it is warm outside (<i>e</i> 1), it is sunny (<i>c</i>)
(2) If the sprinklers are on (<i>c</i> 2), the streets are wet (<i>e</i>)	If there are shadows (<i>e</i> 2), it is sunny (<i>c</i>)
(3) The streets are wet	It is sunny
(4) How likely is “it rains”?	How likely is “it is warm outside”?
(5) $P(c1 e, B)$	$P(e1 c, B)$
(6) The sprinklers are on	There are shadows
(7) How likely is “it rains”?	How likely is “it is warm outside”?
(8) $P(c1 e, c2, B)$	$P(e1 c, e2, B)$
(9) $P(c1 e, c2, B) < P(c1 e, B)$	$P(e1 c, e2, B) = P(e1 c, B)$
(10) Discounting	Independence

The premises (1) to (3) were first presented in each case, representing the background knowledge, *B*, from which the probability ratings (4)—corresponding to the probabilities in (5)—need to be derived. Participants were then provided with the additional premise in (6) and then asked for the same probability rating again (7) corresponding to the probability in (8). In the causal case, the possibility of rain should be discounted as a possible cause of the streets being wet because it is now known that the sprinklers are on (9). In the diagnostic case, if it is sunny then both common effects should occur, and so no discounting should happen (9). These differing predictions arise from the different causal structures described by premises (1) and (2) in the causal and diagnostic cases. For the causal case this is a “collider” structure where the two causes converge on a common effect. This contrasts with the diagnostic case where the underlying causal structure diverges from a common cause of two effects. In Ali *et al.*'s [89] experiment, using 7 and 8 year old children, exactly the predicted pattern of effects was observed.³

Besides our causal Bayes approach, there is also a related approach by Sloman and Lagnado [61]. These authors, however, argued that conditional reasoning differs in quality from causal reasoning, based on evidence showing that causal descriptions and corresponding conditional descriptions produced various systematic differences. In particular, conditional statements tended to produce weaker effects than causal statements. However, we believe there are several issues with Sloman and Lagnado's [61] experiments, to do with an important element of Causal Bayes models, the distinction between structure and strength [94].

We believe that conditionals, “*if c then e*,” are structure building operators in the same way as “*c causes e*,” i.e., they suggest that some dependency exists between *c* and *e*. Different information about enablers, common causes or

³ However, for adults Ali, Chater, & Oaksford (in prep) found discounting in both cases. They argued that in the diagnostic case, adults may make the pragmatic inference from the fact that the premises only state that one of the common effects occurred (6) that this was an abnormal situation in which these effects were not correlated, e.g., at the poles when it is sunny there are shadows but it is not warm and indoors warmth and shadows are not correlated.

effects, will alter the local mental structure people build in working memory. However, to perform further inferences, the various parameters of these structures, or mental mechanisms [90], must be set from linguistic or environmental cues or from prior world knowledge. The mental structures built will be the same for causal and other dependencies, reflecting the idea in situation semantics [95] that causal dependencies are regarded as the core meaning of the conditional. However, the parameters of these models, their strength, and whether additional machinery, perhaps to do with utilities, are needed will vary between different types of conditional sentences.

We argue that the differences between causal and conditional reasoning that Sloman and Lagnado [61] observed are due to variations in strength but not structure. We look at two cases. First, in their Experiment 1, a structure $A \rightarrow B \rightarrow C$ is introduced concerning three billiard balls. This structure is described using causal terminology, e.g. Ball 1's (A) movement causes Ball 2 (B) to move, causal conditionals, e.g., If Ball 1 moves, then Ball 2 moves, or logical conditionals, which use the same conditionals but with the preamble, "Someone is showing off her logical abilities. She is moving balls without breaking the following rules." Participants are then told that there is an intervention on B which prevents it from moving, this should lead to the structure $A \rightarrow C$. So if asked, "Imagine that Ball 2 could not move, would Ball 1 still move?" participants should say "Yes," but if asked, "Imagine that Ball 2 could not move, would Ball 3 still move?" participants should say "No." For all three descriptions participants said that Ball 3 would not move but for the "logical" conditionals about 45% said that Ball 1 would still move whereas 90% endorsed this statement for causal and causal conditional descriptions.

