The psychology of counterfactual thinking
Roese, Neal J.; Morrison, Mike

Empfohlene Zitierung / Suggested Citation:

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more information see:
https://creativecommons.org/licenses/by/4.0

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-286680
The Psychology of Counterfactual Thinking

Neal J. Roese & Mike Morrison*

Abstract: »Die Psychologie kontrafaktischen Denkens«. Counterfactual thinking refers to mental constructions of alternatives to past events. In this overview of the psychological basis of counterfactual thinking, we examine how such thoughts influence emotions and carry benefits for everyday behavior. Two psychological mechanisms, contrast effects and causal inferences, can explain many of the effects of counterfactual thinking reported by psychologists. We then consider how counterfactuals, when used within expository but also fictional narratives (for example, in alternative histories), might be persuasive and entertaining.

Keywords: counterfactual thinking, causal inference effect, contrast effect.

The Psychology of Counterfactual Thinking

What might your life be like if you had made key choices differently? What if you had attended a different college, chosen a different career, married someone else? Most people ponder such possibilities at least once in awhile; sometimes they are haunted by the apparent failings such musings reveal. These sorts of thoughts are termed counterfactual, meaning that they are mental representations of alternatives to past factual events. Psychological researchers have found that counterfactual thoughts play an important role in mental life, informing decisions, shaping emotions, and placing knowledge into context.

According to the definition commonly agreed upon by psychologists, counterfactual thoughts refer to mental representations that are explicitly contrary to facts or beliefs (Byrne, 2005; Epstude & Roese, 2008; Roese, 1997). Some aspect of perceived reality is taken as the starting point, and the counterfactual embodies a juxtaposition against this reality. Typically, an event that is negative or unusual (e.g., an automobile accident) triggers the process of counterfactual thinking. The resultant counterfactual then focuses on how the event might have been different. Counterfactual thoughts often take the form of conditional statements, embracing both an antecedent (“If only Bob had kept his eyes on the road”) and a consequent (“he would have avoided the accident”). In the typical thoughts that abound in daily life, the antecedent constitutes an

* Address all communications to: Neal J. Roese / Mike Morrison, Department of Psychology, University of Illinois, 603 East Daniel St., Champaign IL 61820, USA; e-mail: roese@illinois.edu; mmorris8@cyrus.psych.uiuc.edu.
This research was supported National Institute of Mental Health grant R01-MH55578, awarded to Neal Roese.
action or decision by an individual, and the consequent describes a state of being, often framed in evaluative terms (Markman & McMullen, 2003). Counterfactuals can focus, therefore, on alternative outcomes that are better than actuality (“upward counterfactuals”) or worse than actuality (“downward counterfactuals”). Psychologists are interested in counterfactual thinking because they seem to be intimately related to emotion, social perception, and self-understanding.

The exploration of counterfactuals has been initiated independently by scholars in diverse disciplines, including philosophy, history, economics, political science, linguistics, computer science. The study of the psychological basis of counterfactual thinking began in the 1970s with studies of the basic memory properties of counterfactual versus factual inferences (e.g., Carpenter, 1973; Fillenbaum, 1974). A seminal paper by Kahneman and Tversky (1982) reframed the study of counterfactuals in terms of biased judgment and decision-making. By assessing counterfactuals within their everyday life context, such as consumer choice, monetary decisions, or career plans, Kahneman and Tversky initiated a new wave of research that has connected counterfactual thinking to a wide range of psychological and behavioral outcomes. For example, counterfactual thinking has been linked to difficulty in coping with misfortune, judgments of blame and responsibility, depression and anxiety symptoms, feelings of regret, superstitious beliefs, overconfidence regarding the predictability of the past, and expectations for future occurrences (e.g., Alicke, Buckingham, Zell, & Davis, 2008; Gilbar & Hevroni, 2007; Markman & Miller, 2006; Miller & Taylor, 1995; Roese & Maniar, 1997; Roese & Olson, 1993; Tetlock & Lebow, 2001).

