Episodic Counterfactual Generation Prioritizes Plausible and Vivid Imagined Alternatives

Ricardo Morales-Torres (rim11@duke.edu)[1] & Felipe De Brigard (felipe.debrigard@duke.edu) [2]

[1] Center for Cognitive Neuroscience, Duke University

[2] Department of Philosophy, Duke University

Abstract

Episodic counterfactual thinking refers to the capacity to imagine different ways in which the personal past might have unfolded. However, when generating imagined alternative versions of personal past events, people face a potentially unlimited number of ways to alter them. Current research on modal cognition has posited that when generating imagined alternative versions of events, people tend to sample more plausible alternatives first. In the current study, we asked whether this sampling mechanism is engaged when people generate alternatives for their personal past. To address this question, we asked participants to recall an episodic autobiographical event and then sequentially generate four alternative versions of that event. Employing Bayesian mixed models, our results showed credible evidence that participants sampled more plausible and vivid alternatives first, while showing no credible evidence that simulation difficulty change across sequential generations.

Keywords: Episodic Counterfactual Thinking, Plausibility, Vividness, Difficulty.

Introduction

Episodic counterfactual thinking (ECT) refers to the capacity to imagine alternative ways in which one's personal past might have been different (De Brigard & Parikh, 2019). One common aspect of ECT is that it can vary widely in plausibility: people can imagine alternative versions of events that are very likely to have occurred, or they can imagine versions that are extremely unlikely. For example, when revisiting a memory in which I tripped, fell in front of my high school crush and she laughed at me, I can generate a plausible counterfactual in which she was not looking in my direction. Conversely, I can imagine an implausible alternative in which I tripped over a thief who was mugging her and thus became her hero. An influential proposal argues that counterfactual plausibility depends on the difficulty of generating the mental simulation (Kahneman & Tversky, 1982b, 1982a). Similarly, another classic proposal posits that the vividness of a simulation directly influences its perceived plausibility (Anderson, 1983; Carroll, 1978). However, the cognitive mechanisms driving the perceived plausibility of ECT remain poorly understood.

Current theories regarding modal cognition posit that when people generate imagined alternatives for events—as in ECT—they tend to sample the most plausible ones first (Knobe & Cushman, 2023; Phillips et al., 2019). This sampling process has been posited as an answer to how people can generate alternative versions of events, as they are faced with an a priori unlimited space of alternatives. We argue that this same problem arises for ECT. When people need to generate an alternative for a past autobiographical experience, there are potentially unlimited ways in which those events might have been different. Thus, the main objective of this study is to explore whether this sampling process is employed when subjects engage in ECT.

To support the general claim that the perceived plausibility in ECT is sensitive to sampling order, the current paper has two specific aims. The first aim is to investigate whether generating ECT engages a sampling process in which specific mental simulations of imagined alternative ways in which past personal events might have been different are differentially prioritized. More specifically, we propose that—similar to other forms of modal cognition (Phillips & Cushman, 2017)—the sequential generation of ECT involves a sampling process that prioritizes the initial selection of imagined plausible alternatives to past events. Thus, we expect that when people mentally generate alternative versions to remembered autobiographical events, the first imagined alternative they generate will tend to be the most plausible one.

The second aim of this study is to test whether this proposed sampling mechanism influences other key features of ECT, namely the vividness and difficulty of the mental simulation. As mentioned before, these two features have been posited as influencing counterfactual plausibility. By including these two measures, we can examine, (1) whether changes in difficulty necessarily vary with changes in plausibility—as predicted by the view that that plausibility depends on difficulty (Kahneman & Tversky, 1982a, 1982b)—and (2) whether this sampling process would not only recruit the more plausible alternatives first but also the more vivid ones, which would be consistent with current models of mental simulation according to which vividness relates to the congruency of a mental representation with prior knowledge (Riley & Davies, 2023). Such a result would then suggest that plausibility too may be related to the congruency between the simulated content and prior knowledge (Connell & Keane, 2006).

To this end, we sourced data from a companion study conducted by Morales-Torres et al. (2026) in which participants were asked to recall autobiographical events and generate imagined alternative versions of those past events (see the Methods section for more details). The objective of

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

the current study is to expand on the frequentist analysis reported in Morales-Torres et al. (2026) by employing a Bayesian framework to quantify the evidence that repeated mental simulation influences plausibility, vividness, and difficulty ratings.