The only difference in the logical condition relates to the possible causes of Ball 1 moving. In the "causal" cases, a causally open system is described where the normal causes of Ball 1 moving are in operation which should be unaffected by Ball 2 being prevented from moving. However, for the "logical" case a closed system is described in which the only cause of balls moving and so of Ball 1 moving is an intentional action of the person showing off her logical abilities. With no knowledge of the rule governing whether she moves Ball 1, participants assume she moves it at random, i.e., the probability is approximately .5. So this is not an instance of people adopting a fundamentally different interpretation of conditional and causal statements. The "logical" preamble simply changes the interpretation of the initiating causes of Ball 1 moving, from an open system to a closed system concerning the intentional actions of an agent.

The second case we look at involves their Experiment 3 ([61], p.20). Again we argue that the observed differences between causal and conditional reasoning can be explained by the difference between strength and structure:

"Germany's undue aggression has caused France to declare war. Germany's undue aggression has caused England to declare war. France's declaration causes Germany to declare war. England's declaration causes Germany to declare war. And so, Germany declares war.

- (1) If England had not declared war, would Germany have declared war?
- (2) If England had not declared war, would Germany have been aggressive?"

Thus the causal statements categorically assert that Germany's undue aggression was the cause of France going to war. They are in the past tense describing events that have already occurred. This is in contrast with the conditional statements, using materials like: "if Germany is unduly aggressive, then France will declare war" The consequent of the conditional statement is in the future tense describing an event that may happen in the future given the antecedent event occurs. The future is uncertain in a way that the past is not. Consequently these descriptions differ in strength but relate to the same causal structure. To compare like with like would involve using causal statements like "Germany's undue aggression will cause France to declare war." The difference in strength accounts fully for the lower endorsement of questions 1 and 2 for the conditional case. In sum, Sloman and Lagnado's [61] data are consistent with a Causal Bayes net approach to both conditional and causal reasoning.

Consistent with the conclusions of recent approaches in the semantics of conditionals, most recent research on conditional inference has been on the various effects of prior knowledge on causal conditional reasoning. The original work on causal reasoning, in which cotenable conditions and alternative causes were manipulated, directly investigated these effects [49, 50]. Failure of a cotenable condition for *if p then q* just is a situation in which *p* may be true but *q* is false, i.e., a counterexample. There is now a body of work showing that the efficiency of retrieval of cotenable or "disabling" conditions directly affects inference [96-98]. There is also work showing that asking people to generate alternative causes leads to suppression of MP and MT thought only to be affected by cotenable or disabling conditions [99]. Markovits and Potvin [99] interpret this result to indicate that accessing alternative causes activates the whole semantic frame in which a causal conditional is embedded including alternative causes and cotenable conditions. A conclusion wholly consistent with the causal Bayes net approaches we have just discussed. There is also evidence showing that the frequency of counterexamples, i.e., cotenable conditions or alternative causes, matters more in suppressing inferences than the range of different types of counterexample [100]. Moreover, there is evidence that people can inhibit the process of retrieving counterexamples [101-103] when knowledge and logic conflict or when told to give a logical response, in order to arrive at the responses that look more "logical."

This empirical work continues to be described in the psychology of reasoning in terms of standard logic. The perpetuation of the influence of standard logic derives from three kinds of result. First, in a major modelling paper, Oberauer [104] showed that a probabilistic approach based on equations 3.9 – 3.12, produced worse fits to two large scale web based data sets, than *dual process* models that allowed a role for logic based theories like mental models or mental logic while also allowing that some other knowledge based processes may also be invoked. Second, more specifi-

cally, Weidenfeld, Oberauer, and Hörnig [105] showed, in a structural equation model of another large scale web based data set, that there was a direct path between the occurrence of cotenable conditions and whether people draw MP and MT inferences. That is, this path was not mediated by the conditional probability, $P(q/p)$, as one might expect from a purely probabilistic approach as articulated in equations 3.9 – 3.12. Third, De Neys and Franssens [106] present evidence using a lexical decision task that people can actively inhibit the representations of prior beliefs that may lead to drawing non-logical conclusions. De Neys [102] directly applied this paradigm to conditional inference and showed that cotenable conditions and their semantic associates are inhibited for MP and MT under instructions to make a binary valid vs invalid judgement. However, alternative causes and their semantic associates were facilitated for DA and AC under the same instructions.

