Paradigms, Possibilities, and Probabilities: Comment on Hinterecker et al. (2016)

Mike Oaksford*

Birkbeck College, University of London

David Over

University of Durham

Nicole Cruz

Birkbeck College, University of London and

École Pratique des Hautes Études, Paris, France

DOI: 10.1037/xlm0000586

*Please address correspondence to Mike Oaksford, Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK. E-mail: mike.oaksford@bbk.ac.uk. We thank two anonymous reviewers for their helpful and constructive comments.

Abstract

Hinterecker et al. (2016) compared the adequacy of the probabilistic new paradigm in reasoning with the recent revision of mental models theory (MMT) for explaining a novel class of inferences containing the modal term "possibly". For example, *the door is closed or the window is open or both*, therefore, *possibly the door is closed and the window is open* (*A or B* or both, therefore, *possibly(A & B*)). They concluded that their results support MMT. In this comment, it is argued that Hinterecker et al. (2016) have not adequately characterised the theory of probabilistic validity (*p*-validity) on which the new paradigm depends. It is unclear how *p*-validity can be applied to these inferences, which are anyway peripheral to the theory. It is also argued that the revision of MMT is not well motivated and its adoption leads to many logical absurdities. Moreover, the comparison is not appropriate because these theories are defined at different levels of computational explanation. In particular, revised MMT lacks a provably consistent computational level theory that could justify treating these inferences as valid. It is further argued that the data could result from the non-colloquial locutions used to express the premises. Finally, an alternative pragmatic account is proposed based on the idea that a conclusion is possible if what someone knows cannot rule it out. This account could be applied to the unrevised mental model theory rendering the revision redundant.

Keywords: New paradigm, mental models, modal inference, probability logic, epistemic modals,

Paradigms, Possibilities, and Probabilities: Comment on Hinterecker et al. (2016)

Over the last 25 years, a new paradigm has emerged in the psychology of reasoning that has adopted probability theory as a normative standard against which to compare human reasoning performance (Elqayam & Over, 2013; Evans, Handley, & Over, 2003; Chater & Oaksford, 1999; Douven & Verbrugge, 2010; Fugard, Pfeifer, Mayerhofer, & Kleiter, 2011; Hattori, 2016; Oberauer & Wilhelm, 2003; Oaksford & Chater, 1994, 2007, 2009; Oaksford, Chater, & Larkin, 2000; Over, Hadjichristidis, Evans, Sloman, & Handley, 2007). This new paradigm is part of the general Bayesian turn in the cognitive and brain sciences (e.g., Chater & Oaksford, 2008; Chater, Tenenbaum, & Yuille, 2006; Clark, 2013; Doya, Ishii, Pouget, & Rao, 2008; Friston, 2005; Friston, Schwartenbeck, FitzGerald, Moutoussis, Behrens, & Dolan, 2013; Gershman, Horvitz, & Tenenbaum, 2015; Griffiths, Kemp, & Tenenbaum, 2008; Oaksford & Chater, 1998, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011), which began with Anderson's rational analysis approach (Anderson, 1990, 1991). Bayesian rational analyses of reasoning were developed quite early on in the progress of Bayesian cognitive science (Oaksford & Chater, 1994, 1996). The psychology of reasoning studies the inferences that people draw using the logical terms of natural language, examples being *and*, *or*, *not*, *if…then*, *all*, and *some*. The core of the Bayesian new paradigm is to study reasoning from uncertain premises, and to treat even reasoning on logical tasks as probabilistic rather than binary (Adams, 1998; Bennett, 2003; Edgington, 1995). This proposal requires that we assess inferences for probabilistic validity, *p*-validity, rather than classical binary validity. Against the standard of *p*-validity more of people's reasoning appears rational rather than subject to systematic bias.

A new paradigm presupposes an older paradigm that it has pretensions to replace. This older paradigm in the psychology of reasoning was based on binary truth-functional logic, where propositions of any complexity can be assigned a binary truth value, true or false. The most widely applied account of this type was mental models theory (MMT, Johnson-Laird & Byrne, 1991, 2002). On this account, people represent the possibilities in which a complex sentence is true. These complete, or "fully explicit", representations (see (FE) below) licensed as "valid" all and only the inferences sanctioned by classical binary logic. *Mental* models are cut down versions of these fully explicit models that eliminate the negated terms (see (MM) below) and then typically reduce to a single preferred possibility (Hinterecker, Knauff, & Johnson-Laird, 2016, p. 3). These mental models can account for some of the systematic errors seen in peoples' reasoning. This theory has recently been fundamentally revised and extended to account for ordinary people's reasoning with the modal term "possibly" (Hinterecker et al., 2016; Johnson-Laird, Khemlani, & Goodwin, 2015; Khemlani, Hinterecker, & Johnson-Laird, 2017). Hinterecker et al. (2016) compared the adequacy of this revised MMT with *p*-validity for explaining a novel class of inferences containing "possibly" (◊).[1] For example*, the door is closed or the window is open or both*, therefore, *possibly the door is closed and the window is open* (*A or B or both*, therefore, ◊*(A & B)*). They concluded that their results support MMT and not *p*-validity.

In this paper, we argue that before drawing such a conclusion the following questions need to be addressed. First, have Hinterecker et al. (2016) adequately characterised the competitor theory? In particular, can *p*-validity be directly applied to these inferences, and do these results challenge the core of *p*-validity, or are they a peripheral issue? Second, is the revision of MMT that allows these predictions well motivated? Third, is the comparison between theories appropriate given the computational level at which each theory is framed? Fourth, in their own terms, how successfully do these results discount the probabilistic approach? Finally, is there a better account of their results?

Probabilistic Validity

Have Hinterecker et al. (2016) adequately characterised the competitor theory? The core of the new paradigm probabilistic approach to human reasoning is the Equation (Adams, 1998; Bennett, 2003;

[1] Because the term "possibly" will recur many times in this paper, from here on we use the traditional logical symbol ◊ for this term.

Edgington, 1995). The Equation states that the probability of a natural language indicative conditional, *if you turn the key, then the car starts*, is the conditional probability of the car starting given you turn the key, Pr(*if A then B*) = Pr(*B*|*A*). A conditional that satisfies the Equation is a probability conditional (Adams, 1998), and it stands in contrast to the material conditional of classical binary logic, $A \supset B$, which is true if and only if its antecedent is false ($\neg A$, using ¬ for "not") or its consequent (*B*) is true. Our example would then be true if and only if you do not turn the key or the car starts. Consequently, the probability of the material conditional, $\Pr(A \supset B)$, is the probability that the antecedent is false or the consequent is true, that is, $\Pr(A \supset B) = \Pr(\neg A$ *or* $B)$.