The present paper provides a brief overview of the psychology of counterfactual thinking, examining the determinants and consequences of such thoughts in daily life. We then apply these ideas to shed light on how counterfactuals in the form of “alternative realities” are used in fiction, film, and other forms of narrative entertainment. As we will show, the psychological basis of counterfactual thinking illuminates their role as dramatic devices (Roese, 2005).

Some Consequences of Counterfactual Thinking

Psychological research in the 1980s and 1990s emphasized the negative consequences of counterfactual thinking. The emotion of regret is a negative feeling that hinges on a counterfactual inference, specifically the recognition that a decision, if made differently, would have resulted in a better outcome. “Upward” counterfactuals, i.e., those that focus on how past events might have been better, can produce this feeling of regret. If an individual repeatedly dwells on past failings by musing on how things might have been better, the individual is at risk for depression and anxiety disorders (Kocovski, Endler,
Rector, & Flett, 2005; Markman & Miller, 2006; Monroe, Skowronski, MacDonald, & Wood, 2005; Roese et al., in press). In other research, long-term rumination (repetitive intrusive thoughts about past events) on upward counterfactuals was found to interfere with coping with negative life events (Davis & Lehman, 1995). Another form of negative consequence was biased decision-making. After making a particular investment, for example, a decision-maker might see that he would be better off by 25% had he invested elsewhere. To the extent that this counterfactual realization produces sharply negative emotion, the decision-maker may be driven to make subsequent non-optimal decisions by re-assessing strategy and switching investments (Roese, 1999; Zeelenberg & Pieters, 1999). All else being equal, investors tend to “over-switch” (Barber & Odean, 2000), which is to say that they do not remain within a particular investment long enough to realize maximal benefit, instead incurring unnecessary transaction costs by switching to new investments. The counterfactual emotion of regret exacerbates this tendency to over-switch.

Counterfactual thinking also influences judgments of blame and responsibility (Alicke et al., 2008). Consider the case of Harry, who is attacked late at night while walking near his flat. To the extent that it is easy to imagine Harry taking an alternative route, or staying inside and not walking at all, the resulting counterfactual emphasizes Harry’s decision-making role as a causal input into the attack. Although it is utterly clear to most observers that primary blame rests on the assailant and not the victim, there is nevertheless a tendency to blame the victim more in light of a counterfactual that “undoes” the outcome by focusing on how Harry’s actions might have been different. A variety of studies have thus connected counterfactual thoughts to blame judgments in criminal and other legal decision-making contexts (Alicke et al., 2008; Catellani & Milesi, 2001; Macrae, 1992; Macrae, Milne, & Griffiths, 1993; Turley, Sanna, & Reiter, 1995).

This brief overview of the consequences of counterfactual thinking only scratches the surface. Several hundred studies have demonstrated the impact of counterfactual thinking on a wide range of judgments, decisions, emotions, and behavior. These various consequences may be understood in terms of two basic psychological mechanisms.

Two Mechanisms

Two main mechanisms have been argued to underlie the majority of psychological consequences of counterfactual thinking (Epstude & Roese, 2008; Roese, 1997; Roese & Olson, 1995).
Contrast Effects

Counterfactual thoughts may influence emotions and judgments by way of a contrast effect, which is based on the juxtaposition of reality versus what might have been. For example, winning $50 feels nice, but if one came close to winning $100 instead of $50, it does not feel quite as nice. This effect of counterfacts on emotion and satisfaction is an example of a widely observed psychological principle, that of the contrast effect. Contrast effects occur when a judgment is made more extreme via the juxtaposition of some anchor or standard (Sherif & Hovland, 1961). Contrast effects can apply to any sort of judgment, including physical properties, such as heaviness, brightness, loudness, or temperature. For example, ice cream feels especially cold immediately after sipping hot tea. A suitcase may feel especially light if one has just been moving furniture. Contrast effects also apply to subjective appraisals of value, satisfaction, and pleasure. Thus, by the same token, a factual outcome may be judged to be worse if a more desirable alternative outcome is salient, and that same outcome may be judged to be better if a less desirable alternative outcome is salient.

