INSIGHTS · GENERATING INSIGHT

# Using Counterfactual Thinking to Improve Interpretation

How asking "what would have happened otherwise?" sharpens the interpretation of data, covering the psychology of counterfactual thinking, its role in causal inference, methods for estimating counterfactuals, and structured techniques that help analysts avoid hindsight and outcome bias.

ModQu research article · 2,253 words · 26 references (APA 7th edition)

Every claim that something worked contains a hidden comparison. When a marketing team says a campaign increased sales by 12%, it is comparing what happened with what would have happened without the campaign. When an analyst says a price change caused customers to leave, the comparison is with a world in which the price did not change. These alternative worlds, which did not happen and cannot be observed, are called counterfactuals. Making them explicit is one of the most useful habits an analyst can develop.

This article explains what counterfactual thinking is, how it shapes ordinary judgement, why it lies at the heart of causal inference, and how it can be used deliberately to improve the interpretation of data and the evaluation of decisions.

## Counterfactual thinking in everyday judgement

People constantly imagine how things might have turned out differently. Kahneman and Tversky (1982) described the "simulation heuristic", the process by which people construct mental scenarios of alternative outcomes and use the ease of imagining them to judge events. They showed that people feel more regret about outcomes that were narrowly missed, such as missing a flight by five minutes rather than thirty, because the alternative is easier to imagine.

Roese (1997) reviewed research on counterfactual thinking and described its two main directions. Upward counterfactuals imagine how things could have been better ("if only we had launched a month earlier"). Downward counterfactuals imagine how things could have been worse ("at least we did not launch during the recession"). Upward counterfactuals tend to make people feel worse but help them prepare for the future; downward counterfactuals tend to make people feel better. Medvec et al. (1995) found a striking example: Olympic bronze medallists appeared happier than silver medallists, apparently because silver medallists compared themselves with winning gold, while bronze medallists compared themselves with not winning a medal at all.

Epstude and Roese (2008) proposed that counterfactual thinking serves a practical function: by identifying what could have led to a better outcome, it helps people learn and plan. Byrne (2005) studied the patterns in how people imagine alternatives and found that they tend to change some things more readily than others, such as actions rather than

inactions, controllable events rather than uncontrollable ones, and recent events rather than earlier ones in a sequence. These tendencies mean that spontaneous counterfactual thinking is selective. It does not consider all the ways things might have gone differently, only those that come easily to mind.

## Counterfactuals and causation

Counterfactuals are central to the definition of causation. The philosopher Lewis (1973) analysed causation in counterfactual terms: an event causes another if, had the first not occurred, the second would not have occurred. Statisticians formalised a similar idea. Rubin (1974) described causal effects as the difference between the outcome a unit would experience with a treatment and the outcome it would experience without it, the "potential outcomes" framework.

Holland (1986) pointed out the "fundamental problem of causal inference": for any single unit, only one of these potential outcomes can be observed. A customer either received the discount or did not; we can never see the same customer at the same moment in both conditions. Causal inference therefore always involves estimating an unobserved counterfactual, usually by finding comparable units or periods that did not receive the treatment. Imbens and Rubin (2015) set out the methods for doing so, and Pearl and Mackenzie (2018) gave an accessible account of how counterfactual reasoning underlies the science of cause and effect.

The practical implication for insight work is simple but often neglected. A finding about cause or impact is only as good as the counterfactual it implies. "Sales rose 12% after the campaign" says nothing about the campaign's effect unless we know what sales would have done without it. If sales were rising 10% anyway because of seasonal demand, the campaign's effect was much smaller than it appears.

