American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/19/$12.00

2019, Vol. 45, No. 2, 298–301
http://dx.doi.org/10.1037/xlm0000627

COMMENTARY

How to Infer Possibilities: A Reply to Oaksford et al. (2018)

Thomas Hinterecker
Max Planck Institute for Biological Cybernetics

Markus Knauff
Justus Liebig University

P. N. Johnson-Laird
Princeton University and New York University

Individuals draw conclusions about possibilities from assertions that make no explicit reference to them. The model theory postulates that assertions such as disjunctions refer to possibilities. Hence, a disjunction of the sort, *A or B or both*, where *A* and *B* are sensible clauses, yields mental models of an exhaustive conjunction of *possibly A*, *possibly B*, and *possibly both A and B*, which each hold in default of information to the contrary. Oaksford, Over, and Cruz (this issue) are critical of the model theory and defend a probabilistic approach to reasoning. In this reply, we deal with their three main claims: (a) *Our results concern only the periphery of their probabilistic theory.* We show that they refute their theory insofar as it applies to possibilities. (b) *The model theory leads to logical absurdities.* We rebut this criticism as it applies to the model theory in Hinterecker, Knauff, and Johnson-Laird (2016), and explain why standard modal logics, which concern possibilities, do not set appropriate norms for inferences about them. (c) *The algorithm for reasoning based on models needs a normative theory.* In fact, it has such a theory, but the demand for "a specification of a sound, complete, and decidable normative system" is chimerical for everyday reasoning.

Keywords: disjunctive reasoning, mental models, modal reasoning, possibilities, probabilities

Suppose you know that you are slightly deaf in your left ear or your right ear. Is it therefore possible that you are slightly deaf in your left ear? Of course it is. Everyone makes such inferences. They also make similar ones that we abbreviate here, using *A* and *B* to stand for sensible propositions, and *not-A* to denote the negation of *A*:

A or B or both.

Therefore, it is possible that A.

Therefore, it is possible that B.

Therefore, it is possible that A and that B.

Therefore, it is impossible that not-A and that not-B.

We reported such inferences in a previous article (Hinterecker, Knauff, & Johnson-Laird, 2016). You may well think that they are unremarkable. We agree. But they matter, because some theorists take for granted that the primordial laws of logic are built into human reasoners, and that inferences that do not follow these precepts are impossible or absurd. This view is mistaken: there are infinitely many modal logics, which deal with possibilities, but they differ in their axioms and interpretations. Yet, in all of them, the inferences above about possibilities are invalid. For example, if *A* is impossible, but *B* is true, then the premise *A or B or both* is nevertheless true in all modal logics, but the conclusion, *it is possible that A*, is false. So, the inference is invalid, because validity demands that a true premise yields a true conclusion.

So, what mental processes underlie these inferences? Our explanation comes from the theory of mental models, which has a long history from Johnson-Laird (1983) onward, but which has undergone recent revisions (see Khemlani, Hinterecker, & Johnson-Laird, 2017). It proposes that a disjunction, *A or B or both*, has mental models of a conjunction of three possibilities:

A	
	B
A	B

They each hold by default, that is, only if there is no information to the contrary. Because the disjunction refers only to these possibilities, reasoners agreed that *not-A and not-B* is impossible (Hinterecker et al., 2016).

Thomas Hinterecker, Human Perception, Cognition, and Action, Max Planck Institute for Biological Cybernetics; Markus Knauff, Department of Experimental Psychology and Cognitive Science, Justus Liebig University; P. N. Johnson-Laird, Emeritus Professor, Psychology, Princeton University, and Visiting Scholar, Psychology, New York University.

Correspondence concerning this article should be addressed to Thomas Hinterecker, Human Perception, Cognition, and Action, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, D-72076 Tübingen, Germany. E-mail: thomas.hinterecker@tuebingen.mpg.de

Oaksford, Over, and Cruz (this issue) take exception to the model theory. We thank them for their critique. We all want to understand reasoning, and their expert criticisms helped us to clarify the theory. They propose a probabilistic approach to reasoning, which is part of a general turn to probabilities in thinking about cognition. The model theory too elucidates reasoning about probabilities, whether numerical (e.g., Khemlani, Lotstein, & Johnson-Laird, 2015) or non-numerical (Fontanari, Gonzalez, Vallortigara, & Girotto, 2014). People say that it is barely possible that Trump will be reelected, and they can give a numerical estimate of the probability; for example, the probability is 20%. A major puzzle is where the numbers in such estimates come from. The model theory postulates that they are derived from models of possibilities, which are a simple and fundamental sort of uncertainty (Khemlani et al., 2015). They are built into the process of human reasoning and are a precursor to numerical probabilities.