According to the Equation, a probability conditional is not truth-functional, and its valid inferences are not identical with the valid inferences for the material conditional. For a logically valid inference in binary logic, if the premises are true the conclusion must be true. The probability conditional is inexpressible in binary logic, and so inferences containing it cannot be assessed there for validity, but they can be evaluated for probabilistic validity, *p-validity* (Adams, 1998). A single premise argument is *p*-valid if and only if the probability of its premise cannot be coherently (consistent with probability theory) greater than the probability of its conclusion: Pr(*premise*) ≤ Pr(*conclusion*). For example, take the logically valid inference: supposing a car is *black and a Citroen,* it is *black* (*and*-elimination). This inference is also *p*-valid because the probability of a car being black and a Citroen must be less than (or equal to) the smaller of the probability of being black or the probability of being a Citroen: Pr(*black & Citroen*) ≤ min{Pr(*black*), Pr(*Citroen*)}. To violate this constraint would be to commit the conjunction fallacy (Tversky & Kahneman, 1983). Consequently, *and*-elimination is both classically and probabilistically valid.

However, some inferences that are classically valid for the material conditional are not *p*-valid for the probability conditional (Adams, 1998; Baratgin, Douven, Evans, Oaksford, Over, & Politzer, 2015; Bennett, 2003; Oaksford & Chater, 2007). For an important example, strengthening the antecedent, *if A then B, therefore if* (*A & C*) *then B,* is not *p*-valid for the probability conditional (though it is for the material conditional) because Pr(*B*|*A*) > Pr(*B*|*A & C*) in some cases. It is perfectly coherent to make the following probability assignments, Pr(*the car starts* | *you turn the key*) = .9 and Pr(*the car starts* | *you turn the key* & *the fuel tank in empty*) = 0. But this means that Pr(*the car starts* | *you turn the key*) > Pr(*the car starts* | *you turn the key* & *the fuel tank is empty*), proving that strengthening the antecedent is not *p*-valid for the probability conditional. It is because of the *p*-invalidity of strengthening the antecedent that the probability conditional is non-monotonic: adding information to its antecedent can cause its probability to fall from a high level to zero.[2]

It must be stressed that *p*-validity is itself monotonic, since adding premises to a *p*-valid inference cannot make it *p*-invalid (Baratgin et al., 2015). Transparently for single-premise *p*-valid inferences, if Pr(*premise 1*) ≤ Pr(*conclusion*), then Pr(*premise 1 & premise 2*) ≤ Pr(*conclusion*), on pain of committing the conjunction fallacy. There is no mystery here. Mathematical probability theory establishes the non-monotonicity of the probability conditional, and the axiomatic system for *p*-validity, which is sometimes called *System P* (Pfeifer & Kleiter, 2009), is provably sound (more general than consistent), complete, and decidable (Adams, 1998). The probability conditional is at the centre of the logic that is System P, and it is non-monotonic in the sense we have just explained, but *p*-validity is at the centre of the *metalogic* of System P, and it is monotonic.

The distinction between a logic and its metalogic is critical to analysing logical systems. The metalogic provides the account of how the whole system behaves which depends on the logical language (or object language) and the consequence relation. In System P, the probability conditional is part of the object language, and we describe its properties at the meta-level via the system's consequence relation ($\vdash_p$). So we could express the fact that strengthening the antecedent is not *p*-valid in System P, as *if A then B* $\nvdash_p$ *if A & C, then B.* Strengthening does not hold for a probability conditional *if A then B,* making

[2] There is a further sense in which non-monotonicity can arise when interpreting the conditional as in the Equation. This is when new information suggests that the conditional probabilities themselves have changed. This is called dynamic inference (Oaksford & Chater, 2007, 2013)

it non-monotonic, but the consequence relation, $\vdash_p$, of System P is monotonic, just like the consequence relation for classical logic, $\vdash$.

Problems for Hinterecker et al.'s Exposition

There are a variety of problems created by Hinterecker et al.'s discussion of the probabilistic approach.

Non-Monotonicity. We have been careful introducing *p*-validity because Hinterecker et al. (2016, p. 1608) have not been. They argue that "…p-logic is monotonic (Chater & Oaksford, 2009; Over, 2009) contrary to Adams's avowed purpose in developing it (Adams, 1998, p. 3), though it has inspired nonmonotonic systems of reasoning…" We are not sure how Hinterecker et al. (2016) use the term "p-logic". But if it is used to describe any logic that includes a probability conditional, then p-logic is non-monotonic, contrary to Hinterecker et al. (2016). Hinterecker et al. seem to be confounding, the logic of the probability conditional, System P, with its metalogic, where *p*-validity is defined, again as we have just explained. Using "p-logic" for both the logic and metalogic of System P will only cause confusion in the psychology of reasoning.

Core and Periphery. Having a fuller understanding of the new paradigm probabilistic approach shows that Hintercker et al. (2016) are only criticising it at the periphery, where, in general, it agrees with standard logic. They do not focus their criticisms on the core, i.e., the probability conditional. Nonetheless, their results are potentially interesting because of the attempt to relate modal notions like possibility and necessity to probabilities.

Probabilities of Possibilities. To apply the normative concept of *p*-validity to inferences like *A or B or both*, *therefore*, ◊(*A & B*), requires a normative understanding of how we should assign probabilities to possibilities. Can we establish normatively, for *p*-validity, that Pr(◊(*A & B*)) cannot be strictly less than Pr(*A or B or both*)? For non-modal premises and conclusions, we can do this deductively. It is possible to infer deductively a coherence interval for the probability of the conclusion given the premises. For example, for the *and*-elimination inference, *Citroen and black*, *therefore*, *black*, the probability of the conclusion, Pr(*black*), must be in the inclusive interval between Pr(*Citroen and black*) and 1 (Pr(A) ∈ [Pr(A & B), 1]). These coherence intervals follow from the axioms of probability theory as a matter of deductive logic. Once inferred, we can show whether the lower bound is greater than or equal to the probability of the premises (Gilio & Over, 2012; Oaksford & Chater, 2017; Pfeifer & Kleiter, 2009; Politzer, 2016). However, as Hinterecker et al. (2016, p. 1607) observe, there is no modal logic (Garson, 2016) in which the inference from *A or B or both* to ◊(A & B) is valid. Consequently, the constraints the premise imposes on the coherence interval for the conclusion are unknown.