These results, while counting against an account that relies only on calculating the strength of a dependency, fails as counterevidence to a Causal Bayes net or a neural network approach [64]. In such approaches, the existence of structural relations in a mental representation indicating that a dependency exists is a separate judgement to the determining the strength of that relationship [94]. That is, in a causal Bayes net or a neural network, the decision to include a link between two variables/nodes is distinct from the process of the learning the strength of the relationship between them. Griffiths and Tenenbaum's [94] causal support theory articulated how people learn different causal structures from data, i.e., it addresses the question of whether a causal dependency exists or not. As we argued above, the assertion of a conditional, directly proposes the existence of a causal or related dependency between the antecedent and the consequent. Prior knowledge of the specific content, i.e., about the specific dependency described in the conditional, is then required to assign values to the appropriate parameters and to embed it in further structural relationships determined by prior knowledge of alternative causes and cotenable conditions.

Oaksford and Chater [64, 65] provide a constraint satisfaction neural net implementation of conditional inference which also directly implements the inhibitory processes emphasised by De Neys and colleagues [102, 106]. In particular, any account of the mental representations that allow interventions like causal Bayes nets and neural networks, automatically implement mechanisms of inhibition: interventions can be inhibitory or excitatory. In Oaksford and Chater's [64, 65] constraint satisfaction model, cotenable conditions are implemented as inhibitory links between a set of context units and the effect. If the context nodes representing cotenable conditions are inhibited by an instruction to only indicate what logically follows [106], i.e., to ignore other factors, then they cannot inhibit the cause/antecedent node turning on the effect/consequent node. Consequently, more "logic-like" results will emerge. This inhibitory action will mean that semantically related nodes will become less active yielding the effects observed by De Neys [102] in his lexical decision tasks. In sum, the fascinating results emerging in recent psychology of conditional reasoning are understandable when we interpret conditionals as the verbal ex-

pression of dependencies or methodological policies mentally represented as edges in dynamic mental mechanisms analogous to causal Bayes nets or neural networks in which the ability to intervene *via* inhibitory or excitatory processes are basic operating principles.

CONCLUSION

In this paper, our goal has been to trace the connections between causation and the conditional across the two main paradigms used in studying conditional reasoning, the selection task and the conditional inference paradigm. We began with the hypothesis testing in the selection task. We noted that the problems identified by Quine and Goodman in the philosophical literature about the material conditional interpretation of causal laws did recur in the psychological data. Moreover, the fact that causal relations are always embedded in a larger causal structure involving alternative causes and cotenable conditions features in some early work using the conditional inference paradigm. We then looked at the possible worlds semantics for the counterfactual conditional and traced out the relationship between these conditionals, which provide Hume's second definition of causation, and the indicative conditional. We also concluded that this relationship was effectively lost in the main extant theory of counterfactual reasoning in the psychology of reasoning, i.e., mental models. We then moved on to look at recent probabilistic approaches to conditional inference based on conditional probability. These accounts provide measures of the strength of a dependency between the antecedent and the consequent of a conditional in terms of the relevant conditional probability. However, determining the relevant probabilities to draw inferences that go beyond MP requires more information from world knowledge, regarding the base rate of the antecedent and consequent. Moreover, when these are available there is sufficient information to calculate measures of argument force, such as the likelihood ratio, which provide indices of deviation from independence and which have been used as measures of causal strength in the literature on causal learning. There is also evidence that such measures do influence judgements of the probability of a conditional. However, in the recent literature there are a variety of findings that suggest that in conditional inference people are not only influenced by the strength of the dependency but also by the existence of the structural relationship, the broader causal framework in which a dependency is embedded, and the inhibitory and excitatory processes require to implement Causal Bayes nets or neural networks. These findings in conditional inference and their plausible explanation using the tools of current theories in causal learning suggest a convergence of work which we argue should benefit both areas of research.

