Learning to Plan from Actual and Counterfactual Experiences

Justin Yang & Tobias Gerstenberg
{justin.yang, gerstenberg}@stanford.edu
Department of Psychology, Stanford University

Abstract

Our ability to plan and make effective decisions depends on an accurate mental model of the environment. While learning generally requires novel external observations, people can also improve their understanding by reasoning about past experiences. In this work, we examine whether counterfactual simulation enhances learning in environments where planning is straightforward but encoding new information is challenging. Across two studies, participants navigated gridworlds, learning to avoid hazardous tiles. Some participants were asked to engage in counterfactual simulation, constructing alternative plans after observing navigation outcomes. Others learned purely from experience. While counterfactual paths contained fewer hazards than actual paths, we found reliable evidence across both studies that counterfactual simulation conferred no measurable advantage in either navigation performance or explicit environment learning. These findings shed new light on the scope of learning by thinking – suggesting that the mechanism by which counterfactual reasoning enhances learning might not be by encouraging deeper encoding of past experiences.

Keywords: counterfactual simulation, learning, thinking, planning

Introduction

Imagine that you've just lost a game of chess. Frustrated and determined to redeem yourself, you spend the next few days preparing for a rematch. What can you do to make yourself a more formidable opponent? One option is to gain *experience* by playing more games. Another is to reflect on past mistakes through *counterfactual simulation* – thinking about what went wrong and what could have led to a better outcome (Byrne, 2016; Gerstenberg, 2024). This is an instance of "learning by thinking": somehow, without gaining any new external information, people may gain new insights by reasoning internally about what they already know (Lombrozo, 2024).

Learning by thinking is both pervasive and impactful (Schwartz & Black, 1999). For example, prior work has shown how self-explanations can help people understand the limits of their knowledge (Rozenblit & Keil, 2002) and facilitate deeper understanding in classroom settings (Chi et al., 1989). Thought experiments have played an important role throughout the history of science (Gendler, 1998; Norton, 1991). But how can learning occur without new external input?

Lombrozo (2024) suggests that learning by thinking makes some forms of knowledge more *accessible* to the learner in a process termed "representational extraction" (see also Cushman, 2020). Accordingly, people learn by transforming knowledge into a more explicit format. For example, a chess player can use their knowledge of how pieces move to learn how a move would affect the coverage of their pieces without external feedback. Is this the only way that thinking might lead to new knowledge?

Prior work in social psychology characterizes the functional role of one kind of thinking – counterfactual reasoning – into two main categories: upward and downward (Epstude & Roese, 2008; Roese, 1994). Downward counterfactual thinking compares what actually happened to *worse* alternatives. These potentially serve a regulatory role, reducing the feeling of regret about negative outcomes (Parikh et al., 2022). In this way, future learning may be facilitated indirectly through emotion regulation and motivation. Upward counterfactual thinking compares what happened to *better* alternatives. These counterfactuals are thought to serve a *preparatory role*, guiding us toward better decisions in the future. This kind of thinking often follows failure or disappointment and can help with taking better actions the next time (Markman et al., 1993; Roese, 1997; Roese & Olson, 1993; Smallman & McCulloch, 2012).

Here, we investigate whether counterfactual thinking enhances learning in a planning task where success depends on both acquiring an accurate model of the environment and planning effectively with that model. Specifically, we designed a task in which participants navigate between varying start and goal locations in a gridworld, learning to avoid hazardous tiles that slow them down (see Figure 1). While planning an efficient route given a known model of the environment is straightforward here, learning the probabilistic structure of the environment is challenging. We explore whether asking participants to simulate counterfactual alternatives in this task may result in improved learning and performance.

To foreshadow our main result, we found across two studies that participants who engaged in counterfactual simulation did *not* perform better than participants who learned purely from experience. Even though participants generated counterfactual simulations that would have resulted in better outcomes in a given situation, doing so did not lead them to construct a better model of the environment. or to plan better with their learned model. These findings highlight the boundaries of learning by thinking, suggesting that counterfactual rea-