Methods

Participants

The current study sourced data from Morales-Torres et al. (2026). More specifically, we combined the samples employed for Experiment 2a and Experiment 2b into one dataset. Data was obtained through Prolific, participants were based in the US, were native English speakers, and had an approval rating of at least 90%. We removed participants with at least one invalid counterfactual description (e.g., overly general responses like "petting my dog," invalid responses such as "bread," blank responses, or nonsensical strings like "asdf"). After exclusions, the final sample consisted of 468 participants (191 males, 271 females, 6 did not report gender; Mean age = 38.1, SD = 10.3). The study was approved by the Duke University Campus IRB.

Procedure

The experimental procedure consisted of two stages (**Figure 1**). First, participants were asked to recall a specific autobiographical event not older than 10 years. After recalling an event, participants provided: (1) a brief description of the event; (2) a short title for the event; (3) the date, location, a person, and an object associated with the event. After annotating these details, participants rated the valence (1: Very negative to 5: Very positive) and the vividness (1: Vague with no/few details to 5: Vivid and highly detailed) of the recalled event.

The second stage consisted of a sequential counterfactual generation task that came immediately after the first stage. Participants were asked to think about a different way in which the event they recalled might have happened. During this stage, participants were cued with the title of their memory and were given 12 seconds to generate the mental simulation. They were then asked to provide a brief description of the content of their simulation. Following this, participants rated the personal subjective plausibility (1: Not likely to have happened to me to 5: Very likely to have happened to me), vividness (1: Vague with no/few details to 5: Vivid and highly detailed), difficulty (1: Very easy to 5: Very hard), valence (1: Highly negative to 5: Highly positive), and frequency of thought for that counterfactual (1: The first time I think about this to 5: I have thought about this several times). This second stage was repeated three more times, such that participants sequentially generated a total of four alternative versions for the same memory. Participants were not informed that they would be generating more than one counterfactual.

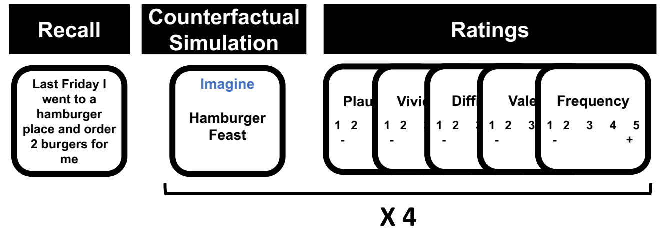

Figure 1: Experimental Procedure. Participants recalled an autobiographical memory and sequentially generated four different alternative versions of that event. After generating each counterfactual simulation, participants rated its plausibility, vividness, difficulty, valence, and frequency of prior thought.

Analysis

As a first frequentist analysis, for each of the three key measures (Plausibility, Vividness, and Difficulty) we computed a difference score by subtracting each participant's rating at the fourth simulation from their rating at the first simulation. These difference scores were compared against zero using one-sample t-tests. We then directly tested whether the difference scores were equivalent to zero (with an equivalence region corresponding to a small effect size: d = 0.3), using a Two One-Sided Test (TOST) employing the equivalence R package (Robinson & Robinson, 2016).

Then, we calculated three different Bayesian Linear Mixed Effects models, each one with one of the three key measures (Plausibility, Vividness or Difficulty) as the main predicted variable —with the other 2 serving as covariates. As the main predictor variable, we included Simulation Order (1 to 4). Frequency of Thought and the Valence of the simulation were included as covariates. Simulation Order was also modeled as a random slope, and subjects were modeled as random intercepts. Priors for each beta value were centered at zero to provide an unbiased direction for the effect; while the prior for the intercept was set at 2.5 (indicating the middle of the 5-point scale).

The models were computed employing the brms R package (Bürkner, 2017). For each model we ran four chains with 16000 iterations and 2000 warm-ups each (all models showed good convergence with $\hat{R} < 1.01$). For each beta estimate of Simulation Order, we report: (1) 95% High Density Interval (HDI); (2) Bayes Factor computed employing the Savage-Dickey density ratio (Wagenmakers et al., 2010), which compares the posterior density to the prior density: (3) The probability of direction (pd), which indicates the proportion of posterior draws that have the same sign as the median (a probability of direction of 0.975 is equivalent to two-sided p-value of .05).