REFERENCES

- [1] Comrie B. Conditionals: A typology. In: Traugott EC, Ter Meulen A, Reilly JS, Ferguson CA, Eds. *On conditionals*. Cambridge: Cambridge University Press 1986; pp. 77-99.
- [2] Stalnaker RC. A theory of conditionals. In: Rescher N, Ed. *Studies in logical theory*. Oxford: Blackwell 1968; pp. 98-112.
- [3] Lewis D. *Counterfactuals*. Cambridge, MA: Harvard University Press 1973.
- [4] Pearl J. *Causality: models, reasoning, and inference*. Cambridge, UK: Cambridge University Press 2000.

- [5] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. New York: Springer 1993.
- [6] Hitchcock C. The intransitivity of causation revealed in equations and graphs. *J Philos* 2001; 98: 273-99.
- [7] Hitchcock C. Prevention, preemption, and the principle of sufficient reason. *Philos Rev* 2007; 116: 495-532.
- [8] Woodward J. Making things happen: a theory of causal explanation. Oxford: Oxford University Press 2003.
- [9] Woodward J, Hitchcock C. Explanatory generalizations. Part I: A counterfactual account. *Nous* 2003; 37: 1-24.
- [10] Salmon WC. Causality and explanation. Oxford: Oxford University Press 1971.
- [11] Suppes P. Probabilistic metaphysics. Oxford: Oxford University Press 1984.
- [12] Adams EW. Probability and the logic of conditionals. In: Suppes F, Hintikka J, Eds. Aspects of inductive logic. Amsterdam: North Holland 1968; pp. 265-316.
- [13] Adams E. The logic of conditionals: an application of probability to deductive logic. Dordrecht: Reidel 1975.
- [14] Adams EW. A primer of probability logic. Stanford: CLSI Publications 1998.
- [15] Evans J.St.BT, Over DE. If. Oxford, England: Oxford University Press 2004.
- [16] Oaksford M, Chater N. Rationality in an uncertain world: Essays on the cognitive science of human reasoning. Hove, Sussex: Psychology Press 1998.
- [17] Oaksford M, Chater N. Probability logic and the Modus Ponens-Modus Tollens asymmetry in conditional inference. In: Chater N, Oaksford M, Eds. The probabilistic mind: Prospects for Bayesian cognitive science. Oxford: Oxford University Press 2008; pp. 97-120.
- [18] Oaksford M, Chater N. The uncertain reasoner: bayes, logic and rationality. *Behav Brain Sci* 2009; 32: 105-20.
- [19] Oaksford M, Chater N, Larkin J. Probabilities and polarity biases in conditional inference. *J Exp Psychol Learn Mem Cogn* 2000; 26: 883-9.
- [20] Wason PC. Reasoning about a rule. *Q J Exp Psychol* 1968; 20: 273-81.
- [21] Popper K. The logic of scientific discovery. New York: Basic Books (originally 1935) 1959.
- [22] Wason P, Johnson-Laird P. Psychology of reasoning: structure and content. Oxford England: Harvard Unary Press 1972.
- [23] Rips LJ. The Psychology of proof. Cambridge, MA: MIT Press 1994.
- [24] Duhem P. The aim and structure of physical theory. Princeton, NJ: Princeton University Press (English trans. 1954) 1914.
- [25] Quine WVO. From a logical point of view. Cambridge, MA: Harvard University Press 1953.
- [26] Goodman. Fact, fiction, and forecast. London: The Athlone Press 1954.
- [27] Rescher N. Conditionals. Cambridge, MA: MIT Press 2007.
- [28] Kvat I. Theory of Counterfactuals. Indianapolis, IN: Hackett Publishing Company Inc. 1985.
- [29] Holland JH, Holyoak KJ, Nisbett RE, Thagard PR. Induction. Cambridge, MA: MIT Press 1986.
- [30] Hempel CG. Studies in the logic of confirmation I. *Mind* 1945a; 54: 1-26.
- [31] Hempel CG. Studies in the logic of confirmation II. *Mind* 1945b; 54: 97-121.
- [32] Howson C, Urbach P. Scientific reasoning: the Bayesian approach 2nd ed. La Salle, Illinois: Open Court Publishing Company 1993.
- [33] Wetherick N. Eliminative and enumerative behaviour in a conceptual task. *Q J Exp Psychol* 1962; 14: 246-9.
- [34] Nickerson R. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998; 2: 175-220.
- [35] Ebbinghaus H, Meyer M. Psychology: an elementary text-book. Boston, MA: D C Heath & Co Publishers 1908.
- [36] Wertheimer M. Productive thinking. Oxford England: Harper 1959.
- [37] Oaksford M, Chater N, Stenning K. Connectionism, classical cognitive science and experimental psychology. *AI Society* 1990; 4: 73-90.