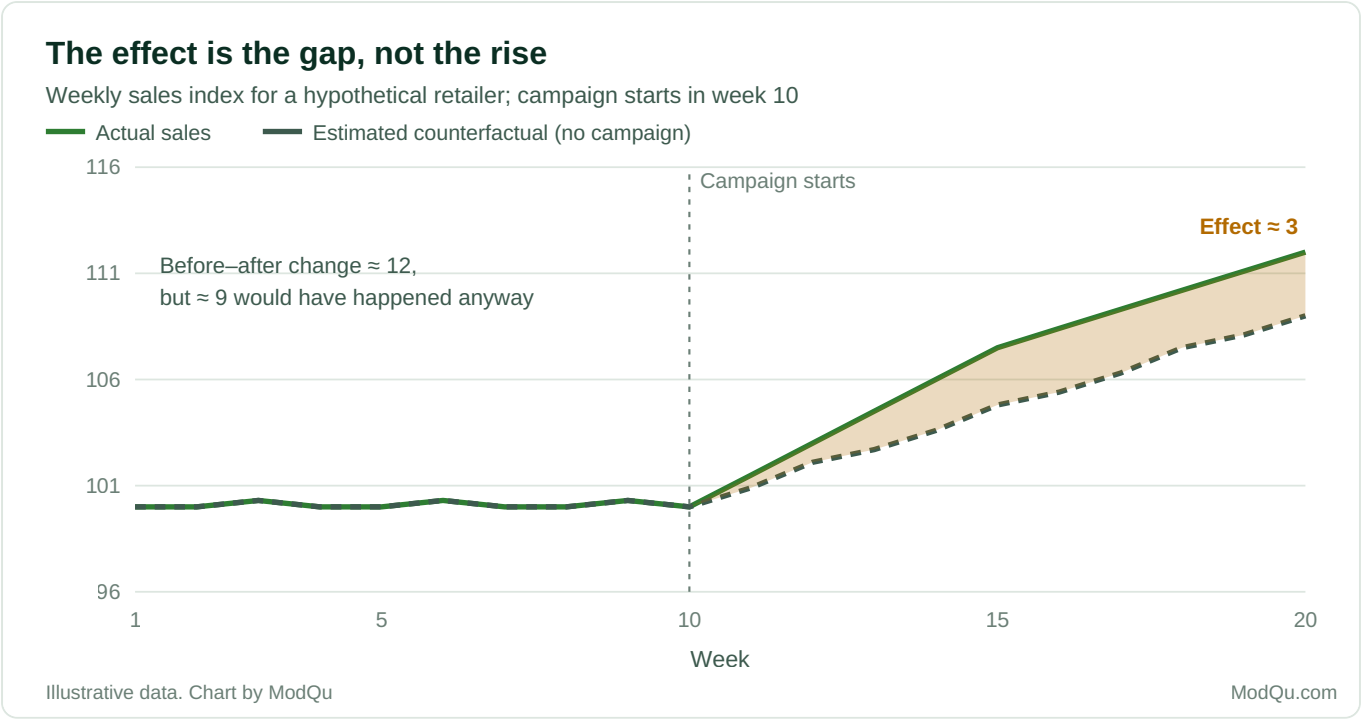



Figure 1. The effect of an intervention is the gap between what happened and what would have happened otherwise, not the change from before to after. Illustrative data, chart by ModQu.

## Estimating counterfactuals in practice

Because the counterfactual cannot be observed directly, it must be estimated. The best method is a randomised experiment, in which chance decides who receives the treatment, so that the control group provides a direct estimate of what would have happened to the treated group. Where experiments are not possible, several methods construct a counterfactual from other data.

**Comparison groups.** Card and Krueger (1994) estimated the effect of a minimum wage increase in New Jersey by comparing changes in fast-food employment there with changes in neighbouring Pennsylvania, where the minimum wage did not change. Pennsylvania served as the counterfactual for what would have happened in New Jersey. This "difference-in-differences" logic is widely used in business: compare the change in regions, stores or customer groups that received an intervention with the change in similar ones that did not.

**Synthetic controls.** When no single comparison unit is a good match, a counterfactual can be built from a weighted combination of several. Abadie et al. (2010) developed this synthetic control method and used it to estimate the effect of a tobacco control programme in California by comparing the state with a combination of other states that matched its pre-programme trends.

**Forecast-based counterfactuals.** Brodersen et al. (2015) described a method that uses time-series models to predict what an outcome would have been after an intervention, based on its own history and on related series that were not affected. The difference between the prediction and the actual outcome estimates the effect, with an indication of uncertainty.

