

Cognitive Processes in Regret for Actions and Inactions

Ruth M.J. Byrne and Alice McEleney

Department of Psychology
University of Dublin, Trinity College
Dublin 2, Ireland
rmbyrne@tcd.ie

Abstract

Reasoning about matters of fact and reasoning about matters of possibility and impossibility may depend on the same sorts of mental representations and processes. We illustrate a mental model theory of counterfactual thinking with reference to the action effect (the tendency to regret actions more than inactions) and we describe an experiment which examines the effects of short-term and long-term perspectives on regret for actions and inactions.

Counterfactual Thinking

Everyday thinking focuses on facts or possibilities. It may focus on present facts, e.g., "Clinton is president of the US" or past facts, e.g., "Kennedy was president of the US", as Table 1 shows. It may focus on present possibilities (that could happen given the actual state of the world), e.g., "Clinton resigns" or past possibilities (that could have happened given the actual state but did not), e.g., "Clinton resigned in 1995" (see Johnson-Laird & Byrne, 1991, Chapter 4). It may even focus on impossibilities (that could never happen in the past or present), such as, "Clinton is president of Australia".

	Present	Past
Factual	Clinton is president	Kennedy was president
Non-factual Possibilities	Clinton resigns	<i>Clinton resigned in 1995</i>
Impossibilities	<i>Clinton is president of Australia</i>	<i>Clinton was president of Australia</i>

Table 1: Examples of factual and non-factual events in the past and the present time (with counterfactual events in italics).

Thinking about past possibilities and past or present impossibilities is often called counterfactual thinking¹ and it has been addressed by researchers interested in the psychology of counterfactual thinking (e.g., Kahneman & Tversky, 1982; Roese & Olson, 1995), as well as the philosophy (e.g., Lewis, 1973; Stalnaker, 1968), linguistics (e.g., Dudman, 1988) and artificial intelligence (e.g., Ginsberg, 1986) of counterfactual thinking.

Counterfactual thinking is important in many aspects of cognition. For example, thinking about counterexamples to conclusions helps people to make deductive inferences (e.g., Johnson-Laird & Byrne, 1991); and thinking about whether an outcome would have happened without its potential cause helps people to make causal inferences (e.g., Chisholm, 1946; Kahneman & Miller, 1986). Thinking about what would be possible helps people to construct sub-goals in problem-solving (e.g., Ginsberg, 1986; Keane, 1997) and in creative discovery (e.g., Hofstadter, 1985). Thinking about counterfactual possibilities helps people to improve their performance (e.g., Markman, Gavanski, Sherman, & McMullen, 1993). People think about the way things might have been to learn from mistakes (e.g., Roese & Olson, 1995) and these thoughts can result in a range of emotions including regret, relief, blame, guilt, and so on (e.g., Davis, Lehman, Wortman, Silver, & Thompson, 1995; Gilovich & Medvec, 1994).

The counterfactual scenarios that people generate tend to make minimal changes to the factual scenario (Pollock, 1986), although some counterfactual scenarios may be more similar to the factual scenario than others, or more accessible from it (Lewis, 1973; Stalnaker, 1968). The mutations may be along the "joints" of reality (Kahneman & Tversky, 1982), where reality is at its most "slippable" (Hofstadter, 1985). The mutability of aspects of reality may be guided by primary categories of mental life, such as space, time, causality, intentionality, and so on (Miller & Johnson-Laird, 1976). For example, temporal order affects mutability: people undo the most recent event more

¹ Thinking about counterfactual possibilities in the *future* is similar to thinking about counterfactual possibilities in the past but it also displays some intriguing vagaries (see e.g., Lewis, 1979).

decisions: John opts to stay where he is, and Paul decides to transfer. Suppose their decisions turn out badly for both of them: at the end of the year, John is even more unhappy where he is and wishes he had transferred, and Paul is even more unhappy at his new college and wishes he had stayed where he was. However it is too late for either of them to reverse his decision. Who do you think would regret his decision more on learning it was a mistake?

(see Byrne, McAlinney, & McEleney, 1997; Gilovich & Medvec, 1994). Most people believe that the individual who acted, Paul, would feel greater regret than the individual who did not act, John. The action effect occurs not only when people judge emotions from the perspective of imaginary characters but also when they recall events that they personally regret (Gilovich & Medvec, 1994).