- [38] Wason PC, Shapiro D. Natural and contrived experience in a reasoning problem. *Q J Exp Psychol* 1971; 23: 63-71.
- [39] Johnson-Laird PN, Legrenzi P, Legrenzi M. Reasoning and a sense of reality. *Br J Psychol* 1972; 63: 395-400.
- [40] Griggs R, Cox J. The elusive thematic-materials effect in Wason's selection task. *Br J Psychol* 1982; 73: 407-20.
- [41] Lindley DV. On a measure of information provided by an experiment. *Ann Math Stat* 1956; 27: 986-1005.
- [42] Oaksford M, Chater N. A rational analysis of the selection task as optimal data selection. *Psychol Rev* 1994; 101: 608-31.
- [43] Oaksford M, Chater N. Rational explanation of the selection task. *Psychol Rev* 1996; 103: 381-91.
- [44] Oaksford M, Chater N. Optimal data selection: revision, review and re-evaluation. *Psychon Bull Rev* 2003; 10: 289-318.
- [45] Roberge J. A study of children's abilities to reason with basic principles of deductive reasoning. *Am Educ Res J* 1970; 7: 583-96.
- [46] Roberge J. Some effects of negation on adults' conditional reasoning abilities. *Psychol Rep* 1971; 29: 839-44.
- [47] Oaksford M, Chater N. Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press 2007.
- [48] Byrne R. Suppressing valid inferences with conditionals. *Cognition* 1989; 31: 61-83.
- [49] Cummins DD. Naive theories and causal deduction. *Mem Cogn* 1995; 23: 646-58.
- [49] Rips LJ. Two kinds of reasoning. *Psychol Sci* 2001; 12: 129-34.
- [50] Cummins DD, Lubart T, Alksnis O, Rist R. Conditional reasoning and causation. *Mem Cogn* 1991; 19: 274-82.
- [51] Thompson V, Mann J. Perceived necessity explains the dissociation between logic and meaning: the case of 'Only If'. *J Exp Psychol Learn Mem Cogn* 1995; 21: 1554-67.
- [52] Rumain B, Connell J, Braine M. Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Dev Psychol* 1983; 19: 471-81.
- [53] Nute D. Conditional logic. In: Gabbay D, Gunthner F, Eds. Handbook of Philosophical Logic Amsterdam: Kluwer Academic Publishers 1984; vol. 2: pp. 387-439.
- [54] Ramsey FP. The foundations of mathematics and other logical essays. London: Routledge and Kegan Paul (originally 1931) 1990.
- [55] Stalnaker RC. Inquiry. Cambridge, MA: MIT Press 1984.
- [56] Oaksford M. Cognition and inquiry: the pragmatics of conditional reasoning. Unpublished doctoral dissertation. Centre for Cognitive Science, University of Edinburgh 1989.
- [57] Johnson-Laird PN. Mental models. Cambridge: Cambridge University Press 1983.
- [58] Byrne RMJ. Mental models and counterfactual thoughts about what might have been. *Trends Cogn Sci* 2002; 6: 10: 426-31.
- [59] Byrne RMJ. The rational imagination: how people create alternatives to reality. Cambridge, M.A: MIT Press 2005.
- [60] Bennett J. A philosophical guide to conditionals. Oxford England: Oxford University Press 2003.
- [61] Sloman SA, Lagnado DA. Do we do? *Cogn Sci* 2005; 29: 5-39.
- [62] Dehghani M, Iliev R, Kaufmann S. Effects of fact mutability in the interpretation of counterfactuals. In: Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society, Nashville, TN 2007.
- [63] Dehghani M, Iliev R, Kaufmann S. Causal Explanations in Counterfactual Reasoning. In: Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands 2009.
- [64] Oaksford M, Chater N, Eds. Cognition and conditionals: Probability and logic in human thinking. Oxford: Oxford University Press 2010.
- [65] Oaksford M, Chater N. Conditionals and constraint satisfaction: Reconciling mental models and the probabilistic approach? In: Oaksford M, Chater N, Eds. Cognition and conditionals: Probability and logic in human thinking. Oxford: Oxford University Press 2010.
- [66] Sober E. Intelligent design and probability reasoning. *Int J Philos Religion* 2002; 52: 65-80.
- [67] Sobel JH. Probable modus ponens and modus tollens and updating on uncertain evidence. Unpublished manuscript, Department of Philosophy, University of Toronto, Scarborough 2004. Available from: www.scar.toronto.ca/~sobel/ConfDisconf.pdf.
- [68] Wagner CG. Modus tollens probabilized. *Br J Philos Sci* 2004; 55: 747-53.
- [69] Jeffrey R. The Logic of decision. 2nd ed. Chicago: University of Chicago Press 1983.

