MIT Press • Open Encyclopedia of Cognitive Science

Counterfactual Thinking

Felipe De Brigard[1]

[1]Duke University

MIT Press

Published on: Sep 23, 2025

DOI: https://doi.org/10.21428/e2759450.078ac821

What if I had not missed my meeting this morning? What if the axis of the earth was tilted three degrees to the left? What if Napoleon had won Waterloo? These questions exemplify the all-too-human psychological tendency to think about alternative ways in which the world could have been. This imaginative process is known as *counterfactual thinking*, because the thoughts entertained are contrary-to-fact, or *counterfactual*. Far from being confined to mere fantastic musings, counterfactual thoughts are extremely common and profoundly useful in ordinary life. We engage in counterfactual thinking when making causal judgments, for instance, when we judge that if a particular event had not occurred, then its effect would not have occurred either. Judgments of responsibility often require counterfactual thinking too, for we often must consider how an agent could have acted otherwise. And even when we reflect on our own past actions and think about how things could have been better or worse for us, we engage in counterfactual thinking.

History

The term *counterfactual* was introduced by philosopher Nelson Goodman (1947), evolving from Roderick Chisholm's (1946) more convoluted *contrary-to-fact* notion, and it was initially used to refer to conditional statements with false antecedents (e.g., "if kangaroos had no tails, they would topple over"; Lewis, 1973). Although the concept behind the term was present much earlier, featuring in the work of Newton, Leibniz, and Hume, research on the meaning of counterfactual statements remained confined to philosophy until roughly the 1980s. Then, in an influential piece in *Scientific American* in 1982, cognitive scientist Douglas Hofstadter (1982) suggested that creativity arises from our ability to mentally simulate alternative scenarios—what he called *subjunctive hypothetical replays*—and referenced the contemporaneous work of Kahneman and Tversky (1982), in which they introduced the *simulation heuristic*, as igniting empirical research into how and when people engage in counterfactual thinking.

Kahneman and Tversky's simulation heuristic suggested that people mentally alter past events in specific ways: *uphill* (adding unlikely events), *downhill* (removing unlikely elements), or *horizontal* (substituting equally likely events). Importantly, this work also indicated that people prefer downhill counterfactuals, likely because they are easier to mentally simulate. Later work with Miller (Kahneman & Miller, 1986) introduced *norm theory*, positing that people are more likely to generate counterfactuals about abnormal events than about normal ones. For example, if someone takes an unusual route home and gets into an accident, they are more likely to think "if only I had taken my usual route." This tendency—called the *exceptionality effect*—demonstrates that counterfactual thoughts often revert extraordinary events back to normative expectations.

While norm theory was developed to explain when people are likely to engage in counterfactual thinking, Roese and Olson's (1995) functional theory aimed to explain why they do so. According to this theory, counterfactual thoughts serve two primary functions: *preparative* and *affective*. Upward counterfactuals—imagining how things could have gone better—help individuals prepare for future situations, often by

eliciting regret that motivates self-improvement. For example, a soccer player might think "if I had aimed more to the left, I would have scored," which could inform future gameplay. In contrast, downward counterfactuals—imagining how things could have been worse—serve an emotional function by offering comfort and eliciting relief. A person who walks away from a car accident might think "at least I was not seriously injured" and feel better about the outcome. However, subsequent findings challenged the initial theory—for instance, not all upward counterfactuals elicit regret, and not all downward counterfactuals elicit relief (Markman & McMullen, 2003). These observations led to revisions of the functional theory, with the most recent version (Roese & Epstude, 2017) proposing two distinct pathways for counterfactual generation: a content-specific pathway, in which the counterfactual content is directly used in guiding future behavior, and a content-neutral pathway, in which emotions and motivation mediate future responses.

Core concepts

To understand contemporary work on counterfactual thinking, it is useful to think about three core concepts that have been fundamental to recent developments in the field.

Mutations

In addition to abnormal situations, researchers have identified other types of content that people are more likely to mentally mutate during counterfactual thinking. For example, individuals tend to imagine alternative outcomes more often for controllable than for uncontrollable events—a phenomenon known as the controllability effect (Girotto et al., 1991). Similarly, people are more likely to imagine undoing an action than doing one, even when the outcome would have been the same (Kahneman & Tversky, 1982). Another well-documented tendency is the temporal order effect: People prefer to mentally change events that occurred later in a sequence rather than earlier ones, even when either change would have led to the same result (Miller & Gunasegaram, 1990).

Mental models

What comes to mind when one engages in counterfactual thinking? An influential account (Byrne, 2002), which builds upon the mental-models framework (Johnson-Laird, 1986), holds that people construct two mental representations when thinking counterfactually: one of the actual event and one of the imagined alternative. For instance, when someone thinks "if kangaroos had no tails, they would topple over," they mentally represent both the real-world scenario ("kangaroos have tails and they do not topple") and the counterfactual one ("kangaroos have no tails and topple over"). Typically, people represent only the minimal set of possibilities necessary to make sense of the counterfactual statement, excluding less intuitive combinations. While more possibilities may come to mind given additional time, the initial focus on just two contrasting scenarios reflects the structure of the mental model used to interpret counterfactuals.

Sampling and mental simulations

Recent work (e.g., Phillips & Cushman, 2017; Phillips et al., 2019) suggests that, when considering counterfactual alternatives, possibilities are narrowed down by favoring those that seem both likely and valuable. This model involves three steps: First, one partitions possible outcomes. Next, one samples from this set, weighing each option's probability and utility. Finally, one selects from this smaller group to reason and decide. A key question, however, is what determines the initial partitioning of possible outcomes? Here, memory research offers insight. Neuroscience shows that imagining episodic future and counterfactual thoughts both rely on neural structures associated with episodic recollection, as all three processes likely require the recombination of details from past experiences [see Episodic Future Thinking] (De Brigard et al., 2013).