There would appear to be two possibilities for how we should interpret Pr(◊(A & B)). First, the standard view is that probability is a refinement of the modality of possibility. If X is impossible, Pr(X) = 0, if X is necessary Pr(X) = 1, and if X is possible, Pr(X) > 0. Hinterecker et al. do not consider this possibility, in which Pr(◊(A & B)) is a type of iterated modality (a relevant reference is Demey, Kooi, & Sack, 2017, on modal probability logic). Second, participants could interpret the question as, “what is the probability that a contingent proposition is possible?”, and not, “what is the probability of a possibility?” A contingent statement X is always logically possible, even if ¬X is true in fact. All it means to say that X is logically possible is that X is not logically inconsistent. Thus for contingent statements, Pr(◊X), and the special case of Pr(◊(A & B)), should be assigned a probability of 1 for logical possibility. Hinterecker et al. are unclear about this point, but they are apparently not referring to the extremely weak notion of logical (or alethic) possibility as the fundamental modality in their system. Khemlani et al. (2017) state, “The present paper focuses on a different notion of possibility: ‘epistemic’ possibilities concern possibilities that are consistent with a reasoner’s personal knowledge.”

Even so, Hinterecker et al. (2016, p. 1611) do countenance the possibility that Pr(◊(A & B)) could be assigned a probability of 1. They note that ◊(A & B) holds “… in any of the contingencies in the JPD, and so its probability could be as high as 100%.” But if this is the correct normative position, then Hinterecker et al. are incorrect in central statements that they make about the *p*-validity of the modal inferences they study. Hinterecker et al. (2016, p. 1611) claim, for example, that it is not *p*-valid to infer ◊(A & B) from *A or B but not both*. This inference, however, would be *p*-valid in their materials for judgments about logical possibility. The reason is that Pr(◊(*A* & *B*)) would always be 1 for their jointly contingent materials. Consequently, Pr(*A or B but not both*) ≤ Pr(◊(*A* & *B*)) = 1, making the inference *p*-valid and not *p*-invalid. In summary, without a normative theory of how people *should* interpret Pr(◊(*A* & *B*)) in their materials, Hinterecker et al. have no account of when inferences using epistemic possibilities are *p*-valid or not.

Empirical Probabilities. The problem we have just identified perhaps motivated Hinterecker et al. to also ask participants for empirical estimates of the probabilities Pr(*premise*) and Pr(*conclusion*). Ostensibly this was to check whether people regard these inferences as *p*-valid or not. But all we could learn from these data, *if we had a normative theory of p-validity for these inferences*, is whether these probability assignments were coherent or not. Without this normative theory, these empirical probabilities are just studies of people's intuitive probability estimates in the absence of a theory of whether they are right or wrong.

A question that we can address is, if participants' estimates are such that Pr(*premise*) ≤ Pr(*conclusion*), do they accept the inference and reject it otherwise? That is, regardless of the actual *p*-validity of the inference, do participants show some sensitivity to the impact of judging that Pr(*premise*) ≤ Pr(*conclusion*) on whether they say the conclusion follows from the premises? Hinterecker et al. (2016, p 1618) say, "The participants' estimates of the probabilities of the premises and conclusions did not determine their acceptance or rejection of inferences." While this was true of the limited set of four inferences investigated in Experiments 1 and 2, when broadened out to twelve further inferences in Experiment 3, "*p*-valid inferences were accepted reliably more often than *p*-invalid inferences" (Hinterecker et al., 2016, p 1616). Inspection of their Table 8 also suggests that inferences judged *p*-invalid were rejected reliably more often than inferences judged *p*-valid. Perhaps unsurprisingly, this result held primarily for the non-modal inferences from the premise *A* & *B*. In summary, despite Hinterecker et al.'s claims to the contrary, by their own admission, they detected significant sensitivity to the import of judging whether Pr(*premise*) ≤ Pr(*conclusion*) for the acceptability of inferences.

Summary. Hinterecker et al. conflate the non-monotonic probability conditional in System P with the monotonic meta-logical concept of *p*-validity. They compare MMT with *p*-validity only at the periphery, not for the core probability conditional. They provide no normative account of the *p*-validity of the modal inferences they study, so these data cannot address whether people conform to *p*-validity or not. However, contra Hinterecker et al., participants do seem sensitive to the import of judging that Pr(*premise*) ≤ Pr(*conclusion*) for some inferences.

One might argue that the lack of a normative account of *p*-validity for these inferences is all the worse for the new probabilistic paradigm. However, this can only count in favour of the alternative MMT account if the recent revision that makes predictions for these modal inferences is well motivated.

Revised Mental Models Theory

The original version of MMT was tied to binary extensional logic (e.g., Johnson-Laird & Byrne, 1991, 2002). The theory represented the fully explicit meaning of a compound sentence, for example, *A or B or both*, by what was equivalent to its logical *disjunctive normal form* (DNF): (A & $\neg B$) *or* ($\neg A$ & B) *or* (A & B). Each disjunct (connected by *or*) is the conjunction (&) of literals (variables or their negations) that make the compound sentence true, and these conjunctions, therefore, correspond to the lines of a truth table in which the compound sentence is true. The fully explicit model represents each disjunct in a row and removes the conjunctions as in:

(FE)

$A \quad \neg B$

$\neg A \quad B$

$A \quad B$

Mental models do not represent the negated cases as in:

(MM)

A

B

$A \quad B$

This version of MMT accepted classical validity: if the premises are true, then the conclusion must be true.

This earlier version of MMT faced a serious problem. It accepted all the valid inferences for the material conditional ($A \supset B$) as also "valid" for the natural language conditional *if A then B*. Consequently *if A then B* supposedly followed "validly" from *not*-*A* and *B* as separate premises. Logicians refer to these inferences as "the paradoxes of the material conditional (material implication)". They are paradoxical because, for example, they entail that *if you become a chain smoker, then your health will improve* follows validly from *you will not become a chain smoker.* These inferences are not *p*-valid for the probability conditional (Adams, 1998), which therefore strongly challenged the MMT account of the natural language conditional (Evans & Over, 2004; Oaksford & Chater, 2007).