This counterfactual contrast effect was vividly illustrated in a study of Olympic athletes and their reactions to medal-winning performances. Medvec, Madey, and Gilovich (1995) found that Silver medalists were less satisfied than were Bronze medalists. Most observes would assume that second place confers greater joy than third place. But in several studies, these authors showed that the counterfactual that “I almost came in first” is salient to the Silver medalist (i.e., an upward counterfactual), whereas the counterfactual that “I might have come in fourth and missed getting a medal” is salient to the Bronze medalist (i.e., a downward counterfactual). These differing counterfactuals (upward vs. downward) were found to influence the degree of satisfaction felt by the athletes, as indicated by their facial displays of rejoicing immediately after the competition, and also by questionnaires in which the athletes gave ratings of their subjective pleasure.

Causal Inference Effects

Counterfactual thoughts may also imply causal inferences, which may have psychological consequences that are independent of contrast effects. By virtue of their conditional structure and implicit reference to a parallel factual statement, counterfactual propositions exemplify the logic of Mill’s method of difference (see Mill, 1872). For example, consider the counterfactual statement that “If only Ellen had bought the insurance, she would not have been in such a dire financial situation after the fire.” This alternative scenario is implicitly connected to the parallel facts that Ellen did not buy insurance and that Ellen was in a dire financial situation. If the counterfactual world in which Ellen does buy insurance is identical in all other respects to factuality, then the only thing
that can account for the difference in financial situations between the counterfactual world and the factual world is Ellen’s insurance decision. As a result, this counterfactual points to the conclusion that Ellen’s decision was a causal force in bringing about her financial situation. In current psychological theorizing, it is assumed that counterfactual thoughts do not create or evoke the causal inference, but rather that they dramatize, underscore, or illuminate a causal inference that is already thought by the individual to reasonably plausible (cf. Epstude & Roese, 2008; Spellman & Mandel, 1999).

Via their influence on causal inferences, counterfactual thoughts may influence the blaming effects discussed in the previous section, but they may also underlie overconfidence in predicting the past (one may feel certain that the outcome of a football match was predictable in hindsight because a particular causal explanation, such as the role of a star player, is salient, e.g., Roese & Maniar, 1997). By the same token, predictions of future outcomes may also be influenced by this counterfactual-induced causal inference (if a football victory was due to the actions of the star player, then one may predict future victories in games featuring that same star player).

These two mechanisms of counterfactual represent a basic, lower-level of conceptual analysis. At a higher-level, we may apply these ideas of mechanism to a broader assessment of “why” human beings generate so many counterfactual thoughts.

**Functions of Counterfactual Thinking**

Counterfactual thoughts may serve important psychological functions for the individual (Epstude & Roese, 2008; Roese, 1997). That is, counterfactual thoughts may be construed not only as negative and as sources of bias. Rather, they may also be construed as useful or adaptive for certain purposes and under certain circumstances. In a functional analysis, such global benefits may explain the genesis, pervasiveness, and situational variation of counterfactual thinking.

Two distinct functions of counterfactual thinking have been supported by a range of research. First, such thoughts may serve a preparative function; that is, they may illuminate means by which individuals can prepare for the future and, accordingly, improve their lot. Thus, if a student who failed an exam realizes that he would have passed if only he had studied more, he has identified a causally potent antecedent action that may be subsequently deployed to enhance future performance (Markman & McMullen, 2003; Roese, 1994; Roese & Olson, 1997). Second, counterfactual thoughts can serve an affective function; that is, they may be used to make oneself or another person feel better. This affective function relies not on any causal information, but rather on the contrast effect mechanism. That is, a given outcome is judged more favorably to the extent that a less desirable anchor is made salient (Roese, 1999). For
example, a person who experiences a negative event (e.g., an automobile accident) may console herself by thinking that the event might have been even worse (e.g., she might also have been seriously injured).