One explanation of the action effect is that it arises because actions seem to depart from the status quo (Kahneman & Miller, 1986): actions are "abnormal" in this regard. Departures from normality may be more easily mutated than normal events because the abnormal event spontaneously recruits its corresponding norm (Kahneman & Miller, 1986). We suggest that people may construct models that represent actions explicitly, and inactions implicitly. For example, John and Paul are both in college A and this starting point may be represented in the models, along with the outcome that they are both unhappy. The action of switching to a different college may be represented explicitly, whereas the inaction of staying in the same college may be represented implicitly:

<i>John</i>	factual:	in college A	unhappy
<i>Paul</i>	factual:	in college A	unhappy
		switch to B	unhappy

Working memory limitations may prevent people from being able to construct all the possible counterfactual models: Multiple models are difficult to keep in mind. Instead people may construct a partial set of the counterfactual models on the basis of what is most readily mutated in their representation of the factual situation. People may find it easier to construct counterfactual scenarios by subtracting events rather than adding events (Kahneman & Tversky, 1982; Roese, 1994). Subtracting a model from the set (e.g., eliminating the second model in the factual set of models for Paul to construct a counterfactual set in which Paul is still in college A) results in fewer models to keep in mind than adding a model (e.g., inserting a second model in the factual set for John to construct a counterfactual set in which John switched to college B). Because people flesh out the counterfactual model for Paul, they judge that he will regret his action, not staying in college A, more than John will regret his inaction, staying in college A. Regret depends on comparing the way a situation turned out with an alternative way in which it might have turned out better (e.g., Landman, 1987). Because an alternative is constructed more readily for Paul, he is judged to experience more regret.

We suggest that the action effect results from the properties of the mental representations that people construct (Byrne, 1997). First, they construct an initial set of models that represents as little information as possible explicitly because of the limitations of working memory: they mentally represent the action explicitly in their models and the inaction implicitly. Second, they do not construct all the possible counterfactual models. Third, aspects of the factual situation that are represented explicitly in models are easier to mutate than aspects that are not represented explicitly. Finally, the number of models that must be kept in mind is an important constraint in counterfactual thinking, just as it is in factual thinking. A counterfactual scenario may be more readily constructed by eliminating models than by fleshing out models. These simple principles underlie the representation of the factual situation and the generation of a counterfactual situation based on it. These principles can account for the observation that people tend to regret their actions more than their failures to act.

The model theory of the action effect also explains the related phenomena that comprise the core of the action effect. For example, the action effect holds for good outcomes too. Most people believe that an individual who acted would feel better about the action than an individual who did not act (Landman, 1987). However, the action effect for good outcomes is not as strong as the action effect for bad outcomes (e.g., Gleicher, Kost, Baker, Strathman, Richman, & Sherman, 1990). For bad outcomes, people judge that an individual who acted will feel worse than an individual who failed to act regardless of whether they are given an explicit counterfactual alternative ("He now finds out that he would have been better off if he had switched to college B.") or not ("He now finds out that he is worse off because he stayed in college A."). But for good outcomes, people judge that the individual who acted will feel worse than the individual who failed to act, only when an explicit counterfactual alternative is provided (Gleicher et al, 1990). In the absence of bad outcomes, the counterfactual models may not be fleshed-out for either protagonist. But, when people are given counterfactual alternatives, the alternatives are represented explicitly in the set of models. Once they are prompted to think of the counterfactual negative outcome in this way, the action effect for good outcomes becomes as high as it is for bad outcomes. The provision of the explicit counterfactual alternative has no effect when the factual outcome is bad because these situations spontaneously bring to mind their good counterparts.

Long Term Perspectives on Action and Inaction

Although people regret their actions more than their failures to act in the short-term, they appear to regret their failures to act more than their actions in the long-term (Gilovich & Medvec, 1994). When people are asked to think back over their own lives and say what they regret most, they tend to think of things they failed to do, such as not pursuing hobbies or educational opportunities, not

spending enough time with family and friends, not 'seizing the moment' (Gilovich & Medvec, 1994). Their memories for things they regretted from their past lives shows an *inaction* effect. Gilovich & Medvec (1994) demonstrated this inaction effect from a long-term perspective with the sort of college-choice scenario described earlier. Most of the people who were asked the question: 'Who do you think would regret his decision more upon learning that it was a mistake?' indicated that the individual who had acted would feel more regret. But most of the people who were asked the question 'Who do you think would regret his decision more in the long run?' thought that the individual who failed to act would feel more regret.