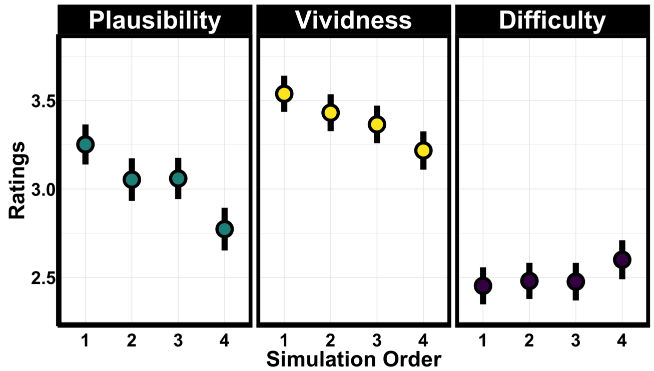

Figure 2: Average rating (and standard errors) as a function of Simulation Order and Feature type. Error bars represent Standard Error of the Mean.

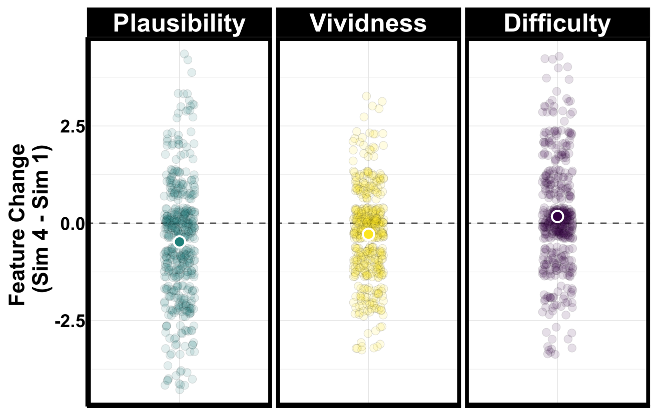

Figure 3: Feature Change: Difference scores as a function of Feature Type. Small dots represent individual participants, whereas large dots represent Feature Change averages.

Results

Before conducting our main analysis, we visually inspected the ratings (Figure 2), which revealed a pattern in which vividness and plausibility ratings decreased across subsequent simulations, while difficulty ratings appeared to increase. To test the consistency of this pattern, we conducted one-sided t-tests and a TOST on the difference scores (Figure 3). These analyses revealed that Plausibility ratings showed a significant decrease from the first to the fourth simulation ($M = -0.48$, $SE = 0.07$; $t(467) = -6.80$, $p < .001$, Cohen's $d = -0.31$; TOST lower $p = .6$; TOST upper $p < .001$). Similarly, vividness ratings also showed a significant decrease from the first to the fourth simulation ($M = -0.32$, $SE = 0.06$; $t(467) = -5.70$, $p < .001$, Cohen's $d = -0.26$; TOST lower $p = .214$; TOST upper $p < .001$). There was a significant difference between the first and fourth simulation on difficulty ratings ($M = 0.15$, $SE = 0.06$; $t(467) = 2.31$, $p = .021$, Cohen's $d = 0.11$); however, this difference was within the equivalence boundary (TOST lower $p < .001$; TOST upper $p = .022$).

Next, we focused on the results from the Bayesian Linear Mixed Effects model. Our results showed that there was credible evidence that Plausibility ($b =-0.07$, $SE = 0.02$, 95% HDI = [-0.11 -0.03], pd = 0.99, BF10 = 27.23) and Vividness ratings ($b = -0.05$, $SE = 0.02$, 95% HDI = [-0.09 -0.02], pd =0.99, BF10 = 6.32) decreased across sequential simulations. By contrast, there was no credible evidence that difficulty ratings decreased across sequential simulations ($b = 0.02$, $SE = 0.02$, 95% HDI = [-0.02 0.06], pd = 0.82, BF10 = 0.06).

