Intuitive Theories of Truth

Kerem Oktar[1#], Isaac J. Handley-Miner[2#],
Max Kleiman-Weiner[3,4], Kevin J. S. Zollman[5], & Liane Young[2]

# K. Oktar and I. J. Handley-Miner contributed equally to this work.

[1]Meta Fundamental AI Research (FAIR)
[2]Department of Psychology and Neuroscience, Boston College
[3]Department of Marketing and International Business, Foster School of Business,
University of Washington
[4]Paul G. Allen School of Computer Science & Engineering, University of Washington
[5]Department of Philosophy, Carnegie Mellon University

*Correspondence: oktar.research@gmail.com (K. Oktar), isaac.handley-miner@bc.edu
(I. J. Handley-Miner)

Abstract

Cognitive science has recently begun exploring how people conceptualize and reason about truth. We offer the field a framework that can guide inquiry on intuitive theories of truth centered on three core questions: how people judge whether statements could be true, are true, and whether to assert them as true.

People regularly evaluate the 'truth' of advertisements, news reports, promises, predictions, and more. But what is 'truth'? Centuries of philosophical debate have shaped the normative question of how people *ought to* define truth, but there has been far less focus among cognitive scientists on the descriptive question of how people *in fact* conceptualize and reason about truth [1-3]. Much as people's understanding of physics has been productively cast as *intuitive theories* that practically track (but do not exactly capture) natural laws [4], we propose that people have *intuitive theories of truth* that guide everyday truth judgments. These theories enable people to evaluate truth without engaging in academic epistemology, which can cause lay truth judgments to diverge from normative expectations. For instance, it is possible for people to share the same exact evidence about a piece of news, yet disagree on whether it should be considered true (see Figure 1) [5]. Beyond theoretical interest, understanding intuitive theories of truth can help shed light on the fraught role 'truth' plays in public debate around topics like artificial intelligence (see Box 1), misinformation, and science.

Cognitive science should focus on three core questions to shed light on intuitive theories of truth: First, how does someone decide whether a statement *can* be true? Second, how does someone decide whether a statement *is in fact* true? And finally, how does someone decide whether to *assert* a statement as true? These questions carve the structure of the intuitive theory into three distinct and interacting components, each posing unresolved questions to subfields of cognitive science.

Judging what can be true

'Truth-aptness' refers to whether a statement permits truth judgment. For instance, questions do not permit truth judgments, while statements in formal logic clearly do. Research has focused on two drivers of truth-aptness judgments: objectivity and knowability [6].

When a domain is construed as subjective rather than objective (e.g., food preferences versus mathematics), people may treat statements about that domain as poor targets for truth claims, as a shared 'truth' with respect to subjective claims can be ill-defined. Morality offers a much-debated example: Some take moral statements to simply express subjective attitudes, and hence deny the possibility of moral truths; others take moral statements to be akin to scientific observations, which are either true or false [7]. These judgments can have important consequences. If moral judgments are not truth-apt, for instance, it may be a mistake to argue about the 'truth' of moral claims, such as whether abortion is immoral, or to attempt to learn about moral 'truths' from others' opinions [8].

Similarly, when people expect domains to be unknowable—to lie beyond empirical investigation or lack discoverable answers (e.g., questions about divinity)—they may be judged as poor targets for truth judgments [6]. This may occur because empirical observations are a canonical method for ascertaining truth across many domains such as science. If empirical inquiry into a statement is impossible in these domains, people may perceive the statement itself as not being truth-apt.

Considerations such as objectivity and knowability thus influence whether people think a statement can have a determinate truth value and whether it would be rational to evaluate its truth. Nevertheless, there are no principled theories of truth-aptness judgments, and scholars at the intersection of philosophy and psychology are well-positioned to address this theoretical gap.

Judging what is true

Once a statement is taken to be truth-apt, people need to decide whether it is true or not. Philosophers have proposed several normative accounts of truth that can serve as hypotheses for how people judge the truth of statements—in particular, *correspondence*, *coherence*, and

pragmatist theories of truth [9]. These accounts are a productive starting point for cognitive scientists as they translate to cognitive theories that make behavioral predictions.

