

MIT Press • Open Encyclopedia of Cognitive Science

Reasoning and Argumentation

Hugo Mercier¹

¹Institut Jean Nicod, ENS-PSL University, EHESS, CNRS

MIT Press

Published on: May 28, 2025

DOI: <https://doi.org/10.21428/e2759450.96cf440a>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/).

By a common definition, reasoning is the private mental act of weighing reasons to decide whether to accept a given conclusion. In argumentation, reasons in favor of a conclusion are exchanged between at least two people, who take turns producing and evaluating them. In spite of what seems to be an obvious link between reasoning and argumentation, cognitive science had until recently focused on reasoning, neglecting argumentation. More recently, experimental and theoretical work has renewed interest in argumentation and strengthened the ties between reasoning and argumentation. Work on reasoning and argumentation promises to shed light on the cognitive mechanisms underpinning a critical human-specific ability—exchanging reasons—and, since argumentation is more easily observable than reasoning, to facilitate the study of reasoning.

History

In the Western tradition at least, reasoning and argumentation were originally, and until the modern era, largely seen as synonymous, with Aristotle defining “deduction [as] a discourse” ([Aristotle, 350 BCE/2009](#); see also [Isocrates, 1928](#)). By contrast, an *individualist* tradition, which sees reasoning as distinct from, and more important than, argumentation, developed later, following Descartes in particular.

The perspective that conflates reasoning and argumentation has exerted some influence on modern research. In social psychology, Michael [Billig \(1996\)](#) drew attention to the importance of arguing in our social and cognitive lives. Jean Piaget’s early suggestion that “logical reasoning is an argument which we have with ourselves, and which reproduces internally the features of a real argument” ([Piaget, 1928](#), 204) has influenced work in developmental psychology ([Doise & Mugny, 1984](#); [Perret-Clermont, 1980](#)). The individualist tradition, however, proved vastly more influential (including on Piaget’s later work). Until very recently, the experimental study of reasoning was dominated by traditions—mental models ([Johnson-Laird, 1983](#)) and mental logic ([Rips, 1994](#))—which largely ignored argumentation (although for a notable exception situated outside the mainstream of work on reasoning, see [Kuhn, 1991](#)). On their side, argumentation scholars showed little interest in the cognitive underpinnings of argumentative skills (e.g., [Van Eemeren & Grootendorst, 2004](#)).

First, researchers have developed a fruitful research program extending to argumentation the Bayesian framework used to model internal reasoning (for review, see [Hahn & Oaksford, 2007](#)). Second, there has been a renewed interest in children’s ability to give and to evaluate reasons within a broadly cognitive framework. Third, the interactionist theory of reason offers a framework in which argumentation—and the social exchange of reasons more generally—is the function for which reasoning would have evolved.

Core concepts

Over the past 20 years, *dual process theories* have played an important role in the psychology of reasoning ([Evans, 2003](#); [Kahneman, 2003](#); [Sloman, 1996](#)). In these models, the mind is composed of *System 1* processes, which work unconsciously, quickly, and effortlessly, delivering broadly reliable but also

systematically flawed outputs. *System 2*, or reasoning, would have evolved to process information more slowly, systematically, and consciously, in order to fix the mistakes made by System 1 processes. For instance, according to these theories, it is System 1 that drives individuals towards intuitive but mistaken answers in tricky logical or mathematical problems such as the bat and ball (a bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? Intuitive answer: 10 cents; correct answer: 5 cents). Dual process theories focus nearly entirely on internal reasoning processes in the individual.

The *interactionist theory of reason*, by contrast, seeks to give the social exchange of reason a central place ([Mercier & Sperber, 2011, 2017](#)). Grounded in evolutionary theorizing, it criticizes dual process models for the implausible function ascribed to reasoning, seen as a silver bullet able to perform any task better than System 1 mechanisms. Instead, it suggests that reasoning is another System 1 mechanism, with a specific function: finding and evaluating reasons in a social context—for instance, when people exchange arguments over what the correct answer to the bat and ball problem is. Reasoning would perform this task well, quickly, effortlessly, and also largely unconsciously, but it would perform quite poorly when required to perform any other task.

Questions, controversies, and new developments

Does individual reasoning work as internalized argumentation?

Dual process theories suggest that reasoning helps to discard mistaken outputs from System 1 processes (for some contradictory findings, see [Bago & De Neys, 2017](#)). By contrast, the interactionist theory of reason suggests that individual reasoning mostly serves either to find arguments for preexisting beliefs, giving rise to the confirmation bias (or myside bias, see [Mercier, 2016](#)), or to estimate which of several possible choices is easiest to defend with arguments (reason-based choice, see [Shafir et al., 1993](#)).

How well do people reason together?

Early results in social psychology suggested that group performance in a variety of tasks tended to be disappointing, either because participants lost their motivation ([Karau & Williams, 1993](#)) or because the group dynamic led to worse outcomes compared to individual performance, such as polarization ([Isenberg, 1986](#)). However, studies also revealed that under some conditions, groups could outperform individuals, as those with the best ideas were able to convince other group members (a phenomenon known as truth wins, see [Laughlin, 2011](#); [Moshman & Geil, 1998](#)). Overall, it seems that group discussion improves performance when participants are motivated to exchange arguments in an open-minded but also somewhat adversarial manner—that is, when the group has common incentives but also disagrees about something ([Mercier, 2016](#)).