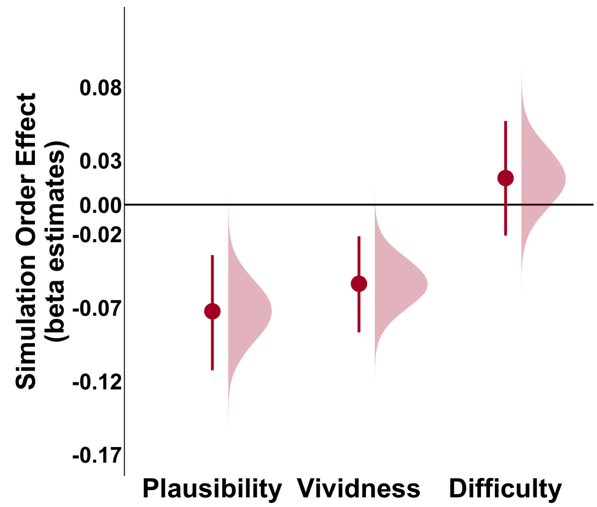

Figure 4: Beta estimates derived from the Bayesian LMM in which Feature Type was predicted by Simulation order. Dots represents the median of the posterior, Error bar represents 95% High Density Intervals, while the distribution represent the posterior density.

Discussion

In the current study, we sought to test whether the generation of episodic counterfactual simulations—imagined alternative ways in which past personal events might have been different (De Brigard & Parikh, 2019)—engages a sampling process over the space of possible alternative representations. To do this, we employed a paradigm in which we asked participants to recall an autobiographical event and then generate four different alternative ways in which that event might have happened. Our results revealed that participants generated more plausible and vivid alternatives first. Conversely, the different counterfactual alternatives were equally difficult to simulate. We discuss these results in the following paragraphs, as well as some limitations of the current study.

Current theories of modal cognition posit that the generation of alternative versions of events is influenced by a sampling mechanism that prioritizes more plausible alternatives first (Phillips et al., 2019; Phillips & Cushman, 2017). Our results extend this finding to ECT: when participants generate alternative versions of their past, they first generate more plausible and vivid alternatives.

Moreover, the first counterfactuals were not only more plausible but also more vivid. Crucially, we controlled for the familiarity of each simulation by including the frequency of thought as a covariate, indicating that our results are not due to mere familiarity with the simulations.

We interpret our results as supporting that ECT are generated through a sampling mechanism that is strongly influenced by what participants know about their lives and the world—that is, what is likely to have happened. Given that this mechanism would sample representations that fit better with prior knowledge, these representations would be rated as more plausible (Connell & Keane, 2006) as well as more vivid (Riley & Davies, 2023), as both features have been shown to be strongly influenced by their correspondence with prior knowledge. However, this interpretation is speculative and should be supported by future experiments in which prior knowledge is manipulated.

Our second finding showed that the sequential generation of episodic counterfactuals did not involve any change in simulation difficulty: the first and last episodic counterfactual simulations were equally difficult to generate. This result, combined with the changes in plausibility, contrasts with classical theories positing that plausibility is determined by simulation difficulty (Kahneman & Tversky, 1982a, 1982b), which has been a highly influential proposal across several strands of psychology (e.g., Bennett et al., 2022; Evans, 2008; Galinsky & Moskowitz, 2000). In contrast, our results indicate that episodic counterfactual plausibility is best captured by a sampling process rather than being tied to simulation difficulty.

One important aspect of our methodology is that the current study collected subjective measures of plausibility, vividness, and difficulty. Thus, the subjective nature of these ratings could raise two potential concerns: first, regarding the comparability of ratings across participants and, second, regarding the validity of these measures. To address the first concern, our study employed linear mixed-effects models in which sequence effects were modeled as random slopes and participants as random intercepts, allowing us to model effects that were consistent across participants regardless of how each subject interpreted the 5-point scale. Nevertheless, future studies could employ more objective measures of the simulation, such as the time needed to simulate an alternative event (Phillips & Cushman, 2017). Regarding the second point, several studies have found that subjective plausibility ratings modulate cortical and subcortical neural activity (De Brigard et al., 2013; Khoudary et al., 2022) as well as having downstream effects on subsequent memory of counterfactual details (Morales-Torres et al., 2025). Thus, even though we asked participants to provide subjective ratings of plausibility, there is reasonable evidence to suggest that these ratings carry meaningful functional implications for behavior.