Oaksford and his colleagues propose a pragmatic account of possibilities in which "a conclusion is possible if what someone knows cannot rule it out." This claim cannot be the whole story. It has nothing to say about how people formulate their own conclusions about possibilities. And consider this assertion:

> It is possible for Trump to be reelected, but not possible that he will be.

According to the pragmatic account, its first clause implies that the speaker cannot rule out the possibility of Trump's reelection, whereas its second clause implies that the speaker can rule out the possibility. So, the assertion ought to be a self-contradiction, just as it is in any normal modal logic. Yet, it isn't, and it strikes us as true. Another problem for modal logics, and perhaps for probabilities, is illustrated in this assertion:

> Hillary lost, but she might not have.

It is a common claim and could be true. It asserts that Hillary lost, which rules out the possibility that she did not lose. It is a self-contradiction in normal modal logics, because they fail to distinguish between possibilities and counterfactual possibilities, that is, those that were once possible but that did not happen (see Byrne, 2005; Johnson-Laird & Byrne, 2002). Without this distinction, neither modal logics nor the pragmatic account can provide normative prescriptions for human reasoning.

Oaksford and colleagues make three main arguments against the model theory, which we now consider.

Our Results Concern Only the Periphery of Their Probabilistic Theory

They argue that the probabilistic-conditional is the core of their approach, and that as our article does not deal with conditionals—assertions of the sort, *if A then B*—it is peripheral to their theory. But, the probabilistic-conditional is a part of a broader probabilistic logic, which, as our article showed, cannot explain inferences from one sort of disjunction to another:

> A or B but not both.
>
> Therefore, A or B or both.

It is probabilistically valid (p-valid), because as they write: "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)." But, the participants in our experiment rejected the inference on 97% of trials. You might suppose that the clash between the tags, "but not both" and "or both," inhibits the inference, but reasoners make the converse inference, which has the same clash, reliably more often—and the model theory explains why (see Hinterecker et al., 2016).

In Hinterecker et al. (2016), reasoners violated probability logic in more fundamental ways. It is axiomatic that the probabilities of mutually exclusive and exhaustive events should sum to a probability of 1, or equivalently 100%. When our participants estimated the probabilities of the four cases in the partition of disjunctions, *A or B or both*, most of them made estimates of each of its four cases, *A and B*, *A and not-B*, *Not-A and B*, and *Not-A and not-B*, that summed to well over 100%: The overall means summed to 191% (for similar errors, see Khemlani et al., 2015). Our participants were sensitive to probabilities: They were biased against inferring conclusions that they judged to be improbable, and yet, from their own estimates of probabilities, they often inferred conclusions that were less probable than the premises, contrary to probabilistic logic.

If these errors do not strike at the center of the probabilistic theory, they nonetheless refute it. And they accord with claims attributable to other proponents of the turn to probabilities: "Bayesian brains need not represent or calculate probabilities at all and are, indeed poorly adapted to do so. . . . [W]ith finite samples, [the brain] systematically generates classic probabilistic reasoning errors" (Sanborn & Chater, 2016, p. 883). The program implementing the model theory of probabilities has a mechanism that predicts these fallacies (Khemlani et al., 2015), which we describe below.

The Model Theory Leads to Logical Absurdities

A typical dog barks, has fur, wags its tail, and has many other such characteristics. So, when you hear about a dog called Fido, you can infer any of these properties by default. If you then learn that Fido is mute, bald, and tail-less, you do not infer that he is not a dog. You merely withdraw your default inferences about his doggyhood. It is similar with the default inferences of possibilities from disjunctions. When one default possibility is denied, you withdraw it, but not the disjunction. Denial transforms a conjunction of default possibilities into a disjunction, and its refutation takes a denial of all of them. That is why conjunctions in logic should not be confused with conjunctions of defaults. You should no more use orthodox logic to assess the model theory's inferences than you should rely on the laws of cricket to adjudicate a game of baseball.

