Inaugural Issues: Essays from Eminent Scholars and Practitioners (Invitation Only) – Theoretical Contributions

POSSIBILITY STUDIES & SOCIETY

Possibilities and human reasoning

Possibility Studies & Society
1–8

DOI: 10.1177/27538699231152731
journals.sagepub.com/home/pst
SAGE

PN Johnson-Laird
Princeton University, USA
New York University, USA

Abstract

This article describes how human reasoning differs from standard logics. It tells the story of three sorts of inference, for example: the possibility of rain implies the possibility of no rain; a forecast of rain or snow implies the possibility of rain and the possibility of snow; and a forecast of frost does not imply a forecast of frost or snow or both. People accept these inferences, which each violate the semantics of standard logics. But, they are predictions from the theory that assertions refer to mental models of possibilities that each hold in default of knowledge to the contrary. The moral of the story is that inconsistencies in human reasoning, unlike those in standard logics, have only local consequences.

Keywords
Logic, mental models, possibilities, reasoning

Once upon a time—around 50 years ago—there was a consensus among cognitive scientists that human reasoning depends on a standard logic. Reasoning itself is the basis of rationality and intelligence; individuals differ in ability, and inferences differ in difficulty. Some theorists still maintain the hypothesis of a natural logic (e.g. O'Brien, 2021). It is implicit in Aristotle and his followers. And, in the 19th century, Boole (1854) argued that the laws of logic *are* the laws of thought. Standard logics, by definition, embody the semantics of his calculus, and so they include:

- sentential logic in which the meanings of *if, and*, and *or*, are defined as true or false depending only on the truth or falsity of the clauses they connect,
- predicate logic, which includes sentential logic, and adds rules for *all* and *some*,
- modal logics for *possible* and *necessary*, which each include sentential or predicate logic,
- logic for the elementary arithmetic of natural numbers (0, 1, 2, . . .), which includes predicate logic.

Hence, standard logics are comprehensive. Piaget (1972), the great student of children's intellectual development, also argued that sentential logic underpins human reasoning.

There were some anomalies, but we all pushed them out of the logical paradise, and tried to pin down the standard logic underlying human reasoning. We knew that it couldn't tell us what conclusion to draw from a set of premises, verbal or visual, because infinitely many follow validly from any premises. Most of them are ludicrous, such as a conjunction of a premise with itself five times. Standard logics do

Corresponding author:
PN Johnson-Laird, Department of Psychology, Princeton University, Princeton, NJ 08544, USA.
Email: phil@princeton.edu

not legislate about ludicrosity or rationality. They capture valid inferences instead:

> An inference is *valid* if its conclusion is true in every case in which its premises are true (Jeffrey, 1981, p. 1).

So, a valid inference has no counterexamples in which its premises are true but its conclusion is false. This concept seems sensible, but it too yields an anomaly.

No standard logic has any way to allow the withdrawal of a valid conclusion. Suppose you learn:

Viv was bitten by a black mamba,

and you know that if a person has such a bite, the person dies. You infer:

So, Viv will die.

You then learn that Viv had the antivenom for the snake, so you withdraw this conclusion. But, no matter what premises you add, they cannot undo the validity of its inference. To force its withdrawal, suppose you add the premise: Viv will not die. You have created a set of inconsistent premises: they cannot all be true. Alas, an inconsistency yields a valid inference of any arbitrary conclusion whatsoever, including the one you want to withdraw. Bertrand Russell assured a critic that this principle was true. "In which case," the critic said, "prove that *one plus one equals one* implies that I am the Pope." Russell replied: "You are one and the Pope is one; one plus one equals one; and so you and the Pope are one." Very witty, but as Russell knew, the real reason that any inference in a standard logic follows from an inconsistency is that the inference cannot have a counterexample: there can be no situation in which its premises are all true (and the conclusion is false). Likewise, if a logic itself is inconsistent, it too has the same catastrophic consequences, spreading a contagion of arbitrary but valid conclusions. Russell (1902) wrote to Frege, a founder of predicate logic, to point out such an inconsistency in another standard logic that that the latter had devised. Frege replied: Arithmetic is tottering.