In the new, revised version of mental models theory, a compound sentence like a disjunction is held to be true, not when one of the disjuncts in its DNF holds, but only "…provided that each of these three cases is possible" (Johnson-Laird et al., 2015, p. 204). Thus the representation of *A or B* becomes equivalent to a *conjunction of possibilities* (COP), which can be written (see, Khemlani et al., 2017) in the following way: $\Diamond(A \,\&\, \neg B) \,\&\, \Diamond(\neg A \,\&\, B) \,\&\, \Diamond(A \,\&\, B)$. This representation is the default, because "modulation" can "block" a conjunction that is impossible (Hinterecker et al., 2016). Hinterecker et al. (2016) write the possibilities in a COP in a list in the just same way as in the old version of *mental models* omitting negated variables (see (MM)). Doing so masks the radical change from a DNF to a COP. This change renders "invalid" some inferences formerly regarded as valid in the old version of the theory. In particular, agreeing with *p*-validity, the paradoxes of the material conditional are no longer considered valid. This revision also allows MMT to make predictions about people's modal inferences from disjunctions and other compound sentences (Hinterecker et al., 2016; Khemlani et al., 2017). Given the change to interpreting (FE) as a COP, inferences like *A or B or both*, *therefore*, ◊(*A & B*) follow by definition.

Problems with the Revised MMT

This radical revision of MMT faces immediate problems, some of which we have raised already (Baratgin et al., 2015).

Baratgin et al. (2015). Baratgin et al. (2015) observed that the new interpretation of *A or B* in terms of COPs immediately lead to range of logical absurdities, some of which we list here (see also Cruz, Over, & Oaksford, 2017).

1. *A or B* and *it is possible that A or B* must share the same mental representation, but surely the latter is weaker than the former.
2. Where *A* and *B* are jointly contingent, just about every ordinary disjunction must be true for logical possibility (and for what Khemlani et al., 2017, seem to mean by "epistemic possibility").
3. An exception is the classical tautology *A or* ¬*A*, which is false because *A &* ¬*A* is not possible, not logically and not epistemically.
4. The fundamental inference of *or*-introduction (*A*, *therefore*, *A or B*), which is *p*-valid, is rendered invalid because "The premise [*A*] does not establish that the second [¬*A & B* in (FE)] and third [*A & B* in (FE)] cases are possible" (Johnson-Laird et al., 2015, p. 205). But again these cases are always logically possible for jointly contingent *A* and *B*.
5. If *or*-introduction is invalid, then *A* and ¬(*A or B*) are consistent. However, ¬(*A or B*) is equivalent to ¬*A* & ¬*B*, and it is absurd to say that this is consistent with *A*. It does not have a model in common with *A*, and MMT itself defines consistency as having at least one mental model in common.
6. People can never commit the disjunction fallacy of judging that Pr(*A*) > Pr(*A or B*) (Bar-Hillel & Neter, 1993). Responding that Pr(*A*) > Pr(*A or B*) is not a fallacy if it is invalid to infer *A or B* from *A*.

These cases illustrate that the revised MMT creates a range of logical absurdities. As we now show, they do not exhaust the problems for the revised MMT.

***Or*-Introduction is judged *p*-valid.** People generally respect the *p*-validity of inferring *A or B* from *A* by responding that $\Pr(A) \leq \Pr(A \text{ or } B)$ (Cruz, Baratgin, Oaksford, & Over, 2015; Cruz et al., 2017). Johnson-Laird et al. (2015) cite evidence that people "baulk" at accepting this inference (Orenes & Johnson-Laird, 2012). But in this paper, a pragmatic explanation for this effect is provided because it was written before MMT was revised when *or*-introduction was still considered valid. Their earlier account seems more consistent with the evidence than the revised MMT (Cruz et al., 2017).

Inclusive and exclusive *or*. An important discriminatory prediction for Hinterecker et al. (2016, p. 1611, Table 1) is that *A or B but not both, therefore, A or B or both* is not valid in the revised MMT. The failure of this inference was a key prediction in Experiments 1 and 2. If this is the case, then ¬(*A or B or both*) must be consistent with *A or B but not both*, with the immediate consequence that ¬*A* & ¬*B* must be consistent with *A or B but not both*, which is absurd.

***A or B, therefore,* ◊*A*.** This inference was investigated in Hinterecker et al.'s Experiment 3, and it is regarded as "valid" in revised MMT. Like the exclusive-*or* to inclusive-*or* inference in the last section, the claim that this is a valid inference leads to some absurdities. So, if valid, *A or B* is inconsistent with ¬◊*A*, but *A or B*, ¬◊*A*, *therefore B* is a common inference pattern for epistemic possibility, and not because the premises are inconsistent. For example, suppose we hear from a reliable source that Mary has an essay to write, and is in the library or working on her computer (or both). Then we learn from the library website that it is not possible that Mary is in the library: it is closed today. This example provides a clear epistemic use of possibility as defined by Khemlani et al. (2017): it is based on our "personal knowledge" of the website. But we can now clearly conclude that Mary is working on her computer. Moreover, given that *A or B* is inconsistent with ¬◊*A* for this epistemic use, the following derivation must be "valid" in revised MMT: ¬◊*A*, *therefore* ¬(*A or B*), *therefore,* ¬*A* & ¬*B*, *therefore,* ¬*B*. Hence from the fact that Mary, by our personal knowledge, cannot possibly be in the library, we can infer ¬*B* for any arbitrary *B* in revised MMT, for example, $1 + 1 \neq 2$.