However, not all counterfactuals engage the same brain systems. Research has shown that counterfactual simulations that involve objects as opposed to people (De Brigard et al., 2015), as well as those that involve semantic facts (e.g., "what if New York was the capital of the US?") as opposed to episodic autobiographical events (Parikh et al., 2018), tend to recruit brain regions associated with semantic and conceptual processing more so than those involved in episodic memory. Likewise, recent work has shown that episodic counterfactual simulations that involve mutating an aspect of the external context rather than a different action one could have taken (Khoudary et al. 2022) also tend to recruit brain areas associated with semantic memory more so than those engaged in episodic recollection. As such, it is likely that the set of possibilities considered when generating counterfactual thoughts may depend on whether they rely more on semantic or conceptual knowledge versus autobiographical experiences.

Questions, controversies, and new developments

Beyond exploring the neural basis of counterfactual thinking, several new developments and open questions deserve attention. One is methodological. Girotto et al. (2007) found an *actor–reader asymmetry*: people imagine different alternatives depending on whether they are judging someone else's actions (as a *reader*) or reflecting on their own. Readers tend to focus on controllable actions, while actors emphasize external, uncontrollable factors. This raises concerns about whether findings from vignette-based studies apply to autobiographical counterfactuals. Additionally, most research has focused on young, Western-educated adults. Future work should examine how counterfactual thinking develops from childhood (Refetseder et al., 2013) to older age and expand to non-Western, less-educated, and nonindustrialized populations (De Brigard et al., 2017).

Broader connections

Work on counterfactual thinking interfaces several other areas of research in cognitive science. One such area is causation. Recent work on causal learning and reasoning [see Causal Learning; Causal Reasoning] strongly suggests that when individuals both learn and reason causally, they engage in the counterfactual simulation of what would have happened to the effect had a putative cause not occurred or had it

occurred differently (Gerstenberg et al., 2021). Likewise, work on moral judgment and free will [see Free Will] involve counterfactual considerations about alternative ways individuals could have acted.

Further reading

- Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology, 67*, 135–157. https://doi.org/10.1146/annurev-psych-122414-033249
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind, 55*, 289–307. https://doi.org/10.1093/mind/LV.219.289
- De Brigard, F, & Parikh, N. (2019). Episodic counterfactual thinking. *Current Directions in Psychological Science*, *28*(1), 59–66. https://doi.org/10.1177/0963721418806512
- Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. In J. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 1–79). Academic Press. https://doi.org/10.1016/bs.aesp.2017.02.001

References

- Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences, 6*(10), 426–431. https://doi.org/10.1016/S1364-6613(02)01974-5 ↩
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind, 55*, 289–307. https://doi.org/10.1093/mind/LV.219.289 ↩
- De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia, 51*(12), 2401–2414. https://doi.org/10.1016/j.neuropsychologia.2013.01.015 ↩
- De Brigard, F., Rodriguez, D. C., & Montañés, P. (2017). Exploring the experience of episodic past, future, and counterfactual thinking in younger and older adults: A study of a Colombian sample. *Consciousness and Cognition*, *51*, 258–267. https://doi.org/10.1016/j.concog.2017.04.007 ↩
- De Brigard, F., Spreng, R. N., Mitchell, J. P., & Schacter, D. L. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *NeuroImage, 109*, 12–26. https://doi.org/10.1016/j.neuroimage.2014.12.075 ↩
- Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936–975.

https://doi.org/10.1037/rev0000281

↩

- Girotto, V., Ferrante, D., Pighin, S., & Gonzalez, M. (2007). Postdecisional counterfactual thinking by actors and readers. *Psychological Science, 18*(6), 510–515. https://doi.org/10.1111/j.1467-9280.2007.01931.x

↩

- Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica, 78*(1–3), 111–133. https://doi.org/10.1016/0001-6918(91)90007-M

↩

- Goodman, N. (1947). The problem of counterfactual conditionals. *The Journal of Philosophy, 44*(5), 113–128. https://doi.org/10.2307/2019988

↩

- Hofstadter, D. R. (1982). Variations on a theme as the essence of imagination. *Scientific American.* 247(4): 20-29.

↩

- Johnson-Laird, P. (1986). *Mental models*. Harvard University Press.

↩

- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

↩

- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, E. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.015

↩

- Khoudary, M., O'Neill, K., Faul, L., Murray, S., Smallman, R., & De Brigard, F. (2022). Neural differences between internal and external episodic counterfactual thoughts. *Philosophical Transactions of the Royal Society B, 377*(1866), 20210337. https://doi.org/10.1098/rstb.2021.0337

↩

- Lewis, D. K. (1973). *Counterfactuals*. Harvard University Press.

↩

- Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review, 7*(3), 244–267. https://doi.org/10.1207/S15327957PSPR0703_04

↩

- Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology, 59*(6), 1111–1118. https://doi.org/10.1037/0022-3514.59.6.1111

  ↩
- Parikh, N., Ruzic, L., Stewart, G. W., Spreng, R. N., & De Brigard, F. (2018). What if? Neural activity underlying semantic and episodic counterfactual thinking. *NeuroImage, 178*, 332–345. https://doi.org/10.1016/j.neuroimage.2018.05.053

  ↩
- Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences, 114*(18), 4649–4654. https://doi.org/10.1073/pnas.1619717114

  ↩
- Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences, 23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

  ↩
- Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

  ↩
- Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. In J. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 1–79). Academic Press. https://doi.org/10.1016/bs.aesp.2017.02.001

  ↩
- Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 1–59). Erlbaum.

  ↩