Correspondence theory argues that a statement is true if it corresponds to the way the world really is. As a psychological theory of truth judgments, *correspondence* predicts that the perceived accuracy of a statement should be the primary determinant of people's truth judgments. People holding this theory should principally be interested in comparing a target statement against the world.

Coherence theory argues that a statement is true if it coheres with a prior set of statements or beliefs. Psychologically, *coherence* predicts that people should judge statements as true when they are compatible with their existing beliefs. People holding this theory should be interested in comparing a target statement against their existing beliefs.

Pragmatist theories argue that a statement is true insofar as it withstands scrutiny and reliably guides action. Psychologically, *pragmatism* predicts that people should consider statements true when they show signs of robustness (e.g., reliance on the belief leading to intended outcomes). People holding this theory should be interested in using the target statement to help guide their action and observing the consequences.

Cognitive psychologists can build on these theories to better understand the cognitive mechanisms underlying truth judgments. These mechanisms may help explain well-studied biases in truth judgments (e.g., mere repetition of a claim affecting truth judgments) [10].

Judging what to assert as true

Beyond developing one's own representation of what is true, people also need to decide whether and when to assert statements as true. This decision to assert that a statement is true or false may be driven by both epistemic and non-epistemic considerations.

On the epistemic front, assertion allows people to inform one another, and thus engender beliefs that transcend individual observations. Whether to assert a statement as true thus depends on its communicative value. For instance, generic statements that are literally untrue (e.g., 'birds lay eggs') may nevertheless be asserted because they can allow listeners to form accurate generalizations about the world [11]. Assertions may also be regulated by social norms that set epistemic criteria, such as scientists agreeing to declare a null hypothesis 'false' if $p < .05$.

On the non-epistemic front, assertion allows people to guide one another's actions. Whether to assert a statement as true thus depends on its practical consequences. For instance, even generic statements that lead to *inaccurate* generalizations may nevertheless be asserted if they lead to good actions (e.g., 'these mushrooms kill' can help a tourist avoid a mildly poisonous mushroom) [12]. Assertions may also be regulated by social norms that set non-epistemic criteria; for example, employees agree to non-disclosure agreements that prevent true assertions which can harm the company.
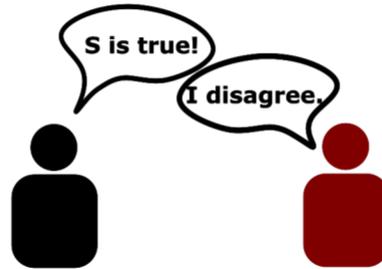
In many cases, both sets of considerations may be at play [12]. Consider recent work which found that people came to opposite truth judgments depending on whether the source of a statement was trying to be deceptive or informative [5]. Importantly, this occurred even though people knew exactly how accurate the statements were. People may want to avoid endorsing an intentionally deceptive statement as true since it might induce a false belief in others. People may also want to avoid endorsing information from a deceptive source as true for fear of lending credibility to an immoral actor.

Social, linguistic, and evolutionary psychologists should draw on research from topics like social norms, linguistic pragmatics, and social signaling to develop theories of how social considerations influence which statements people assert as true.

Concluding remarks

Society's information ecosystem rests on the bedrock of people's intuitive theories of truth. To understand why disagreements and consensus about the truth arise—and how best to resolve them—we need to understand how people decide what counts as true in the first place (see Figure 1). After all, people with the same representations of the world may disagree on what to consider true (e.g., if they adhere to different norms about standards of evidence), and people with differing representations of the world may nevertheless agree on what to consider true (e.g., if they agree that it is practically useful to endorse a claim). Understanding the intuitive theories underlying truth judgments—and how these intuitive theories differ across individuals, languages, and cultures—is an important interdisciplinary project that can help resolve contemporary debates around truth, from polarized attributions of misinformation,[I] to declining faith in institutions tasked with discovering and adjudicating truth.[II,III]

Fig 1. How intuitive theories of truth can cause disagreement

Hypothetical example of how two people's intuitive theories of truth might cause them to disagree about the truth of a statement, even given the same evidence about the event in question. The text under each heading displays the kinds of thoughts each person might have with regards to the three core questions that organize this paper. Note, however, that we do not propose that people serially consider these three questions, nor that they are even aware of their influence. These questions simply provide helpful structure for future research on intuitive theories of truth.