Subadditivity. In Hinterecker's et al.'s experiments participants were also asked to judge the probability of all four cases in the joint probability distribution (JPD) for two propositions *A* and *B*, when described as possibilities (Experiment 1) and when not so described (Experiment 2). They found that people's judgements were highly subadditive, that is, they summed to considerably more than 1. While this result has no immediate bearing on the inferences people should or did draw in these experiments, Hinterecker et al. (2016) argue that this result uniquely supports the revised MMT theory. However, the theory of how people judge the subjective probabilities of the singular events used in their materials (Khemlani, Lotstein, & Johnson-Laird, 2012, 2015) forms no part of the revised MMT. It is an ad hoc and detachable theory to which other psychological theories could just as well appeal.[3]

The finding of subadditivity for judging the possibility of the four JPD cases is also not independent of the lack of a normative account of the probability of possibilities. As we argued in the section, *Probabilities of Possibilities*, the standard account is that if X is possible, then the only constraint on $\Pr(X)$ is that it is greater than 0 and could be 1. So it is unclear how this finding argues against the probabilistic approach. Hinterecker et al. claim further that the finding of subadditivity when the JPD cases were not described as possible (Experiment 2) counts against the probabilistic view. But all of the materials used in Hinterecker et al. were *future contingent* statements (Øhrstrøm & Hasle, 2015) about possible future events, for example, "A nuclear weapon will be used in a terrorist attack in the next decade". It is arguable that people can have some evidence to judge such a statement likely, and some other evidence to judge its negation likely, and so their overall subadditive probability judgments, while incoherent, would be explicable in a probabilistic account.

[3] However, we would argue that recent research in the Bayesian paradigm on using restricted Markov Chain Monte Carlo (MCMC) sampling with typical vs. atypical exemplars provides a more compelling account of the cognitive algorithms underlying subadditivity and superadditivity effects in probability judgement (Dasgupta, Schulz, & Gershman, 2017).

Levels of Explanation

The standard account of computational explanation in the cognitive sciences is multi-levelled (e.g., Anderson, 1990; Marr, 1982). The computational (Marr, 1982) or rational (Anderson, 1990) level specifies a normatively justified and descriptively adequate model for some cognitive phenomenon. The computational level is implemented at the algorithmic or performance level. At this level, the choices of representations and processes to implement the computational level theory may introduce deviations from the prescriptions of the computational level theory. These deviations can be because of the complexity profile of the algorithm, or because of the nature of the representations themselves. System P and standard binary logic are both computational level theories of the inferences that people should draw. The previous version of MMT fitted clearly into this explanatory scheme. The computational level theory of MMT was binary logic. When mental models were complete and fully explicit, they captured the DNF of the premises, and hence licensed all and only the inferences licensed by binary logic. However, *mental* models only partially represent the DNF, and this can lead to systematic deviations from binary logic. MMT was an algorithmic theory.

Problems with Levels

Several consequences follow from considering how the theories compared in Hinterecker et al. (2016) fit into this scheme for computational explanation in the cognitive sciences.

Comparing different levels. Hinterecker et al. (2016) compare a computational level theory, what they called "p-logic," with an algorithmic level theory, revised MMT. This comparison is inappropriate as no one holds that "p-logic" systems are, unmodified, adequate descriptive, algorithmic theories. Evans, Thompson, and Over (2015) and Cruz et al. (2015) set out to discover the conditions under which people may better conform to the requirements of coherence and *p*-validity. These studies presupposed that people do not always conform to the prescriptions of these normative, computational level theories, and indeed they contained two results on incoherence that revised MMT has not so far explained. Evans et al. found subaddivity caused by negation in their materials. And Cruz et al. showed that not even explicit inference helped people to avoid the conjunction fallacy: people violated *p*-validity and became incoherent even when explicitly inferring *A* from *A & B* in the context of the Linda problem known to produce the conjunction fallacy (Tversky & Kahneman, 1983). Hinterecker et al. (2016) mention both Evans et al. and Cruz et al., but do not acknowledge that these researchers did not expect people to conform exactly to "p-logic". In fact, Evans et al. and Cruz et al. provide more convincing evidence for this conclusion than Hinterecker et al. on inferring ◊(A & B) from *A or B*, because Hinterecker et al. fail to specify what Pr(◊(A & B)) should normatively be in "p-logic".

No computational level theory of revised MMT. The lack of a normative, computational level theory of Pr(◊(A & B)) could be argued to count against the probabilistic approach as well as limit the conclusions Hinterecker et al. can draw (see, the summary of the section on *Probabilistic Validity*). However, revised MMT is in the same position; it lacks a normative, computational level theory of the modal inferences licensed by interpreting mental models as COPs. We have already pointed out that many inferential absurdities arise from adopting this interpretation.

It is possible that these absurdities do not arise. But to establish that this is the case, Hinterecker et al. require a specification of the sound, complete, and decidable normative system which revised MMT implements. Previously, fully fleshed out mental models, interpreted as DNFs, respected classical binary logic. For revised MMT, we need to know what normative theory fully fleshed out mental models respect when interpreted as COPs? It cannot be classical binary logic as the COPs interpretation rules out at least three inferences which this normative theory licenses as valid.

Adding valid inferences. A striking feature of revised MMT is that inferences that are invalid in standard normative modal logics are classified as "valid". This classification is in contrast with System P, in which the adoption of the probability conditional implies that some classical inferences for the conditional are invalid. If to derive a plausible psychological theory of reasoning, we restrict the number of valid inferences in a provably consistent normative system, we do not have to worry about the consistency of these restricted procedures. If a system is provably consistent, a sub-system of it must be too. However, when we *add* supposedly "valid" inferences to try to obtain a plausible theory, the question of consistency does arise. Do the new “valid” inferences lead to a contradiction in the extended system? The absurdities we have easily derived from the revised MMT imply that this is a pressing question for MMT.

Validity and “validity.” The problem we have just identified might be defused if Hinterecker et al. argued that they do not mean to assert that these extra modal inferences are normatively valid. But Khemlani, et al. (2017) are perfectly clear that the revised MMT has a normative position on the “validity” of the inference from *A or B or both* to ◊(*A & B*), which they label (7), and the “invalidity” of the inference from *A or B but not both* to *A or B or both*, which they label (8). They point out that (7) is invalid, (8) is valid, in well-known modal logics, but they add, “On our account, reasoners are justified in feeling that the invalidity of (7) and the validity of (8) are counterintuitive and incorrect.” Hinterecker et al. and Khemlani et al. ascribe biases and errors to people (e.g., Hinterecker et al. make a number of references to belief bias), and they claim that the system of fully explicit models can "correct the errors and biases" (Khemlani et al., 2017) of the fast system of mental models. But they do not justify the system of fully explicit models itself as a normative system, which should have, at the very least, a consistency proof, like the systems of modal logic that they discuss. Khemlani et al. refer specifically to modal system T (they also mention system K, but K is too weak to express epistemic modality). There are consistency, soundness, completeness, and decidability results for system T (Chellas, 1980).