Artificial intelligencers have long understood that standard logics do not permit the retraction of valid inferences, and so they devised so-called "nonmonotonic" systems of reasoning that allow tentative conclusions to be drawn and to be withdrawn (Marek & Truszynski, 2013). Nevertheless, as I mentioned, none of the anomalies led us psychologists to depart from our logical paradise. This article is a reaction to our stubborn inertia. It will tell you a story of three sorts of inference, introduce you to a tree of knowledge (sans apple), and lead you, I hope, to a happy ending.

The first inference: From one possibility to another

The first inference is simple and it has a precedent in Aristotle (*De Interpretatione*, 21b34-6, Barnes, 1984). Someone tells you:

I. It may rain.

If that's true, you think:

Well, it may not rain.

People accept such inferences (Ragni & Johnson-Laird, 2021), but they are invalid in all standard modal logics. There is a countable infinity of these logics, which all assume that a fact implies its own possibility. Suppose that it is raining; it follows validly in these logics that it may rain, which if we add the first inference in our story implies that it may not rain. Yet, the fact that it *is* raining hardly implies that it may not rain. Hence, the first inference is incompatible with standard modal logics.

There are ramifications. If you decide that it is false that it may not rain, then it is *certain* to rain. Certainty, however, is not part of standard modal logics. Indeed, possibility and certainty suggest a scale of this sort: *impossible*, *barely possible*, *possible*, *highly possible*, *nearly certain*, and *certain*. It is similar to a scale of subjective probabilities, or even identical to it (Lassiter, 2017). It depends on knowledge, and so by definition it concerns "epistemic" interpretations of possibility. These possibilities are the likely source of probabilities, which in numerate cultures yield numerical estimates (Khemlani et al., 2015). One school of thought proposed that the basis of reasoning was, not logic, but the probability calculus (e.g. Oaksford & Chater, 2007). A problem for this hypothesis is that naive individuals do not know how to compute the probability of a conjunction, and tend to make a rough average of the probabilities of its constituents. So, their estimates of an exhaustive set of probabilities tend to sum to well over 100% (Khemlani et al., 2015).

Possibilities have other interpretations (Johnson-Laird & Ragni, 2019). One such is the "alethic" interpretation, referring to relations between ideas, such as one between premises and conclusions, for example:

> It follows, possibly, that it will rain.
> It follows, necessarily, that it will rain.

A sign that this interpretation differs from an epistemic one is the compatibility of contrasting claims of the two sorts in one and the same sentence:

> It follows necessarily that it may rain.
> One other interpretation of possibilities is common in everyday life, a "deontic" one

concerning the tree of knowledge about what is obligatory, allowed, and forbidden (Bucciarelli & Johnson-Laird, 2019). Assertions can *create* these deontic states, as when a person in authority asserts one of the following:

> It is necessary for you to leave—it's obligatory.
> It is possible for you to leave—it's permissible.
> It is not possible for you to leave—it's forbidden.

The infinitival complement, "for you to leave," has no tense, because these assertions do not refer to a particular act of your leaving. Another phenomenon suggesting that possibilities underlie probabilities rather than the converse is that probabilities cannot capture the meaning of deontic possibilities. A speaker cannot create, say, an obligation using only a statement of probability. "The probability that you leave is 100%" does not oblige you to leave—indeed, you may not leave even if it is obligatory: people do not always obey orders. Now that we have reviewed the three main interpretations of possibility in daily life, we can consider the next inference in our story.

The second inference: From alternatives to possibilities

When Richard Feynman, the physicist, was a graduate student at Princeton, the Dean's wife asked him, "Would you like milk or lemon in your tea?" Feynman, ignorant of tea, said: "Both." As tea-drinkers know, lemon juice curdles milk, and so the Dean's wife was astonished and said: "Surely you're joking, Mr. Feynman!" After he had won the Nobel prize, her reply became the title of one of his popular books. His reasoning about tea is an instance of the "paradox" of free choice permission (e.g. Kamp, 1974). It is invalid in standard logics: a conjunction (*and*) cannot follow from a disjunction (*or*). In fact, the paradox is not limited to permissions. The second inference in our story shows that it also occurs with epistemic possibilities:

> II. Feynman put milk or lemon, or both, in his tea.
> Therefore, he may have put milk and lemon in his tea

In standard logics, the disjunctive premise is true if he didn't put milk in his tea as long as he put lemon in it, and so the inference that he may have put both in his tea is invalid in any standard logic.