Each method relies on assumptions, chiefly that the comparison represents what would have happened to the treated units. Making the counterfactual explicit makes these assumptions visible, so that they can be examined and challenged.

## Common counterfactual traps in business reporting

Several patterns in routine business data produce apparent effects that disappear once the counterfactual is considered.

**Seasonality and trend.** Many measures rise and fall with the calendar or follow an underlying trend. Comparing a period after an intervention with the period before it attributes the seasonal or trend change to the intervention. Comparing with the same period in previous years, or with a comparison group exposed to the same season, provides a better counterfactual.

**Regression to the mean.** When an intervention is targeted at units that performed unusually badly, such as the worst stores or the least satisfied customers, those units will tend to improve on the next measurement even without any intervention, simply because extreme values are partly due to chance and chance does not repeat itself reliably. Barnett et al. (2005) explained how regression to the mean arises and how it can be mistaken for a treatment effect. The counterfactual question, "how would these units have changed if we had done nothing?", is answered by comparing them with similar low performers that were not treated.

**Selection.** Customers who choose to take part in a loyalty programme, attend a webinar or use a new feature differ from those who do not. They are often more engaged to begin with. Comparing their later behaviour with that of non-participants confuses the effect of the programme with the effect of being the kind of customer who joins. The relevant counterfactual is what the participants themselves would have done without the programme, which a simple comparison with non-participants does not provide.

**Displacement.** An intervention may increase one measure by taking from another. A promotion may raise sales of the promoted product while reducing sales of other products or of the same product in later weeks. The counterfactual for total sales, not only promoted sales, shows whether anything was gained overall.

**Attribution to the last change.** When several things change at once, such as a new website, a price change and a competitor's exit, the most recent or most visible change tends to receive the credit. Asking what would have happened with each change alone, even approximately, guards against this.

In each case, the habit of stating the counterfactual explicitly, in a sentence, exposes the assumption on which the claimed effect depends. "Without the programme, retention among these customers would have stayed at last year's level" is a claim that can be examined and tested. "Retention rose after the programme" is not.

## Hindsight and outcome bias

Counterfactual thinking can also correct two common errors in interpreting past events.

Fischhoff (1975) showed that once people know how an event turned out, they overestimate how predictable it was, a tendency known as hindsight bias. Roese and Vohs (2012) reviewed research on hindsight bias and argued that it arises partly because knowing the outcome makes it hard to imagine how things could have turned out differently. Hindsight bias undermines learning: if a failure seems to have been obvious all along, people draw simple lessons and miss the real uncertainty that decision makers faced.

Baron and Hershey (1988) showed that people judge the quality of decisions by their outcomes, rating the same decision as better when it happened to turn out well. This outcome bias leads organisations to reward lucky decisions and punish unlucky ones, regardless of how well they were made.

Counterfactual thinking helps with both. Asking "how else could this have turned out, given what was known at the time?" restores the uncertainty that hindsight removes. Asking "would this have been a good decision if the outcome had been different?" separates the quality of the decision from the result.

## Structured techniques

Several structured techniques use counterfactual thinking deliberately.

### Five counterfactual questions

Questions that test interpretations of data and decisions

| Question | Purpose |
|---|---|
| **What would have happened without it?** | Estimate the true effect, not the before–after change |
| **What else could explain this?** | Consider rival causes |
| **What if the result had been the opposite?** | Would the explanation have been just as convincing? |
| **How else could this have turned out?** | Restore the uncertainty that hindsight removes |
| **If this plan failed, why?** | Find risks before acting (the premortem) |

Diagram by ModQu ModQu.com

Figure 2. Five counterfactual questions that improve the interpretation of data and decisions. Diagram by ModQu.

**Consider the opposite.** Lord et al. (1984) found that asking people to consider whether they would have made the same judgement if the evidence had pointed the other way reduced bias in how they evaluated evidence. Koriat et al. (1980) found that asking people to list reasons why their answer might be wrong reduced overconfidence. For analysts, a useful test is to ask whether an explanation would have seemed just as convincing if the result had been the opposite. If a rise in sales would have been attributed to the campaign, and a fall would have been attributed to the competitor's price cut, neither explanation has really been tested.