Box 1. Do humans and LLMs evaluate truth similarly?

Large language models (LLMs) are becoming increasingly embedded into human information ecosystems, such as web search, scientific research, education, fact-checking, and content creation. This raises the question, do humans and LLMs evaluate truth similarly?

LLMs are optimized for next-token prediction on human-generated text and fine-tuned on human ratings of qualities like helpfulness and accuracy. Given these training processes, it's unlikely that LLMs use the same mechanisms humans use to evaluate truth. For example, LLMs are not given direct perceptual access to the world. This prevents them from evaluating, firsthand, the correspondence between a statement and the state of the world. Nevertheless, it's plausible that LLM training processes could teach models to mimic humans' intuitive theories of truth, such as adherence to social norms about thresholds of evidence. Much as whether LLMs can have 'beliefs' is an open question, the extent to which they can develop coherent representations of truth is a question central to active areas of inquiry in AI research [13].

As cognitive scientists learn more about human intuitive theories of truth, they should collaborate with AI researchers to understand how LLMs evaluate truth, and how this process aligns with and diverges from the intuitive theories humans hold.

Declaration of interests

The authors declare no competing interests.

Acknowledgments

Resources

[I]https://www.pewresearch.org/journalism/2021/02/22/misinformation-and-competing-views-of-reality-abounded-throughout-2020/

[II]https://www.pewresearch.org/short-reads/2025/10/29/how-americans-trust-in-information-from-news-organizations-and-social-media-sites-has-changed-over-time/

[III]https://www.pewresearch.org/science/2023/11/14/americans-trust-in-scientists-positive-views-of-science-continue-to-decline/

References

1. Barnard, R. and Ulatowski, J. (2013) Truth, correspondence, and gender. Rev. Philos. Psychol. 4, 621–638

2. Næss, A. (1953) *An Empirical Study of the Expressions 'True', 'Perfectly Certain' and 'Extremely Probable'*, Jacob Dybwad.

3. Reuter, K. and Brun, G. (2022) Empirical studies on truth and the project of re-engineering truth. Pac. Philos. Q. 103, 493–517

4. Gerstenberg, T. and Tenenbaum, J.B. (2017) Intuitive theories. In *The Oxford Handbook of Causal Reasoning* (Waldmann, M.R., ed.), pp. 515-548, Oxford Univ. Press

5. Handley-Miner, I.J. et al. (2023) The intentions of information sources can affect what information people think qualifies as true. Sci. Rep. 13, 7718

6. Oktar, K. and Lombrozo, T. (2025) How beliefs persist amid controversy: The paths to persistence model. Psychol. Rev.

7. Heiphetz, L. and Young, L. (2017) Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. Cognition 167, 78–90

8. Oktar, K. and Lombrozo, T. (2025) How aggregated opinions shape beliefs. Nat. Rev. Psychol. 4, 81-95

9. Burgess, A.G. and Burgess, J.P. (2011) *Truth*, Princeton Univ. Press

10. Brashier, N.M. and Marsh, E.J. (2020) Judging truth. Annu. Rev. Psychol. 71, 499–515

11. Tessler, M.H. and Goodman, N.D. (2019) The language of generalization. Psychol. Rev. 126, 395–436

12. Sumers, T.R. et al. (2024) Reconciling truthfulness and relevance as epistemic and decision-theoretic utility. Psychol. Rev. 131, 194–230

13. Wang, C. et al. (2025) Survey on factuality in large language models. ACM Comput. Surv. 58, 13