Defenders of standard logics have invoked pragmatics to justify such inferences. The philosopher Grice (1989) argued that the conventions of discourse enable speakers to convey more than the literal meanings of their assertions. So, a speaker who asserts the premise of our second inference creates an *implicature* that the speaker does not know what Feynman put in his tea—milk, lemon, both of them. Otherwise, she would have followed the cooperative convention of discourse, and said what it was. Unlike inferences in a standard logic, implicatures can be cancelled without contradiction, for example:

> Feynman put milk or lemon, or both, in his tea, but I'm not going to tell you which.

So, implicatures can be withdrawn. They can act like tentative premises to allow inferences that otherwise would be impossible. But, what they cannot do is to justify the withdrawal of a valid conclusion.

The interpretation of conjunctions and disjunctions in standard logics depends solely on the truth or falsity of their clauses. That is the central principle of Boolean semantics. The disjunctive premise of our second inference is true provided that at least one of its alternatives is true; otherwise, it is false. But, when you understand this premise, you have more in mind than truth values. You can envisage the situation and its alternatives. This fact leads us to the theory of mental models—the "model" theory, for short.

Mental models have a long history, reaching back to 19th century physics, and even perhaps in the dictum: "Without an image, thinking is impossible" (Aristotle, *On memory*, 450a1, see Barnes, 1984). The first theory of the role of mental models in human reasoning was within standard logics (Johnson-Laird, 1983), but the current theory is no longer consistent with them. It still assumes that mental models are *iconic* representations of the world, that is, insofar as possible their structure mirrors the structure of the world. Visual perception can construct three-dimensional models, and so can verbal comprehension, albeit of a more schematic sort. Hence, models underlie images, and models of abstract entities, such as negation, ownership, and intention, have no images that can capture their contents (Johnson-Laird, 1983, p. 423 et seq.). So, mental processes represent the meaning of the disjunctive premise in the second inference, and then use this meaning to construct models of the epistemic possibilities (e.g. Byrne & Johnson-Laird, 2020; Khemlani et al., 2018). The mSentential program, which implements the theory in Common Lisp, is at https://www.modeltheory.org/models/. Rather than trying to depict actual mental models, it uses words to refer to the three models of what Feynman might have put into his tea:

milk
lemon
milk lemon

These models represent a conjunction of three epistemic possibilities, which each hold unless there is knowledge to the contrary. Because the models are in a conjunction, they yield each of these conclusions:

> He may have put milk into his tea.
> He may have put lemon into his tea.
> He may have put milk and lemon into his tea.

People make such inferences (Hinterecker et al., 2016). Each of them follows as an alethic necessity, which the model theory defines as follows:

> An inference is *necessary* if its premises refer to one or more possibilities, and its conclusion refers only to at least one of these possibilities.

Because possibilities hold only in default of knowledge to the contrary, such knowledge can lead to the withdrawal of conclusions that are otherwise necessary given the premises. The definition of necessary inferences leads us to the third and final inference in our story.

The third inference: From a categorical premise to a disjunctive conclusion

The third sort of inference is the rejection of conclusions, such as the following one:

III. There's a chardonnay in the fridge.
Therefore, there's a chardonnay in the fridge or there's a beer, or both.

Granted the meaning of "or" in standard logics: the inference is valid because it has no counterexample in which the premise is true but the disjunction is false. However, according to the model theory, it is not a necessary inference, because the premise does not refer to the alternative possibility—that there's a beer in the fridge—to which the conclusion refers. So, it is at best only a possible inference.

Psychologists who proposed that human reasoning depends on a natural logic took pains to avoid a direct formal rule of inference for the third inference (e.g. O'Brien, 2021; Rips, 1994). Likewise, an experimental test showed that most participants rejected the inference, except in one case that the model theory predicts (Orenes & Johnson-Laird, 2012). The one exception is of this sort:

There is a chardonnay in the fridge.
Therefore, there is a chardonnay in the fridge or a white wine.