- [70] Field H. A note on Jeffrey conditionalization. *Philos Sci* 1978; 45: 361-67.
- [71] Osherson D. Order dependence and Jeffrey conditionalization 2005. Available from http://www.princeton.edu/_osherson.
- [72] Over DE. New paradigm psychology of reasoning (Review of Bayesian rationality: the probabilistic approach to human reasoning). *Think Reason* 2009; 15: 431-8.
- [73] Hahn U, Oaksford M, Bayindir H. How convinced should we be by negative evidence? In: Bara B, Barsalou L, Bucciarelli M, Eds. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Mahwah, N.J.: Lawrence Erlbaum Associates 2005; pp. 887-92.
- [74] Chater N, Oaksford M. Local and global inferential relations: Response to Over. *Think Reason* 2009; 15: 439-46.
- [75] Evans J.St.BT, Handley SH, Over DE. Conditionals and conditional probability. *J Exp Psychol Learn Mem Cogn* 2003; 29: 321-55.
- [76] Oberauer K, Wilhelm O. The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *J Exp Psychol Learn Mem Cogn* 2003; 29: 680-93.
- [77] Over DE, Hadjichristidis C, Evans J.St.BT, Sloman S, Handley S. The probability of causal conditionals. *Cogn Psychol* 2007; 54: 62-97.
- [78] Johnson-Laird PN, Tagart J. How implication is understood. *Am J Psychol* 1969; 2: 367-73.
- [79] Oaksford M, Chater N, Grainger B. Probabilistic effects in data selection. *Think Reason* 1999; 5: 193-243.
- [80] Oberauer K, Weidenfeld A, Fischer K. What makes us believe a conditional? The roles of covariation and causality. *Think Reason* 2007; 13: 340-69.
- [81] Cheng PW. From covariation to causation: a causal power theory. *Psychol Rev* 1997; 104: 367-405.
- [82] McKenzie C, Mikkelsen L. A Bayesian view of covariation assessment. *Cogn Psychol* 2007; 54: 33-61.
- [83] Hattori M, Oaksford M. Adaptive non-interventional heuristics for co-variation detection in causal induction: model comparison and rational analysis. *Cogn Sci* 2007; 31: 765-814.
- [84] Dowty D, Wall RE, Peters S. *Introduction to montague semantics*. Dordrecht, Holland: Reidel 1981.
- [85] Oaksford M, Hahn U. Induction, deduction and argument strength in human reasoning and argumentation. In: Feeney A, Heit E, Eds. *Inductive reasoning*, Cambridge: Cambridge University Press 2007; pp. 269-301.
- [86] Hahn U, Oaksford M. The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychol Rev* 2007; 114: 704-32.
- [87] Bishop YMM, Feinberg SE, Holland PW. *Discrete multivariate analysis*. Cambridge, MA: MIT Press 1975.
- [88] Dennett DC. Reflections on language and mind. In: Carruthers P, Boucher J, Eds. *Language and Thought: Interdisciplinary Themes*. Cambridge: Cambridge University Press 1998; pp. 284-94.
- [89] Ali N, Schlottmann A, Shaw C, Chater N, Oaksford M. Conditionals and causal discounting in children. In: Oaksford M, Chater N, Eds. *Cognition and conditionals: Probability and logic in human thinking*. Oxford: Oxford University Press (in press).