How does argumentation develop?

Developmental studies have shown that very young children possess at least basic argumentative abilities, allowing them to be more persuaded by strong arguments than by weak ones (e.g., [Castelain et al., 2016](#); [Koenig, 2012](#)) and to provide reasons appropriate to their audience (e.g., [Köymen et al., 2014, 2015](#)). However, much more work remains to be done to better understand how these skills are refined with age and experience [see [Cognitive Development](#)].

Do large language models argue like humans?

Recent studies have shown that discussions with a chatbot can have large and persistent effects on attitudes ([Costello et al., 2024](#); see also [Altay et al., 2021](#)) [see [Large Language Models](#)]. Discussions between chatbots also appear to emulate the positive effects of discussion between humans ([Du et al., 2023](#)). The rapidly evolving performance of large language models will warrant continued attention.

Broader connections

Work linking reasoning and argumentation could help bridge different areas of research at two levels: that of mechanisms and that of function [see [Mental Representation](#)]. Reason is a metarepresentational mechanism; it deals with representations of representations: for instance, a representation of a given reason (itself a representation) as being a good reason. As such, reason relates to other metarepresentational mechanisms such as mentalizing or ostensive communication. All of these metarepresentational mechanisms might be partially overlapping, and their study could be more integrated—while keeping in mind that each serves a different function and should thus also have specific components. Which brings us to the functional level. According to the interactionist theory of reason, attributing a social function to reason helps make sense of some of its features. The same could be true of other metarepresentational or metacognitive mechanisms [see [Metacognition](#)]. In particular, it has been suggested that some metacognitive mechanisms also serve a social communicative function ([Shea et al., 2014](#)). If many, or maybe all, metarepresentational mechanisms serve social functions—even if the functions are not exactly the same—this would further argue for their more integrated study.

Further reading

- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732. <https://doi.org/10.1037/0033-295X.114.3.704>
- Kuhn, D. (1991). *The skills of arguments*. Cambridge University Press.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74. <https://doi.org/10.1017/S0140525X10000968>
- Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, 4(3), 231–248. <https://doi.org/10.1080/135467898394148>

References

- Altay, S., Hacquin, A.-S., Chevallier, C., & Mercier, H. (2021). Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. *PsyArXiv Preprints*.
<https://doi.org/10.31234/osf.io/eb2gt>

↩
- Aristotle. (2009). *Prior analytics* (G. Striker, Trans.). Oxford University Press. (Original work published 350 BCE)

↩
- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90–109. <https://doi.org/10.1016/j.cognition.2016.10.014>

↩
- Billig, M. (1996). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.

↩
- Castelain, T., Bernard, S., Van der Henst, J.-B., & Mercier, H. (2016). The influence of power and reason on young Maya children's endorsement of testimony. *Developmental Science*, 19(6), 957–966.
<https://doi.org/10.1111/desc.12336>

↩
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814. <https://doi.org/10.1126/science.adq1814>

↩
- Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Pergamon Press.

↩
- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. *Proceedings of the 41st International Conference on Machine Learning*, (467), 11733–11763. <https://dl.acm.org/doi/10.5555/3692070.3692537>

↩
- Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454–459. <https://doi.org/10.1016/j.tics.2003.08.012>

↩

- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732. <https://doi.org/10.1037/0033-295X.114.3.704>

↩

- Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50(6), 1141–1151. <https://doi.org/10.1037/0022-3514.50.6.1141>

↩

- Isocrates. (1928). *Antidosis* (G. Norlin, Trans.). In *Isocrates: Volume II*. Harvard University Press.

↩

- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge University Press.

↩

- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>

↩

- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4), 681–706. <https://doi.org/10.1037/0022-3514.65.4.681>

↩

- Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development*, 83(3), 1051–1063. <https://doi.org/10.1111/j.1467-8624.2012.01742.x>

↩

- Köymen, B., Mammen, M., & Tomasello, M. (2015). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, 52(3), 423–429. <https://doi.org/10.1037/dev0000089>

↩

- Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision-making by preschool peers. *Cognitive Development*, 32, 74–85. <https://doi.org/10.1016/j.cogdev.2014.09.001>

↩

- Kuhn, D. (1991). *The skills of arguments*. Cambridge University Press.

↩

- Laughlin, P. R. (2011). *Group problem solving*. Princeton University Press.

↩

- Mercier, H. (2016). Confirmation (or myside) bias. In R. Pohl (Ed.), *Cognitive illusions* (2nd ed., pp. 99–114). Psychology Press.



- Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9), 689–700. <https://doi.org/10.1016/j.tics.2016.07.001>



- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74. <https://doi.org/10.1017/S0140525X10000968>



- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.



- Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, 4(3), 231–248. <https://doi.org/10.1080/135467898394148>



- Perret-Clermont, A.-N. (1980). *Social interaction and cognitive development in children*. Academic Press.



- Piaget, J. (1928). *Judgment and reasoning in the child*. Routledge and Kegan Paul.



- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. MIT Press.



- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1–2), 11–36. [https://doi.org/10.1016/0010-0277\(93\)90034-S](https://doi.org/10.1016/0010-0277(93)90034-S)



- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4), 186–193. <https://doi.org/10.1016/j.tics.2014.01.006>



- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3–22. <https://doi.org/10.1037/0033-2909.119.1.3>



- Van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press.