Our study has several limitations. First, we did not instruct participants to engage in any specific type of counterfactual mutation. Prior research has shown that generating better or worse alternatives for past events have different influences on future behavior (Markman et al., 1993). Whether our results generalize to different types of counterfactual simulations remains an open question for future research. Second, while participants were blind to the repeated nature of the task for the second simulation, for the third and fourth they knew that they were going to be asked to generate alternative counterfactual simulations for that memory. We believe that our results hold even when comparing the first with the second simulation (when participants are unaware of the sequential nature) and the first with the fourth (when participants are aware of it). Nevertheless, future studies could employ between-subjects designs to test whether order effects differ between these two comparisons.

Given the relevance of counterfactual thinking to mental health, particularly in the form of ruminative thinking (De Brigard & Hanna, 2015; Markman & Miller, 2006), understanding the mechanisms that generate them is an important question for cognitive science. Future studies could test whether the sampling mechanism proposed here is modulated in depression. One potential hypothesis to explore is that individuals with depression may preferentially sample counterfactuals that fit more negative alternatives (rather than more plausible ones).

In sum, our results showed that when participants need to generate alternative versions of their past, they tend to generate the most plausible and vivid alternatives first. However, the perceived difficulty of generating the mental simulation does not seem to determine its different levels of plausibility or vividness.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Carroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, *14*(1), 88–96. https://doi.org/10.1016/0022-1031(78)90062-8

Connell, L., & Keane, M. T. (2006). A Model of Plausibility. *Cognitive Science*, *30*(1), 95–120. https://doi.org/10.1207/s15516709cog0000_53

De Brigard, F., & Hanna, E. (2015). Clinical applications of counterfactual thinking during memory reactivation. *Behavioral and Brain Sciences*, *38*, e5. Cambridge Core. https://doi.org/10.1017/S0140525X14000351

De Brigard, F., & Parikh, N. (2019). Episodic Counterfactual Thinking. *Current Directions in Psychological Science*, *28*(1), 59–66. https://doi.org/10.1177/0963721418806512

De Brigard, F., Szpunar, K. K., & Schacter, D. L. (2013). Coming to grips with the past: Effect of repeated simulation on the perceived plausibility of episodic counterfactual thoughts. *Psychological Science*, *24*(7), 1329–1334. https://doi.org/10.1177/0956797612468163

Kahneman, D., & Tversky, A. (1982a). The psychology of preferences. *Scientific American*, *246*(1), 160–173. https://doi.org/10.1038/scientificamerican0182-160

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

Khoudary, A., 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: Biological Sciences*, *377*(1866), 20210337. https://doi.org/10.1098/rstb.2021.0337

Knobe, J., & Cushman, F. (2023). The common effect of value on prioritized memory and category representation. *Trends in Cognitive Sciences*, *27*(10), 892–900. https://doi.org/10.1016/j.tics.2023.06.007

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*(1), 87–109. https://doi.org/10.1006/jesp.1993.1005

Markman, K. D., & Miller, A. K. (2006). Depression, Control, and Counterfactual Thinking: Functional for Whom? *Journal of Social and Clinical Psychology*, *25*(2), 210–227. https://doi.org/10.1521/jscp.2006.25.2.210

Morales-Torres, R., Miceli, K., & De Brigard, F. (2026). Plausibility in episodic counterfactual thinking does not depend on the difficulty of the mental simulation. *Cognition*, *271*, 106424. https://doi.org/10.1016/j.cognition.2025.106424

Morales-Torres, R., Miceli, K., Huang, S., Szpunar, K., & De Brigard, F. (2025). Episodic details are better remembered in plausible relative to implausible counterfactual simulations. *Psychonomic Bulletin & Review*, *32*(4), 1852–1859. https://doi.org/10.3758/s13423-025-02670-0

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

Riley, S. N., & Davies, J. (2023). Vividness as the similarity between generated imagery and an internal model. *Brain and Cognition*, *169*, 105988. https://doi.org/10.1016/j.bandc.2023.105988