In the absence of a consistency proof for revised MMT, there is no explanation of why it has evolved in human beings, or been otherwise acquired, as a reasoning system (Anderson, 1990). It is supposed to be a system for reasoning with *not*, *and*, *or*, *if*, and *possibly*, and at a minimum such an elementary system should be able to detect and eliminate inconsistencies in reasoning with these terms, at least under certain conditions (when the models are fully explicit). This is impossible if the proposed system is itself inconsistent.

Illusions and the computational level. Hinterecker et al. (2016) cannot retreat into descriptivism (Elqayam & Evans, 2010), claiming that they are giving a scientific account of what might be people’s inconsistent reasoning system. First, their scientific account would have to be consistent, and second, this would be to abandon the MMT account of cognitive illusions of reasoning (Cohen, 1981). These illusions are inferences people accept (reject), although logic says that they are invalid (valid) because their partial mental models suggest they are valid (invalid) (Johnson-Laird, 2001; Khemlani & Johnson-Laird, 2017). As Khemlani and Johnson-Laird (2017, p. 11) argue, these partial mental models "…usually represent what is true in a possibility, not what is false. This procedure reduces the load on working memory, and for the most part, it yields valid inferences." We need to know with respect to what provably consistent system revised MMT "for the most part…yields valid [modal] inferences."

What then is the status of the novel modal inferences Hinterecker et al. investigate? We do not know of a normative theory with respect to which they are valid. On the one hand, they could be cognitive illusions. On the other hand, Hinterecker et al. seem to want to regard them as valid inferences, which, as we have seen, leads to many logical absurdities. Another possibility is that the intuitions that their experiments have apparently uncovered are just that, inferential intuitions that require some explanation but which in the absence of a normative theory we cannot identify as valid inferences or cognitive illusions.

Summary. Hinterecker et al. (2016) inappropriately compare "p-logic" with the revised MMT, which are defined at different levels of computational explanation. Proponents of the probabilistic approach have never represented it as a descriptive algorithmic level theory (see, for example, Oaksford & Chater, 2003). Moreover, they have provided better evidence than Hinterecker at al. (2016) that, although a better descriptive theory than classical binary logic, *p*-validity is not fully descriptive of human reasoning. Hinterecker et al. (2016) do not provide a computational level theory for revised MMT, which is problematic given they appear to want to add normatively valid modal inferences. That is, MMTs interpreted as COPs apparently yield valid modal conclusions that Hinterecker et al. want to argue are not cognitive illusions. In the absence of a provably consistent computational level theory, it is impossible to make good on this claim. As things stand, as we have demonstrated in previous sections, the claim that these modal inferences are valid leads to many logical absurdities.

The Experimental Data

Hinterecker et al. present three experiments. In their Experiments 1 and 2, they investigated inferences from inclusive and exclusive disjunctions to modal conclusions: (1) *A or B or both*, *therefore* ◊(*A* & *B*), (2) *A or B or not both*, *therefore* ◊(*A* & *B*), inferences from exclusive to inclusive disjunctions: (3) *A or B or not both*, *therefore, A or B or both*, and inclusive to exclusive disjunctions: (4) *A or B or both*, *therefore, A or B or not both* (we use the same numbering as, Hinterecker et al. 2016: Table 1). Experiment 3 looked at four inferences for each of three different premises: *A or B or both*, *therefore,* ◊*A*, ◊*B*, ◊(*A* & *B*) , or ◊(¬*A* & ¬*B*) (5 - 8)*, A* & *B*, *therefore, A*, *B*, *A or B or both*, or ¬*A* & ¬*B* (9 - 12), *not both A* & *B*, therefore, ◊*A*, ◊*B*, ◊(¬*A* & ¬*B*), or *A* & *B* (13 - 16).

The Materials.

We make a couple of observations about the materials used which leads to a conjecture about the results. First, for the non-modal inferences (9 -12), *p*-validity captured the data well. Moreover, participants judging that Pr(*conclusion*) ≥ Pr(*premise*) was a good predictor of acceptance and rejection rates. Second, endorsement of the remaining modal inferences seems to be far more related to the surface form of the non-colloquial expressions used for inclusive (*A or B or both*) and exclusive-*or* (*A or B or not both*), rather than to deep semantic, COP representations. We checked the occurrence of the expressions "or both" and "or not both" against the British National Corpus (2015) of 100M words of text and spoken language. We also checked "but not both" which is perhaps a more colloquial expression. Out of the 355,615 occurrences of "or" it was accompanied by "both" on 0.23% (819), by "not both" on 0.0006% (2), and by "but not both" on 0.007% (26) of occasions.[4] Overwhelmingly, natural language does not mark inclusive and exclusive *or* in these ways. Context is left to disambiguate between the two

[4] Most of these occurrences were in technical or legal documents.

interpretations. To properly test whether people spontaneously draw these inferences in the real world would involve just using "or" and allowing context to disambiguate.

Surface Form.

The inclusion of these locutions means that the modal conclusions that participants were asked to accept or reject are described explicitly in the surface form of the premises for *A or B or both* ($\Diamond A$ or $\Diamond B$ or $\Diamond(A \,\&\, B)$). This diagnosis applies equally to *not both A & B,* as it only excludes *A & B*. Consequently, participants endorse any other conclusion with which they are presented as possible. We also conjecture that people may also accept the same conclusions in the absence of the modal operator. There is evidence that although people endorse the valid inference *A & B, therefore, A or B* more than the invalid inference *A or B, therefore, A & B*, they still endorse the latter at greater than 50% (Cruz, Over, Oaksford, & Baratgin, 2016). Consequently, it would be useful to repeat these experiments without the modal operators in the conclusions. Of course, if people endorse the non-modal conclusions, then they should also endorse the modal conclusion. This inference is natural because, for example, *it is raining* (X) is a stronger claim than *it is possible that it is raining* ($\Diamond X$), and so *it is possible that it is raining* could be thought of as more probable than *it is raining*, $\Pr(\Diamond X) \geq \Pr(X)$, and then *it is raining, therefore, it is possible that it is raining* (X, *therefore*, $\Diamond X$) would be *p*-valid.

A Final Proposal.

The last conjecture above does not lead to a full theory of epistemic modal inferences. But we should recall that other than redefining mental models as COPs, revised MMT does not provide such a theory either. We, therefore, present a final proposal based on logic programming and natural language semantics that may provide the beginnings of a theory.

