Automating Human Inference

Sangeet Khemlani

US Naval Research Laboratory, Washington DC
sunny.khemlani@nrl.navy.mil

1 Abstract

Researchers of reasoning in computer science and psychology are estranged siblings. The tools they use to investigate patterns of inference seldom overlap, because the goals of each group differ. A computer scientist's primary goal is to efficiently engineer systems based on logical calculi. But, human reasoning systematically violates the constraints imposed by orthodox logic. One logician described the disparity between logical systems and the everyday inferences they are intended to capture as "one of the greatest scandals of human existence" [1]. And so the psychological objective is to discover patterns of reasoning in humans, both normative and fallacious, with the ultimate goal of developing theories capable of predicting human inference.

Despite their diverging purposes, computer scientists and psychologists face similar theoretical challenges: how is knowledge represented and integrated into reasoning processes? What constitutes a normative inference? Why are some inferences more difficult than others? Human reasoning is predictably irrational, but it can also be more productive, flexible, and capable than current automated reasoning systems. Indeed, as some researchers argue, certain kinds of inference – such as reasoning about defaults, non-monotonic inference, explanatory reasoning, and conditional inference – cannot be characterized without reference to how humans make them [14, 16]. A keen understanding of human reasoning therefore has both psychological and computational value.

Present day automated reasoning systems do not reason the way humans do – which may explain their success as inferential tools. The productivity of automated theorem provers (ATPs) has advanced by orders of magnitude since McCune's famous solution to the Robbins problem [13]. ATPs regularly compete against one another at international competitions to efficiently yield proofs of thousands of reasoning problems. For instance, the ATPs that entered the 2004 theorem-proving competition at the annual Conference on Automated Deduction (CADE) [18] were designed to solve two thousand eligible problems. At the same competition ten years later, the number of eligible problems grew to fifteen thousand [17]. As a result of prolonged development, ATPs now routinely serve as productive analytical tools, and they are instrumental in diverse applications, such as the verification of transportation systems, electrical circuitry, and automation systems.

But, ATPs are designed to carry out just one inferential task, i.e., they operate by deriving a valid proof of a conclusion from a given set of premises. This

design constraint allows systems to filter out invalid inferences that may corrupt further processing, but it represents a stark divergence from human thinking. Humans do not spontaneously construct logical proofs when they reason [5]. Indeed, there exists little evidence to suggest that humans make use of any kind of logical form whatsoever [3, 6], and algorithms capable of recovering the logical form of an assertion from its description in natural language remain elusive. Human inference is resistant to logical formalism for three overarching reasons: first, inferences tend to be rapid and intuitive, and they are prone to systematic errors. Reasoners are theoretically capable of correcting their errors through deliberation, but doing so demands cognitive resources. Second, reasoners carry out many sorts of inferential task. For example, they can generate their own conclusions from a set of premises [8], they can consult background knowledge to explain inconsistencies [7, 4], and they can infer probabilities of unique events [11]. Finally, humans adopt different strategies when they reason, and so an automated human reasoning system must be able to account for a variety of human abilities.

mReasoner is a novel automated reasoning system [8]. It is a computational implementation of mental model theory, which posits that when people reason, they construct small-scale mental simulations of the world [2]. Mental models are discrete representations of real, hypothetical, or imaginary possibilities. They are *iconic* in that they mirror the relationships they represent. So, when a mental model represents a set of objects, the model contains multiple tokens representing multiple objects. In this way, mental models cannot be processed through syntactic transformations the way ATPs process formulas. Instead, the theory posits that reasoners build, scan, and revise models by mapping natural language input onto simulated structures. mReasoner makes inferences the way humans do: it heuristically draws initial conclusions by analyzing the structure of mental models. In doing so, it predicts reasoners' systematic errors and explains how they overcome them [10, 15]. The system can carry out multiple inferential tasks, such as assessing whether a given conclusion is possible, necessary, or consistent with the premises [12]. Its parameters affect the size and contents of the models that the system builds, and also the propensity for the system to engage in deliberation, i.e., to search for alternative models and counterexamples. Hence, it can explain individual differences in reasoning too [9].