These two psychological functions of counterfactual thinking help to place into context a wide range of effects of counterfactual thinking that have been observed. More specifically, several key patterns of counterfactual thinking in daily life are compatible with the general idea that counterfactuals can improve performance. First, counterfactual thoughts tend to be idealistic, in that upward counterfactuals outnumber downward counterfactuals. That is, people tend to think spontaneously about how the past could have been better rather than how it could have been worse (Nasco & Marsh 1999; Summerville & Roese, 2008). Second, counterfactual thoughts are situationally reactive, in that they are more likely to appear after failure than after success (Roese & Hur 1997). Third, counterfactual thoughts are problem-focused, in that they tend to focus on fixing problems and achieving goals (Roese, Hur, & Pennington, 1999). Finally, counterfactual thoughts themselves have the effect of improving performance, at least under some circumstances (Markman, McMullen, & Elizaga, 2008; Roese, 1994).

That counterfactual thoughts help to improve performance (the preparative function) is consistent with recent brain imaging evidence. If counterfactual thinking is implicated in learning, deciding, and planning (all of which seem to involve neural pathways that come together in the orbitofrontal cortex of the brain; Kringelbach, 2005), then counterfactual thinking should also involve similar brain activation. Indeed, patients with damage to the orbitofrontal cortex show decreased skill in counterfactual thinking, and tasks that elicit counterfactual thinking involve orbitofrontal activation (Camille et al., 2004; Chandrasekhar, Capra, Moore, Noussair, & Berns, 2008; Fujiwara, Tobler, Taira, Iijima, & Tsutsui, 2008).

This psychological basis of counterfactual thinking can be applied to an understanding of counterfactual ideas and might be more or less convincing when used as an argument, or when used to bolster a narrative. We turn next to a discussion of these ideas.

**Impact of Counterfactuals in Communications and Entertainment**

The psychological basis of counterfactual thinking helps to shed light on why some sorts of counterfactuals come across as more persuasive, more compelling, and more entertaining. As discussed previously, counterfactuals can influence emotions by way of a contrast effect. Accordingly, counterfactuals may be used to make a narrative more stimulating (and perhaps more enjoyable) by way of this contrast effect mechanism. Upward counterfactuals (thinking about how the past might have been better) evoke negative emotions, whereas down-
ward counterfactuals (thinking about how the past might have been worse) evoke positive emotions, by way of juxtaposing against reality. By skillfully injecting a story with counterfactuals, a storyteller may manipulate the emotions of the audience with subtlety and aplomb.

A classic example is the film *It’s a Wonderful Life*, in which the main character is shown what his small town (and the people in it) might have been like had he never been born. The vision shown to the main character is a dark and unpleasant vision – the town that might have been is much worse than reality, and many of the people’s lives that are emptier and lonelier. By drawing out this contrast, the main character, and the audience in turn, come to appreciate his actual life to a greater extent. The film ends with a joyous appreciation of the world as it actually is. In this example, the entire plot of the film hinges on a downward counterfactual.

Counterfactuals may also be used to heighten emotion at specific points in a narrative. The feeling that something bad almost happened creates a momentary tension, followed rapidly by a release of positive emotion when it becomes clear that this something bad did not happen. Winning a race, scoring a big promotion, getting the girl are all positive story outcomes with satisfying endings that leave audiences pleased. But a race won that was nearly lost, a promotion that was nearly a termination, and a girl who came so very close to falling in love with some other guy are all story endings with dramatic flare born of that feeling of “almostness,” and they leave the audience all the more satisfied.

Creating situations in which something else almost happens is a staple of good storytelling. As plots unfold, forks in the road, surprising twists, and the overall recognition of multiple possibilities breathe life into the story. Devices that the author can plant to emphasize the almostness, the palpable alternative that nearly happened, create dramatic tension. Some are blatant, as when an action hero’s sidekick suffers a gruesome death (eaten by an alligator, melted in a vat of acid), a fate spelled out in vivid detail so as to drive home the counterfactual that the very same fate nearly befall the hero. Some are only a little less blatant, as when the fate is not directly portrayed but rather hinted at: a vat of boiling acid or a pit of snakes are shown, the hero nearly falls in, but then does not. We see the acid or snakes, but we are left to imagine on our own how awful it would have been to have fallen into such places.