We have suggested that the action effect arises because the models that people construct of the scenario represent the action explicitly and the inaction implicitly. To construct a counterfactual scenario, it may be easier to eliminate the model of the action, which is represented explicitly in the models of the factual situation for Paul, than it is to add a model of an action to the models of the factual situation for John. But from a long-term perspective, it appears that people construct the counterfactual scenario by adding a model rather than deleting it. In what circumstances is it easier to add a model than to eliminate a model from the set? In the domain of deductive reasoning, content and context can help people to flesh out models (Johnson-Laird & Byrne, 1991). How does temporal perspective facilitate fleshing out models in the domain of counterfactual thinking?

One possibility is that the temporal proximity of events in the short-term may make them mutable, just as it is easier to imagine catching an airplane missed by 5 minutes than an airplane missed by an hour (e.g., Miller & McFarland, 1987). From the long-term perspective, the temporal distance of events may make them seem part of the immutable, pre-supposed background. Temporal proximity may affect actions because they are represented explicitly, but not inactions which are represented implicitly. Temporal proximity may lead actions, because they are represented explicitly, to be considered mutable in the short-term but immutable in the long-term. Inactions, because they are represented implicitly, may remain at a constant lower level of mutability regardless of temporal perspective and so they may be less mutable than actions in the short-term but more mutable in the long-term.

A second possibility is that people maintain a good outlook and flesh out their models to contain possible good outcomes from events. Gilovich and Medvec (1994), in considering several parallel explanations of the reversal of the action effect to an inaction effect in the long term, suggest that people may feel more responsible for their actions. When actions lead to bad outcomes, they experience more 'cognitive dissonance' than when their inactions lead to bad outcomes. As a result of this dissonance, they do more mental work to look on the bright side of the outcomes of their actions: they seek out the silver linings in clouds that arise from their actions

more than they do in ones that arise from their failures to act. Hence, the only events people continue to perceive to lead to bad outcomes from the long-term perspective are their failures to act (Gilovich & Medvec, 1994). It may be that people flesh out their models to contain not only the bad outcomes but also the possible good outcomes from the action, because actions are represented explicitly. They do not flesh out their models to contain possible good outcomes from the inaction, because it is represented implicitly. Over time, they may focus on the model of the action and its good outcome; the model of the action and its good outcome may continue to be represented explicitly, and the model of the action and its bad outcome may be represented implicitly. In the long term, their representation of the factual situation for Paul may contain a model of his action and a good outcome, and their representation of the factual situation for John may contain a model of his inaction and a bad outcome.

To examine these different explanations we carried out an experiment to establish whether the reversal of the action effect to an inaction effect in the long-term also occurs when people think about events that turned out well. Our aim in the experiment was to test whether the reversal of the action effect in the short-term to an inaction effect in the long-term is confined to situations with bad outcomes, such as being unhappy in college, or whether it also occurs for situations with good outcomes, such as being happy in college. The two alternative explanations lead to different predictions. If temporal proximity mediates fleshing out models than the reversal of the action effect to an inaction effect in the long-term should occur for scenarios that lead to good outcomes just as it does for scenarios that lead to bad outcomes. If a good outlook mediates fleshing out models than the reversal of the action effect to an inaction effect in the long-term should occur for scenarios that lead to bad outcomes but not for scenarios that lead to good outcomes, where there is no need to look on the bright side.

We gave four groups of subjects one of four sorts of scenario based on John and Paul and their college choices: a short-term bad-outcome version, a long-term bad-outcome version, a short-term good-outcome version, and a long-term good-outcome version. We constructed the four different sorts of scenarios (described in the Appendix) based on Gilovich and Medvec's (1994) college-choice scenario, which we altered in several ways to clarify the finality of the decision, and especially its long-lasting impact (see Byrne, et al, 1997, for details). We assigned 190 undergraduates from Dublin University, Trinity College, at random to one of four groups, and we gave each group one of the four sorts of scenarios. The participants task was to tick the name of one of the individuals, John or Paul, in response to the question at the end of the scenario.