It is acceptable, because its premise implies both possibilities to which the conclusion refers:

chardonnay is a white wine.

Defenders of standard logics can provide extra-logical reasons for why people reject the conclusion in the third inference, for example, it throws information away (e.g. Johnson-Laird, 1983, p. 38). That is, you know more when you know that a chardonnay is in the fridge, than when you know only that a chardonnay or a beer is in the fridge. Such defenses, however, do not refute the validity of the inference in standard logics. This point leads us to the major difference between human reasoning and standard logics.

A set of assertions is inconsistent when they cannot all be true at the same time. People detect such inconsistencies from the impossibility of constructing a single model of all of the assertions (Johnson-Laird et al., 2000; Legrenzi et al., 2003). As I explained earlier, inconsistencies in standard logics are catastrophic: any arbitrary conclusion whatsoever follows validly from them. It is easy to assert inconsistencies in natural language, such as the well-known "liar" paradox:

This assertion is false.

Given that the assertion refers to itself: if it is true then it follows that is false; and if it is false then it follows that it is true. Wittgenstein lectured on the foundations of mathematics in Cambridge in 1939, and Turing, the main founder of the theory of computability, was a regular attender. They were both skeptical about the catastrophic consequences of inconsistencies.

Wittgenstein proposed that one should never draw conclusions from them. Turing commented that the harm comes only from applications of such conclusions—a bridge could collapse, for instance (Wittgenstein, 1989, p. 214). A distinguished physicist, Wilczek (2002, p. 159) was also skeptical. Contradictions could be fruitful in science, and he judged that they were not "an irremediable catastrophe." That, too, is true for human reasoning in general.

The model theory explains why. When human reasoners are presented with this set of assertions, for example:

If someone pulled the trigger, then the gun fired.
Someone pulled the trigger. But the gun did not fire.

they notice the inconsistency. The first two assertions yield a single model representing that the gun fired. The final assertion holds in a model in which the gun does not fire. Their conjunction is inconsistent, and so it yields the *null* model, which is bound to be false. If it is conjoined with any other model, the result is also the null model. You can infer from the null model that something is wrong with the premises that led to it, but that is all. So, reasoners can withdraw conclusions when, as in the inference above, they run into a brick wall of a contradictory fact. But, unlike the nonmonotonic systems in artificial intelligence, which I mentioned earlier, human reasoners seek an explanation to resolve such an inconsistency (Johnson-Laird et al., 2004). A corollary is that inconsistencies are not a general catastrophe, spreading a contagion of conclusions. The proofs of such arbitrary conclusions in formal logics depend on the third sort of inference in our story. But, they are not necessary inferences in the model theory. So, inconsistencies are local. This phenomenon was perhaps an intimation in the minds of the skeptics, Wittgenstein, Turing, and Wilczek.

The morals of the story

We are coming to the end of our story. The three inferences show that human reasoners infer that when an event is possible its non-occurrence is too, that possibilities follow from a disjunction of categorical alternatives, and that premises do not imply conclusions referring to additional possibilities. You should bear in mind that there are many other relevant results—the story I have told you illustrates an argument rather than clinches it. However, the impact of all the results propels human reasoning from the paradise of a standard logic. In truth, it is no Edenic garden. It is at such a height of abstraction that the air is thin: reasoning is a formal manipulation of abstract symbols cut off from their meanings. These meanings concern the truth or falsity of formulas, right down to their elements, which are true or false, or else satisfied or not, by sets of entities. This separation between the formal system and the semantic system keeps inconsistencies, such as the liar paradox, at bay. But, inconsistencies are part of everyday reasoning.

The three sorts of inference in the story have illustrated that our reasoning is far from formal. It is semantic through and through, and relies on knowledge. It depends on semantic algorithms that unlike, say, machines for computing formal algorithms, have perceptual, motor, and internal systems. They can determine the truth or falsity of assertions about our world, our actions, and our states of mind and body. These systems are limited and fallible, but they work with much richer possibilities than those in modal logics. When the facts corroborate one possibility in a conjunction of them, it becomes a truth, and the other possibility becomes a counterfactual one (Byrne, 2005). Individuals spontaneously consider whether such a counterfactual is true or false, and this evaluation affects their verification of the assertion referring to the original possibilities (Johnson-Laird et al., 2023). For instance, a speaker tells you:

> There is a chardonnay in the fridge or a hippopotamus, or both.