When Oaksford and colleagues argue that the model theory leads to at least six logically absurdities, they may have overlooked this point, because much of their critique relies on orthodox logic. The absurdities do not arise in the model theory or its implementation in the mSentential program, which is described in Khemlani et al. (2017) and its source code is downloadable from http://mentalmodels.princeton.edu/models/. We summarize the critics' six alleged absurdities in their own words, and rebut them:

1. "*A or B* and *it is possible that A or B* must share the same mental representation." Not so, because the models of the latter allow that *not-A and not-B* is possible, whereas models of the former do not.

2. "If *A* and *B* are jointly contingent, just about every ordinary disjunction must be true." Not so, because their joint contingency allows *not-A and not-B* as an outcome, but the ordinary disjunction of *A or B* does not.

3. "An exception [to the previous claim] is the classical tautology A or ¬A, which is false because A & ¬A is not possible, not logically and not epistemically." Not so. The interpretation of *A or ¬ A,* where "¬" symbolizes negation, yields the conjunction of two possibilities: the disjunction's first clause and the negation of its second clause, and the negation of its first clause and its second clause. The conjunction of these two possibilities is bound to be true, that is, it is a tautology:

   A

   ¬ A

   The conjunction of *A and ¬ A* is not possible, and so it yields the null model akin to the empty set, which is a subset of all sets. It has no effect on interpretation.

4. "The fundamental inference of *or*-introduction (*A*, therefore, *A or B*), which is p-valid, is rendered invalid [in the model theory]." Not so. But, this point is subtler than the previous ones. The mental model of the premise is:

   A

   and the mental models of the conclusion refer to three possibilities:

   A

   B

   A B

   The premise matches the first of these possibilities, but provides no support for the possibility of *B*, and so, as the program shows, the premise yields a valid inference:

   Therefore, it is possible that *A or B or both*.

5. "If *or*-introduction is invalid, then *A* and ¬ (*A or B*) are consistent." Not so. The model of the negated disjunction is: *not-A and not-B*, which contradicts the model of the first assertion, *A*.

6. "People can never commit the disjunction fallacy of judging that Pr(A) > Pr(A or B)." That is, they should never infer that the probability of *A* is greater than the probability of *A or B or both*. Not so. The model theory's program for probabilistic reasoning, which we mentioned earlier, implements a primitive mechanism for inferring the probability of a disjunction from estimates of the probabilities of *A* and of *B*: it takes a rough average of the two probabilities. People who have not mastered the probability calculus err in exactly this way (Khemlani et al., 2015), and so it is common for their estimates for the probability of *A* to be greater than those for the probability of *A or B or both*.

The Algorithm for Mental Models Needs a Normative Theory

Oaksford and colleagues make insightful claims about norms and levels of theorizing. As they say, theories of reasoning do need to describe both what the mind computes (a theory at the computational level) and how the mind computes it (a theory at the algorithmic level). The model theory, they claim, needs a normative theory at the computational level. In fact, such a theory has existed for a long time: reasoners drawing their own conclusions should aim to maintain semantic information, to simplify, and to reach a new proposition that is not explicit in the premises (see, e.g., Johnson-Laird, 1983, p. 40, and the section "A theory at the computational level," in Chapter 2 of Johnson-Laird & Byrne, 1991). The maintenance of semantic information guarantees that an inference is valid in that its conclusion is true in every case in which its premises are true (Jeffrey, 1981, Chapter 1). Some psychologists suppose that this definition of validity applies only to classical logic. But, in fact, all you need to know to evaluate inferences is when assertions are true. The model theory gives such an account of the meanings of compound assertions (in terms of exhaustive conjunctions of possibilities), and so the normative account above applies to it too.

Oaksford and his colleagues demand that a normative theory be complete, decidable, and, as Oaksford and Chater (1991) argued, computationally tractable. But, reasoners make inferences in domains that are provably incomplete, undecidable, and intractable (see Jeffrey, 1981, Chapters 6 and 7; Ragni, 2003; Ragni & Knauff, 2013). Sentential reasoning, the focus of the studies under discussion (Hinterecker et al., 2016), is intractable. Its normative account cannot determine the status of every inference in a tractable amount of time.