- [90] Chater N, Oaksford M. Mental mechanisms. In: Fiedler K, Juslin P, Eds. *Information sampling and adaptive cognition*. Cambridge: Cambridge University Press 2006; pp. 210-36.
- [91] Gopnik A, Sobel DM, Schulz LE, Glymour C. Causal Learning Mechanisms in Very Young Children: Two-, Three- and Four-Year-Olds Infer Causal Relations From Patterns of Variation and Covariation. *Dev Psychol* 2001; 37: 620-9.
- [92] Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, Danks D. A theory of causal learning in children: Causal maps and Bayes nets. *Psychol Rev* 2004; 111: 1-31.
- [93] Schulz LE, Gopnik A, Glymour C. Preschool children learn about causal structure from conditional interventions. *Dev Psychol* 2007; 10: 322-32.
- [94] Griffiths TL, Tenenbaum JB. Structure and strength in causal induction. *Cogn Psychol* 2005; 51: 334-384.
- [95] Barwise J, Perry J. *Situation semantics*. Cambridge, MA: MIT Press 1983.
- [96] De Neys WD, Schaeken W, d'Ydewalle G. Causal conditional reasoning and strength of association: the disabling condition case. *Eur J Cogn Psychol* 2003a; 15: 161-76.
- [97] De Neys WD, Schaeken W, d'Ydewalle G. Inference suppression and semantic memory retrieval: every counterexample counts. *Mem Cogn* 2003b; 31: 581-95.
- [98] Markovits H, Quinn S. Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Mem Cogn* 2002; 30(5): 696-706.
- [99] Markovits H, Potvin F. Suppression of valid inferences and knowledge structures: the curious effect of producing alternative antecedents on reasoning with causal conditionals. *Mem Cogn* 2001; 29: 736-44.
- [100] Geiger SM, Oberauer K. Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Mem Cogn* 2007; 35: 2060-74.
- [101] Simoneau M, Markovits H. Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Dev Psychol* 2003; 39: 964-75.
- [102] De Neys WD. Counterexample retrieval and inhibition during conditional reasoning: Direct evidence from memory probing. In: Oaksford M, Chater N, Eds. *Cognition and conditionals: Probability and logic in human thinking*. Oxford: Oxford University Press 2010.
- [103] De Neys W, Vartanian O, Goel V. Smarter than we think: When our brains detect that we are biased. *Psychol Sci* 2008; 19: 483-9.
- [104] Oberauer K. Reasoning with conditionals: a test of formal models of four theories. *Cogn Psychol* 2006; 53: 238-83.
- [105] Weidenfeld A, Oberauer K, Hörnig R. Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *Q J Exp Psychol A* 2005; 58: 1479-513.
- [106] De Neys W, Franssens S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* 2009; 113: 45-61.

Received: December 08, 2009

Revised: January 18, 2010

Accepted: January 18, 2010

© Oaksford and Chater.; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.