In sum, mReasoner is a cognitively plausible automated reasoning system. It eschews logical formalisms in favor of mental models, i.e., discrete, iconic representations of possibilities. The system serves as an analytical tool that mimics both the frailties of human reasoning, e.g., systematic errors, as well as strengths of human inference, e.g., the ability to spontaneously generate relevant conclusions. Future applications in artificial intelligence and computer science will demand automated reasoning systems that interact with human reasoners. Hence, mReasoner – and systems like it – provides a foundation for those interactions.

Acknowledgments

I am most indebted to my longtime mentor and collaborator, Phil Johnson-Laird, for his infectious enthusiasm, perseverance, and creativity. The work I've described would have been impossible without many conversations with Selmer Bringsjord, my mentor in logic, and Greg Trafton, whose abilities in computational cognitive modeling continue to inspire. Finally, I thank Paul Bello, Ruth Byrne, Monica Bucciarelli, Geoff Goodwin, Tony Harrison, Laura Hiatt, Max Lotstein, Robert Mackiewicz, Isabel Orenes, and Marco Ragni for their comments and criticisms.

References

1. Yoshua Bar-Hillel. Colloquium on the role of formal languages. *Foundations of Language*, 5:256–284, 1969.
2. PN Johnson-Laird. *How we reason*. Oxford University Press, USA, 2006.
3. PN Johnson-Laird. Against logical form. *Psychologica Belgica*, 50(3-4), 2010.
4. PN Johnson-Laird, Vittorio Girotto, and Paolo Legrenzi. Reasoning from inconsistency to consistency. *Psychological Review*, 111(3):640, 2004.
5. PN Johnson-Laird, Sangeet Khemlani, and Geoffrey Goodwin. Logic, probability, and human reasoning. *Trends in cognitive sciences*, 19(4):201–214, 2015.
6. GB Keene. *The Foundations of Rational Argument*. Edwin Mellen Press, Lampeter, Wales, 1992.
7. Sangeet Khemlani and PN Johnson-Laird. Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica*, 139(3):486–491, 2012.
8. Sangeet Khemlani and PN Johnson-Laird. The processes of inference. *Argument & Computation*, 2013.
9. Sangeet Khemlani and PN Johnson-Laird. How people differ in syllogistic reasoning. *Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, 2016.
10. Sangeet Khemlani, Max Lotstein, and PN Johnson-Laird. A mental model theory of set membership. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, 2014.
11. Sangeet Khemlani, Max Lotstein, and PN Johnson-Laird. Naive probability: Model-based estimates of unique events. *Cognitive science*, 39(6):1216–1258, 2015.
12. Sangeet Khemlani, Max Lotstein, J Gregory Trafton, and PN Johnson-Laird. Immediate inferences from quantified assertions. *The Quarterly Journal of Experimental Psychology*, 68(10):2073–2096, 2015.
13. William McCune. Solution of the Robbins problem. *Journal of Automated Reasoning*, 19(3):263–276, 1997.
14. Francis Jeffry Pelletier and Renee Elio. The case for psychologism in default and inheritance reasoning. *Synthese*, 146(1-2):7–35, 2005.
15. Marco Ragni, Sangeet Khemlani, and PN Johnson-Laird. The evaluation of the consistency of quantified assertions. *Memory & cognition*, 42(1):53–66, 2014.
16. Walter Schroyens. Logic and/in psychology: The paradoxes of material implication and psychologism in the cognitive science of human reasoning. *Cognition and conditionals: Probability and logic in human thinking, ed. M. Oaksford & N. Chater*, pages 69–84, 2010.

17. Geoff Sutcliffe. The CADE-24 automated theorem proving system competition–CASC-24. *AI Communications*, 27(4):405–416, 2014.
18. Geoff Sutcliffe and Christian Suttner. The CADE-19 ATP system competition. *AI Communications*, 17(3):103–110, 2004.