**Theme and Variation**

Counterfactuals may also enhance the persuasiveness and entertainment value of a narrative by way of their connection to the element of “theme and variation.” If reality is the theme, and counterfactual is the variation, then the juxtaposition of the two embodies a combination of the joy of recognition with surprise at something novel. Hofstadter (1979) argued that “the crux of creativity resides in the ability to manufacture variations on a theme” (p. 249). He
further argued that the ease with which ordinary people generate counterfactual thoughts is an example of the basic creative capacity of the human mind. In an important sense, artists who use variations on a theme are mimicking the natural manner in which the human brain sees the world. Brains comprehend reality by generating benchmarks built of past experience. When the brain sees something surprising, the experience of surprise itself comes from the mental benchmarks that pop to mind and reveal how things could have (or should have) been. Brains are continuously producing creative variations (i.e., counterfactual elaborations of alternatives to current experiences) as we experience the flow of events in our lives.

To become truly great art, theme and variation need a third companion, resolution. In a three-act play, three distinct sections correspond to establishing the setting, introducing a problem, and then presenting a solution to the problem. Like the three-act play, a “plot counterfactual” embraces a triplet structure. The theme is reality as we know it. The variation is the counterfactual, and contained in the counterfactual is some problem that confronts the main characters. The resolution is an ending that reveals some insight about the workings of reality that might otherwise have gone unrecognized. For example, in the film, *It’s a Wonderful Life*, the value of friendship and community is drawn into sharper focus by juxtaposing their factual form to a counterfactual alternative in which they are absent.

A constraint on the persuasive power of a counterfactual is its degree of variation, or the amount of alteration to reality, that it entails. Berlyne (1974) demonstrated experimentally that what strikes many as good art typically involves only slight deviation from expectations. Art that perfectly fits expectations is boring; art involving a great departure from the familiar strikes many as bizarre and repugnant. Somewhere between the extremes of the boring and the bizarre lays a sweet zone of recognition coupled with mild surprise. This principle advocated by Berlyne (1974) applies to counterfactuals as well, whether they are used by artists to influence an audience’s emotions, or if they are used as persuasive arguments to convince someone of a particular point of view. A compelling counterfactual, one that convinces an audience that some alternative might well have happened, must follow a “minimal rewrite rule” (Tetlock & Belkin, 1996). Small, minor changes to reality are acceptable, whereas bigger changes may leave the audience baffled. As psychological research on counterfactual thinking has shown, the regrets with which people kick themselves also follow this minimal rewrite rule (Roese & Summerville, 2005). People typically focus on just one action to alter within the counterfactual. All other aspects of reality remain within the counterfactual exactly as they truly are. In the best stories of the alternate history genre (in which the entire story takes place in a counterfactual world), there are a few key differences between the story’s setting and reality, framed by innumerable similarities, such as the laws of physics and basic characteristics of human nature. Counterfactuals
within narratives that follow this minimal rewrite rule are, we suggest, the most compelling and most persuasive (Lebow, 2000; Tetlock & Belkin, 1996).

Conclusion

Our primary goal has been to describe some research on the psychological basis of counterfactual thinking. This overview emphasized a functional view of such thought processes, which seeks to specify what goals they serve and what benefits they bring for the typical individual on a daily basis. We believe that these factors influence not only lay perceivers, but also scholars in various disciplines. Accordingly, an understanding of the psychological principles underlying counterfactual thinking might enable scholars of history, political science, philosophy, and literature to structure counterfactually based arguments more effectively. On the other hand, scholars might leverage their knowledge of psychological principles so as to create arguments that juxtapose against them. If cognitive biases maintain and reinforce conceptual parochialism, then deliberately avoiding such cognitive constraints may facilitate the realization of more creative, novel, and insightful analyses.

References