You find the chardonnay, but you know that there couldn't have been a hippo there—the counterfactual possibility is false, and so you accept this evaluation of the disjunction:

> It's true but it could have been false.

No standard logic allows such truth values.

Our system of reasoning allows us to create semantic paradoxes and to be inconsistent—if Frege was, who amongst us can be wholly consistent? And, even our reasoning itself yields inconsistencies between our intuitions and our deliberations—a chapter that I had to omit from the story (see, e.g. Khemlani & Johnson-Laird, 2017; for three other missing chapters, see Khemlani & Johnson-Laird, 2022; López-Astorga et al., 2022; Rasga et al., 2022). But, the story does have a happy ending. Any inconsistency is local. It tells us that something is wrong with our premises (or our reasoning), but unlike standard logics it does not imply any conclusion whatsoever. Hence our reasoning is more robust to error than the fragile ecosystem of a formal logic. The story also ends a tradition beginning over 2000 years ago: standard logics are neither a basis nor a guide for human reasoning. They are a supreme achievement of human reasoning, but not its foundation. Our reasoning, rationality, and even perhaps intelligence, depend on our ability to envisage possibilities and to draw conclusions from them.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Barnes, J. (Ed.) (1984). *The complete works of Aristotle* (*Vol. 1*). Princeton University Press.

Boole, G. (1854). *An investigation of the laws of thought*. Macmillan.

Bucciarelli, M., & Johnson-Laird, P. N. (2019). Deontics: Meaning, reasoning, and emotion. *Materiali per una storia della cultura giuridica*, *49*, 89–112.

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

Byrne, R. M. J., & Johnson-Laird, P. N. (2020). *If* and *or*: real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology Learning Memory and Cognition*, *46*, 760–780.

Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.

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.

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

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

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*, 640–661.

Johnson-Laird, P. N., Khemlani, S., & Byrne, R. M. J. (2023). *Truth, verification, and reasoning*. Under submission.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531–532.

Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, *193*, 103950.

Kamp, H. (1974). Free choice permission. *Proceedings of the Aristotelian Society*, *74*, 57–74.

Khemlani, S. S., & Johnson-Laird, P. N. (2022). Reasoning about properties: A computational theory. *Psychological Review*, *129*, 289–312.

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science*, *42*, 1887–1924.

Khemlani, S. S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and 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.

Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2003). Models of consistency. *Psychological Science*, *14*, 131–137.

López-Astorga, M., Ragni, M., & Johnson-Laird, P. N. (2022). The probability of conditionals: A review. *Psychonomic Bulletin & Review*, *29*, 1–20.

Marek, W., & Truszynski, M. (2013). *Nonmonotonic logic*. Springer.

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

Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, *27*, 357–377.

O'Brien, D. P. (2021). Natural logic. In M. Knauff, & W Spohn. (Eds.), *The handbook of rationality* (pp. 215–224). MIT Press.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development*, *15*(1), 1–12.

Ragni, M., & Johnson-Laird, P. N. (2021). Reasoning about epistemic possibilities. *Acta Psychologica*, *208*, 103081.

Rasga, C., Quelhas, A. C., & Johnson-Laird, P. N. (2022). An explanation of *or*-deletions and other paradoxical disjunctive inferences. *Journal of Cognitive Psychology*, *34*, 1032–1051.

Rips, L. J. (1994). *The psychology of proof*. MIT Press.

Russell, B. A. W. (1902). Letter to Frege. In J van Heijenoort. (Ed.), (1967) *From Frege to Gödel* (pp. 124–125). Harvard University Press.

Wilczek, F. (2002). A piece of magic: The Dirac equation. In G Farmelo. (Ed.), *It must be beautiful: Great equations of modern science* (pp. 132–160). Granta.

Wittgenstein, L. (1989). *Wittgenstein's Lectures on the Foundations of Mathematics: Cambridge, 1939*. From the notes of R. G. Bosanquet, N. Malcom, R. Rhess, & Y. Smythies. University of Chicago.