As Table 2 shows, most of the subjects who received the short-term bad-outcome scenario indicated that Paul, the individual who acted, would feel more regret (61%). In contrast, most of the subjects who received the long-

term bad-outcome version indicated that John, the individual who had failed to act, would feel more regret (76%). More subjects judged that Paul who acted would feel more regret in the short-term scenario (61%) than in the long-term scenario (24%) and this difference is reliable ($\text{Chi}^2 = 12.86$, $df = 1$, $p < 0.001$). The experiment thus replicates Gilovich and Medvec's finding of a reversal from an action effect to an inaction effect over time for bad outcomes. As Table 2 also shows, this reversal does not occur for good outcomes. Most subjects who received the short-term good-outcome version indicated that Paul, who acted, would feel better about his decision (70%), and likewise, most subjects who received the long-term good outcome version indicated that Paul, who acted, would feel better about his decision (76%) and there is no reliable difference between them ($\text{Chi}^2 = 0.23$, $df = 1$, $p > 0.05$).

The results show that for bad outcomes people regret their actions from the short term perspective and they regret their failures to act from the long-term perspective; but for good outcomes they feel better about their actions from both the short term and the long term perspective.

	Short term	Long term
Bad outcome		
Action	61	24
Inaction	39	76
Good outcome		
Action	70	75
Inaction	30	25

Table 2: The percentages of judgments of greater emotion for actions and inactions

The reversal to an inaction effect is confined to a long-term perspective on bad outcomes and it does not occur for good outcomes. The data suggest that the reversal to an inaction effect in the long term for bad outcomes is not the result of temporal proximity, but the result of a good outlook. People flesh out their models to contain not only the bad outcome but also the possible good outcomes from the action, because actions are represented explicitly. In time, they focus on the action and its good outcome more than its bad outcome. They do not flesh out their models to contain good outcomes from the inaction, because it is represented implicitly. Their models come to represent the action and its good outcome and the inaction and its bad outcome. For scenarios that lead to good outcomes, there is no need to look on the bright side, and so the reversal does not occur.

Conclusions

People feel more strongly about their actions than their inactions. We suggest that the action effect can be explained by the following principles: people construct an

initial set of models that represents as little information as possible explicitly: they represent the action explicitly in their models. They do not flesh-out all the counterfactual models because of the constraints of working memory. Aspects of the factual situation that are represented explicitly in models are easier to mutate than aspects that are not represented explicitly. The model theory of counterfactual thinking has been advanced to account for the action effect (Byrne, et al, 1997), as well the temporality effect (Byrne, Culhane, and Tasso, 1995), and a new spatial effect predicted by the theory (Byrne, 1997).

A theory of the mental representations and processes that underlie the generation of counterfactual scenarios requires a theory of the mental representations and processes for factual scenarios. Reasoners may construct mental models that correspond to the way the world would be if the assertions were true. They may usually represent what is true in their models; a unique feature of counterfactual thinking is that it requires people to represent explicitly not only what is true but also what is false, temporarily supposed to be true. We suggest that the representations and processes required for reasoning about matters of fact and matters of possibility are essentially the same. What differs is what is represented explicitly in the initial set of models.

Acknowledgments

We thank Phil Johnson-Laird and Mark Keane for their helpful comments on this research. Alice McEleney is supported by a Dublin University Foundation Scholarship.

Appendix

The materials used in the experiment.

First paragraph (common to all four scenarios).

John and Paul do not know each other, but both are enrolled at the same elite university. Both are only moderately satisfied where they are and both are considering transferring to another prestigious university. They are both told they must make their final decision before the end of the year. Each agonizes over the decision, going back and forth between thinking he is going to stay and thinking he will leave. They ultimately make different decisions: John opts to stay where he is, and Paul decides to transfer.

Second paragraph:

1. Long-term bad-outcome:

Suppose their decisions turn out badly for both of them: at the end of the year, John is even more unhappy where he is and wishes he had transferred, and Paul is even more unhappy at his new college and wishes he had stayed where he was. However it is too late for either of them to reverse his decision. As a result they both drop out of college and neither of them ever secures a good job. Who do you think would regret his decision more on looking back on it ten years later?