The human system for reasoning about possibilities, Oaksford et al. argue, should be consistent, because it could not otherwise be acquired or have evolved. Yet, inconsistent inferential systems have evolved (see Shafir, 1994), and no proofs exist for the consistency of most cognitive theories, including the probabilistic theory of how people select evidence to test hypotheses (Oaksford & Chater, 2007, Chapter 6). In contrast, the model theory predicts that inferences about possibilities should be systematically inconsistent. Individuals should tend to accept inferences, such as:

> It is possible that Ann is married and it is possible that Ben is single.
>
> Therefore, it is possible that Ann is married and that Ben is single.

But, they could realize that the inference is invalid, especially with clues from its content, for example, when the second clause is: *Ann is single*. The difference depends on whether they rely on their intuitions alone (implemented in mSentential's system 1) or on deliberations (implemented in its normative system 2). Many such inconsistencies in inferences are robust (Khemlani & Johnson-Laird, 2017). To the best of our knowledge, the model theory is unique in predicting them.

Conclusions

Is it possible that the model theory is wrong? Of course. But, it yields a better fit than the probabilistic theory for the data on syllogistic reasoning (Khemlani & Johnson-Laird, 2013) and for the data on how people select potential evidence to test hypotheses (Ragni, Kola, & Johnson-Laird, in press). It also fits the data on inferring possibilities from disjunctions that make no explicit reference to them (Khemlani et al., 2017). Oaksford and colleagues make many cogent points, but their criticisms of the model theory seem to call for changes in its exposition rather than in its principles. They view the probabilistic conditional as central to their approach, and that may be why, unlike the model theory, they have no algorithm for many sorts of reasoning, such as spatial, temporal, and causal inferences. It would be instructive to carry out an experiment to find out which approach—probabilistic or model-based—makes more accurate predictions about modal inferences. But, it cannot be done, because probabilistic logic has no theory of the domain. Its authors hint at a pragmatic account instead of one based on meanings. But, even if such an approach accommodated counterfactual possibilities, it seems likely to make much the same assumptions about possibilities as the model theory. Possibilities can underpin probabilities (see Khemlani et al., 2015), but probabilities cannot underpin possibilities. They cannot distinguish between certainty and necessity, no more than they can distinguish between *mud causes rain* and *rain causes mud* (Pearl, 2009) or *rain causes floods* and *rain allows floods* (Frosch & Johnson-Laird, 2011).

References

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 17075–17080. http://dx.doi.org/10.1073/pnas.1410583111

Frosch, C. A., & Johnson-Laird, P. N. (2011). Is everyday causation deterministic or probabilistic? *Acta Psychologica, 137,* 280–291. http://dx.doi.org/10.1016/j.actpsy.2011.01.015

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1606–1620. http://dx.doi.org/10.1037/xlm0000255

Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd ed.). New York, NY: McGraw-Hill.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

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

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

Khemlani, S. 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. 663–668). Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument and Computation, 4,* 4–20. http://dx.doi.org/10.1080/19462166.2012.674060

Khemlani, S. S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and Machines, 27,* 11–35. http://dx.doi.org/10.1007/s11023-017-9421-x

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

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6,* 1–38. http://dx.doi.org/10.1111/j.1468-0017.1991.tb00173.x

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198524496.001.0001

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511803161

Ragni, M. (2003). An arrangement calculus, its complexity and algorithmic properties. In A. Günter, R. Kruse, & B. Neumann (Eds.), *Advances in artificial intelligence: 26th Annual German Conference on AI* (p. 580–590). Berlin, Germany: Springer.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120,* 561–588. http://dx.doi.org/10.1037/a0032460

Ragni, M., Kola, I., & Johnson-Laird, P. N. (in press). On selecting evidence to test hypotheses. *Psychological Bulletin*.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences, 20,* 883–893. http://dx.doi.org/10.1016/j.tics.2016.10.003

Shafir, S. (1994). Intransitivity of preferences in honey bees: Support for 'comparative' evaluation of foraging options. *Animal Behaviour, 48,* 55–67. http://dx.doi.org/10.1006/anbe.1994.1211

Received February 22, 2018
Revision received April 15, 2018
Accepted April 18, 2018 ■