Negation as Failure.

Logic programming uses the negation-as-failure procedure (Clark, 1978; Stenning & van Lambalgen, 2005): if A cannot be derived ($\vdash$) from the content of the database consisting of an agent's world knowledge (Γ), then $\neg A$ can be inferred ($\Gamma \nvdash A$, *therefore,* $\neg A$). If we substitute $\neg A$ for A, then we arrive at the classical epistemic fallacy of the argument from ignorance, $\Gamma \nvdash \neg A$, *therefore,* A. For example, Ghosts exist because we can not prove that they do not (Chater, Oaksford, Hahn, & Heit, 2011; Hahn & Oaksford, 2007; Oaksford & Hahn, 2004). The argument from ignorance is an epistemic fallacy because not being able to prove that Ghosts do not exist from what you know does not mean that they do exist. An epistemically more prudent conclusion to draw from not being able to prove that Ghosts do not exist is that *possibly Ghosts exist* ($\Gamma \nvdash \neg A$, *therefore,* $\Diamond A$), that is, you cannot rule it out. If all we know is what we are told in the premises, for example, Γ = *A or B*, then *A or B* $\nvdash \neg(A \,\&\, B)$, *A or B* $\nvdash \neg A$, and *A or B* $\nvdash \neg B$. Hence the epistemically prudent conclusions are $\Diamond(A \,\&\, B)$, $\Diamond A$, and $\Diamond B$ respectively. As we now show, these epistemically prudent conclusions may arise, not from the meaning of a disjunction, but from pragmatic considerations.

Evidentiality and Indirectness Signalling.

Epistemic modals are frequently used in English to mark a source of evidence as indirect and inferred (Aikhenvald, 2004; von Fintel & Gillies, 2010).[5] We suspect that this may be the function they are performing in Hinterecker et al.'s (2016) materials, which is consistent with the analysis we have just provided. For example, on seeing people entering the building shaking their umbrellas, you have indirect evidence to infer "It must be raining", in one ordinary use of "must", but not direct evidence for asserting "It is raining." This use of the modal "must" indicates this claim is inferred from the evidence. This

[5] There is empirical work emerging on evidentiality in the psychology of reasoning, for example, Krzyżanowska, Wenmackers, & Douven (2013) and Karaslaan, Hohenberger, Demir, Hall, & Oaksford (in press).

function captures our example about studying in the library. If the library is closed, one can infer that "it is not possible that Mary is there," therefore, she must be at home. However, what if you know nothing that would allow you to infer she is not in the library? You would have to concede that she might be in the library, that is, it is possible. This natural language evidential function of modals suggests that "Mary is at home" can be a stronger claim than "Mary must be at home" in some uses, whereas in standard modal logic, it is always the other way round. This puzzle has been called "Karttunen's Problem" after the philosopher who first pointed it out (Karttunen, 1972). There are possible semantic solutions to it (von Fintel & Gillies, 2010), but it has been conceded that there remains uncertainty about whether "…the indirectness signal [derives] from conversational principles [it is pragmatic] rather than hardwiring it into the semantics of epistemic modals" (von Fintel & Gillies, 2010, p. 381). That is, this could well be a pragmatic phenomenon. Knowing the disjunction can give people pragmatic grounds for asserting $\Diamond(A \& B)$, because they cannot rule out $A \& B$. They therefore pragmatically make the inference without further information, for example that A is impossible (the library is closed). In this analysis, $\Diamond(A \& B)$ is not a deductively valid consequence of the disjunction.[6]

In summary, there is a more principled and generalizable way of drawing the epistemic modal conclusions Hinterecker et al. (2016) observed other than by introducing COPs in the revised MMT. A participant need only check the DNF to see whether the negation of the conclusion follows. If not, then there is indirect inferred evidence for the claim because it cannot be ruled out, which would commonly be marked by the modal "possibly" in natural language. This option potentially leaves MMT just where it was, with classical binary logic as its computational level theory, but it has a downside. The paradoxes of material implication would remain valid in MMT, despite evidence that people do not endorse them (Cruz et al. 2017).

[6] At the very least, Karttunen's problem shows that modals, like "must" are ambiguous between the *necessity* and the *evidential* reading. It is possible that an appropriate semantics for the evidential reading can be developed, perhaps tied to the interpretation of conditionals (Krzyżanowska et al., 2013). However, as von Fintel and Gillies (2010) observe, it is also possible that this is a pragmatic phenomenon.

Conclusion

In this comment, we have argued that Hinterecker et al. (2016) have not adequately characterised the theory of probabilistic validity (*p*-validity). They do not show how to apply *p*-validity to their modal inferences, which are anyway peripheral to the theory. We also pointed out that revised MMT is not well motivated and its adoption leads to many logical absurdities. We further argued that the comparison between *p*-validity and revised MMT is not appropriate because these theories are defined at different levels of computational explanation. In particular, revised MMT lacks an appropriate computational level theory, with a consistency proof, that could justify treating these inferences as valid. Furthermore, Hinterecker et al.'s data could result from the non-colloquial locutions used to express the premises. Finally, we sketched an alternative account based on logic programming and the evidential signalling function of epistemic modals, which suggests that the modal conclusions Hinterecker et al. observed are the result of these conclusions being indirectly inferred. That is, it is a pragmatic phenomenon. This account could be applied to the unrevised mental model theory, interpreted as DNFs, rendering the revision, using COPs, redundant.

The original appeal of MMT was that, under certain conditions (fully explicit models), this performance theory fully implemented its computational level theory of binary logic. Thus people were, in principle, capable of rational thought but often erred because they represented premises only partially. In revised MMT, interpreted as COPs, this connection is broken. No provably consistent system has been presented that revised MMT implements when fully explicit. This omission makes it impossible to interpret Hinterecker et al.'s claim (along with those of Khemlani et al.) to have identified novel and *normatively valid* modal inferences. We suspect that these inferences are pragmatic, resulting from the indirectness signalling function of epistemic modals. MMT does need to respond to the challenge of the new probabilistic paradigm, but the current revision of MMT does not appear to be the most promising avenue to go down.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford: Oxford University Press

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14,* 471-517. doi:10.1017/S0140525X00070801

Baratgin, J., Douven, I., Evans, J. St.B. T., Oaksford, M., Over, D., & Politzer, G. (2015). The new paradigm and mental models. *Trends in Cognitive Sciences*, *19*, 547-548. doi:10.1016/j.tics.2015.06.013

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford England: Oxford University Press.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: a disjunction fallacy in probability judgments. *Journal of Personality & Social Psychology*. *65*, 1119–1131.