2. Short-term bad-outcome:

Suppose their decisions turn out badly for both of them: at the end of the year, John is even more unhappy where he is

and wishes he had transferred, and Paul is even more unhappy at his new college and wishes he had stayed where he was. However it is too late for either of them to reverse his decision. Who do you think would regret his decision more on learning it was a mistake?

3. Long-term good-outcome:

Suppose their decisions turn out well for both of them: at the end of the year, John is happier where he is and is glad he stayed where he was, and Paul is happier at his new college and is glad he transferred. They both do very well at college and secure good jobs after graduating. Who do you think would feel better about his decision looking back on it ten years later?

4. Short-term good-outcome:

Suppose their decisions turn out well for both of them: at the end of the year, John is happier where he is and is glad he stayed where he was, and Paul is happier at his new college and is glad he transferred. Who do you think would feel better about his decision on learning that it was the right one?

References

- Byrne, R.M.J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D.L. Medin (Ed). *The Psychology of Learning and Motivation*, Vol 37. San Diego, CA: Academic Press. (In press).
- Byrne, R.M.J. & Handley, S.J. (1997). Reasoning strategies for suppositional deductions. *Cognition*, 62, 1-49.
- Byrne, R.M.J. & Tasso, A. (1994). Counterfactual reasoning: inferences from hypothetical conditionals. In Ram, A & Eiselt, K. (Eds). *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum.
- Byrne, R.M.J., Culhane, R., and Tasso, A. (1995). The temporality effect in thinking about what might have been. In J.D. Moore and J.F. Lehman (Eds.) *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum, p. 385-390.
- Byrne, R.M.J., McAlinney, P. and McEleney, A. (1997). Acts and failures to act: Short-term and long-term perspectives. *Manuscript in preparation*.
- Chisholm, R.M. (1946). The contrary-to-fact conditional. *Mind*, LV, 289-307.
- Davis, C. G., Lehman, D. R., Wortman, C. B., Silver, R.C., & Thompson, S.C. (1995). The undoing of traumatic life events. *Personality and Social Psychology Bulletin*, 21, 109-124.
- Dudman, V. H. (1988). Indicative and subjunctive. *Analysis*, 48, 113 - 122.
- Gilovich, T. & Medvec, V.H. (1994). The temporal pattern to the experience of regret. *Journal of Personality and Social Psychology*, 67, 357-365.
- Ginsberg, M. L. (1986). Counterfactuals. *Artificial Intelligence*, 30, 35 - 79.
- Gleicher, F., Kost, K.A., Baker, S.M., Strathman, A.J., Richman, S.A., & Sherman, S.J. (1990). The role of counterfactual thinking in judgments of affect. *Personality and Social Psychology Bulletin*, 16, 284-295.
- Hofstadter, D. R. (1985). *Metamagical Themas: Questing for the Essence of Mind and Pattern*. London: Penguin.
- Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. & Byrne, R.M.J. (1991). *Deduction*. Hove and Hillsdale: Erlbaum.
- Kahneman, D. & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136-153.
- 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). New York: Cambridge University Press.
- Keane, M.T. (1997). What makes an analogy difficult? The effects of order and causal structure in analogical mapping. *Journal of Experimental Psychology: Language, Memory & Cognition*. In press.
- Landman, J. (1987). Regret and elation following action and inaction: affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin*, 13, 524-536.
- Lewis, D. (1973). *Counterfactuals*. Oxford, Blackwell.
- Lewis, D. (1979). Counterfactual dependence and time's arrow. *Nous* XIII, 455-76.
- Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focusing in reasoning and decision-making. *Cognition*, 49, 37-66.
- Markman, K.D., Gavanski, I., Sherman, S.J. & McMullen, M.N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, 29, 87-109.
- 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, 1111-1118.
- Miller, D.T. and McFarland, C. (1987). Counterfactual thinking and victim compensation: a test of norm theory. *Personality and Social Psychology Bulletin*, 12, 513-519.
- Miller, G. and Johnson-Laird, P.N. (1976). *Language and Perception*. Cambridge: Cambridge University Press.
- Pollock, J. L. (1986). *Subjunctive reasoning*. Dordrecht Reidel.
- Roese, N.J. & Olson, J.M. (1995). *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Mahwah, New Jersey: Erlbaum.
- Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford: Basil Blackwell.
- Wells, G.L., Taylor, B.R., & Turtle, J.W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421-430.