**Counterfactual mind-sets.** Galinsky and Moskowitz (2000) found that people who had been prompted to think counterfactually were more likely to consider alternative possibilities in later, unrelated tasks. Kray and Galinsky (2003) found that groups primed with a counterfactual mind-set were more likely to seek information that could disconfirm their initial preference and made better group decisions as a result.

**Prospective hindsight.** Mitchell et al. (1989) found that imagining that an event had already occurred increased people's ability to generate reasons for it. Klein (2007) built on this in the "premortem": before a decision is finalised, the team imagines that it has been implemented and has failed, and each member writes down why. This counterfactual exercise surfaces risks that optimism and group pressure would otherwise suppress.

**Counterfactual history.** Tetlock and Belkin (1996) described how counterfactual thought experiments are used in the study of history and politics, and proposed criteria for judging them, including that the imagined change should be minimal and consistent with what was known. Tetlock (2005) found that experts who considered multiple possibilities, rather than relying on a single explanatory framework, made more accurate forecasts. The same discipline applies to business reviews: a credible counterfactual changes as little as possible and respects what was actually known at the time.

## Counterfactual explanations of models

Counterfactual thinking has also become important in explaining the output of predictive models. Wachter et al. (2018) proposed that automated decisions can be explained by describing the smallest change to the inputs that would have changed the decision, for example "your loan would have been approved if your income had been 5,000 higher". Such explanations are useful to customers and to analysts trying to understand what drives a model's predictions, although they describe the model's behaviour, not necessarily the causes in the real world.

## An example

Consider a hypothetical subscription business that introduced a new onboarding programme for customers in January. By June, customer retention had improved by four percentage points, and the programme was credited with the improvement.

An analyst asks what would have happened without it. A comparison with customers who joined in the same months the previous year shows that retention usually improves in the first half of the year. A difference-in-differences comparison with a region where the programme had been delayed for technical reasons suggests that the programme's effect was about one and a half percentage points, still worthwhile but much smaller than claimed.

The team also asks the "opposite" question: if retention had fallen, would they have concluded that the programme had failed? On reflection, they would probably have blamed a competitor's promotion that ran in the spring. Recognising this, they agree to judge future programmes against a counterfactual defined in advance, with a comparison group built into the rollout plan.

## Key takeaways

Every claim about cause or impact compares what happened with what would have happened otherwise. Making that counterfactual explicit is one of the most powerful ways to improve interpretation. Counterfactuals cannot be observed directly, so they must be estimated through experiments, comparison groups, synthetic controls or forecasts, each with assumptions that should be stated. Counterfactual questions also counter hindsight and outcome bias, and structured techniques such as considering the opposite and the premortem help analysts and teams test their explanations before relying on them.

## References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569–579. https://doi.org/10.1037/0022-3514.54.4.569

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics, 9*(1), 247–274. https://doi.org/10.1214/14-AOAS788

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

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84*(4), 772–793.

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*(2), 168–192. https://doi.org/10.1177/1088868308316091

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology, 36*(4), 384–409. https://doi.org/10.1006/jesp.1999.1409

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge University Press.

Klein, G. (2007). Performing a project premortem. *Harvard Business Review, 85*(9), 18–19.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes, 91*(1), 69–81. https://doi.org/10.1016/S0749-5978(02)00534-4

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*(17), 556–567. https://doi.org/10.2307/2025310

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology, 69*(4), 603–610. https://doi.org/10.1037/0022-3514.69.4.603

Mitchell, D. J., Russo, J. E., & Pennington, N. (1989). Back to the future: Temporal perspective in the explanation of events. *Journal of Behavioral Decision Making, 2*(1), 25–38. https://doi.org/10.1002/bdm.3960020103

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*(1), 133–148. https://doi.org/10.1037/0033-2909.121.1.133

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701. https://doi.org/10.1037/h0037350

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Tetlock, P. E., & Belkin, A. (Eds.). (1996). *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives*. Princeton University Press.

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887.