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

Chater, N., & Oaksford, M. (Eds.) (2008). *The probabilistic mind*: *Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.

Chater, N., & Oaksford, M. (2009). Local and global inferential relations: Response to Over (2009). *Thinking & Reasoning*, *15*, 439-446.

Chater, N., Oaksford, M., Heit, E., & Hahn, U. (2011). Inductive logic and empirical psychology. In D. Gabbay, S. Hartmann, & J. Woods (Eds.), *The Handbook of the History of Logic*: *Inductive Logic* (Volume 10) (pp. 553-624), Amsterdam: North Holland.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*, 287-291. doi:10.1016/j.tics.2006.05.007

Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences. 36*, 181-253

Clark, K. L. (1978). Negation as failure. In H. Gallaire & J. Minker (Eds.), *Logic and databases* (pp. 293-322). New York: Plenum Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4*, 317-370. doi:10.1017/S0140525X00009092

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology, 6*, 1–9.

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

Cruz, N., Over, D. E., & Oaksford, M. (2017). The elusive oddness of *or*-introduction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 663-668). Austin, TX: Cognitive Science Society.

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25.

Demey, L., Kooi, B., & Sack, J. (2017). Logic and probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). Downloaded from <https://plato.stanford.edu/archives/sum2017/entries/logic-probability/>.

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, *117*, 302-318. doi:10.1016/j.cognition.2010.08.015

Doya, K., Ishii, S., Pouget, A., & Rao, P. N. (2006). *The Bayesian brain*. Cambridge, Mass.: MIT Press.

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

Elqayam, S., & Evans, J. St.B. T. (2011). Subtracting ‘ought’ from ‘is’: Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*, 233-248. doi:10.1017/S0140525X1100001X

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by S. Elqayam, J.F. Bonnefon, & D. E. Over. *Thinking & Reasoning, 19*, 249-265.

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

Evans, J. St.B. T., & Over, D. E. (2004). *If*. New York, NY, US: Oxford University Press. doi:10.1093/acprof:oso/9780198525134.001.0001

Evans, J. St.B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology, 6*, 398.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*, 815–36.

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Frontiers in Human Neuroscience*. *7*:598. doi: 10.3389/fnhum.2013.00598

Fugard, J. B., Pfeifer, N., Mayerhofer, B., & Kleiter, G. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*, 635-648.

Garson, J. (2016). Modal logic. E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). Downloaded from <https://plato.stanford.edu/archives/spr2016/entries/logic-modal/>.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds and machines. *Science*, *349*, 273-278. doi: 10.1126/science.aac6076

Gilio, A., & Over, D.E. (2012). The psychology of inferring conditionals from disjunctions: A probabilistic study. *Journal of Mathematical Psychology, 56*, 118-131.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59-100). New York, NY, US: Cambridge University Press. doi:10.1017/CBO9780511816772.006

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114,* 704-732.

Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate mental models and heuristics. *Cognition, 157*, 296-320. doi:10.1016/j.cognition.2016.09.009

Hinterecker, T., Knauff, M., Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology*: *Learning, Memory, & Cognition*, *42*, 1606-1620.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5*, 434–442.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646-678. doi:10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Response to Baratgin et al.: Mental models integrate probability and deduction. *Trends in Cognitive Sciences*, *19*, 548-549. doi:10.1016/j.tics.2015.06.014

Karaslaan, H., Hohenberger, A., Demir, H., Hall, S., & Oaksford, M. (in press). Cross-cultural differences in informal argumentation: Norms, inductive biases, and evidentiality. *Journal of Cognition and Culture.*

Karttunen, L, (1972). *Possible* and *must*. In J. Kimball (Ed.), *Syntax and semantics*, vol. 1, (pp. 1-20). New York; Academic Press.

Khemlani, S., Hinterecker, T., & Johnson-Laird, P. N. (2017). The provenance of modal inference. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual*

Conference of the Cognitive Science Society (pp. 259-264). Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2017). Illusions of reasoning. *Minds &Machines*, *27*, 11-35.

Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science*, *39*, 1216-1258. doi:10.1111/cogs.12193

Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. (2012). The probabilities of unique events. *PLoS ONE*, 7(10), doi:10.1371/journal.pone.0045975

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

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

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

Oaksford, M. & Chater, N. (Eds.) (1998). *Rational models of cognition*. Oxford: Oxford University Press.

Oaksford, M. & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381-391.

Oaksford, M., & Chater, N. (2003). Computational levels and conditional reasoning. *Journal of Experimental Psychology*: *Learning, Memory & Cognition*, *29*, 150-156.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford, England: Oxford University Press.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking and Reasoning*, *19*, 346-379. DOI:10.1080/13546783.2013.808163

Oaksford, M., & Chater, N. (2017). Probabilities and Bayesian rationality. In, L. J. Ball, & V. A. Thompson (Eds.), *International handbook of thinking and reasoning*. Hove, Sussex: Psychology Press.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 883-899.

Oaksford, M., & Hahn, U. (2004). A Bayesian analysis of the argument from ignorance. *Canadian Journal of Experimental Psychology*, *58*, 75-85.

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*(4), 680-693. doi:10.1037/0278-7393.29.4.680

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*, 431–438.

Over, D. E., Hadjichristidis, C., Evans, J. St.B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*, 62-97. doi:10.1016/j.cogpsych.2006.05.002

Øhrstrøm, P., & Hasle, P. (2015). Future contingents. . In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition). Downloaded from <https://plato.stanford.edu/archives/win2015/entries/future-contingents/>.

Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, *27*, 357-377. doi:10.1111/j.1468-0017.2012.01448.x

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, *7*, 206-217.

Politzer, G. (2015). Deductive reasoning under uncertainty: A water tank analogy. *Erkenntnis*, *81*, 479-506. doi:10.1007/s10670-015-9751-0>

Tversky, A., & Kahneman, D. (1983). Extensional vs intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

von Fintel, K., & Gillies, A. S. (2010). Must . . . stay . . . strong! *Natural Language Semantics*, *18*, 351–383. doi:10.1007/s11050-010-9058-2.