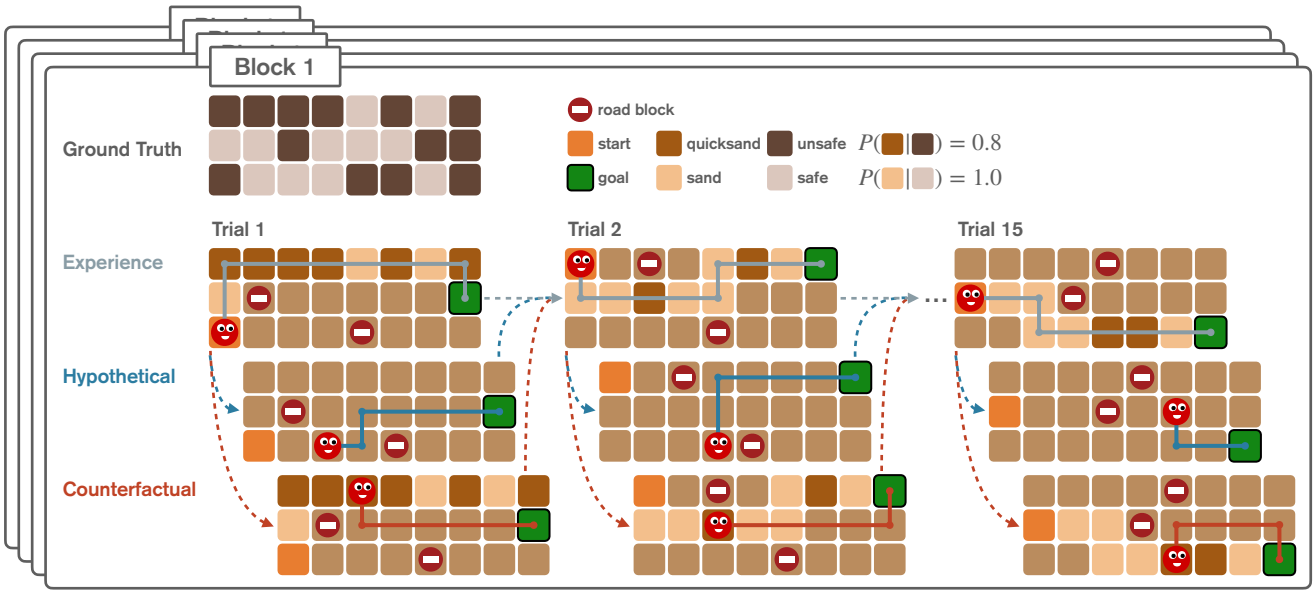

Figure 1: **Experimental paradigm**. Across both studies, participants navigated from start (orange tile) to goal position (green tile) while trying to minimize the number of quicksand tiles encountered. The ground truth grid at the top shows which tiles are safe and which ones are unsafe. Safe tiles are always safe but unsafe tiles have an 80% chance of being quicksand. **Experience**: In the 'experience-only' condition, participants completed 15 navigation trials in each block. Participants first planned a full route to the goal and then watched the agent reveal which tiles in that path were sand or quicksand. Over successive trials (gray arrows) we expect participants to encounter less quicksand. **Hypothetical**: In the 'experience + hypothetical' condition, participants completed a hypothetical trial after each experience trial (blue arrows). They planned a route to the goal from a different starting location but without receiving any feedback. **Counterfactual**: In the 'experience + counterfactual' condition, participants completed a counterfactual trial after each experience trial (red arrows). Participants saw the outcome of the previous experience and planned an alternative route from one of the tiles they had stepped on.

soning does not always enhance learning. We discuss the implications of this work and suggest boundary conditions for learning from counterfactual simulations.

Study environment

We designed an environment where success depends on learning the environment's underlying probabilistic structure.[1] Participants navigate an agent from a start to a goal position in an 8×3 gridworld (called an *experience trial*). Each tile is either *safe* or *unsafe*, which affects how likely it appears as *sand* or *quicksand* on a given trial. Specifically, **safe** tiles always appear as sand, while **unsafe** tiles become quicksand with 80% probability. For an optimal learner, this means that a single observation of quicksand indicates that the tile is unsafe. Stepping on a quicksand tile slows down the agent, impeding progress towards the goal. The objective is to reach the goal while minimizing quicksand encounters. Before observing the agent follow its path, participants must plan their route. Because the sand/quicksand state of a tile is only revealed when stepped on, effective navigation requires learning which tiles are unsafe and which ones are safe.

[1]Materials, data, analyses, preregistrations and links to experiment demos can be found at https://github.com/cicl-stanford/counterfactual_learning_cogsci2025.

Gridworld selection We selected gridworlds where successfully learning the environment would lead to a substantial improvement compared to a baseline which assumes that each tile has a 50% chance of being safe or unsafe. While the gridworlds were selected to be particularly rewarding if participants adopted a planning-based strategy using the probabilistic structure of the environment, they do not effectively discourage using less generalizable forms of task representation like path memorization. To discourage against learning simpler policies like rote memorization of a single path, we varied participants' start and goal locations. Agents start at a random location on the first column of the grid, and were asked to navigate to a goal location randomly selected from the tiles in the last column. We also added randomly placed road blocks on each trial that prevented movements to that spot.

Study 1: Learning from hypothetical vs. counterfactual simulations

The goal of our first study was to explore whether engaging in counterfactual simulation in a memory-intensive gridworld task would improve learning relative to two baselines: an experience-only condition and a hypothetical simulation condition. Including the hypothetical condition allowed us to

distinguish any specific benefits of counterfactual reflection – where participants mentally revise past actions to achieve better outcomes – from general effects of engaging in mental simulation more broadly (Gerstenberg, 2022).

Methods

Participants and design Participants were recruited via Prolific. A total of 60 participants (*age*: median = 33, range = 19-64; *gender*: 34 female, 22 male, 3 non-binary, 1 other; *race*: 6 black/African American, 6 Asian, 1 American Indian/Alaska Native, 7 multiracial, 40 white) completed the pilot study, equally divided between three between-subjects conditions: *experience-only*, *experience + hypothetical*, and *experience + counterfactual*.

All participants were native English speakers residing in the US. Participants were paid $6 for an estimated 30 minutes to complete the study (*mean completion time*: 33.8 mins). In addition, participants were awarded a bonus of up to $2.40 depending on their performance on the task (*mean bonus*: $1.00). Both experiments were programmed using jsPsych (De Leeuw, 2015). This pilot study followed the Stanford University IRB protocol and all participants provided informed consent.

Procedure Participants were instructed that their task was to help an agent navigate deserts filled with quicksand. They were told that quicksand is a hazard, that weather changes whether a tile is quicksand or not on each day (trial), and that regular weather patterns make some tiles always safe and others unsafe, where unsafe tiles were quicksand 80% of the time. After reviewing the instructions, participants had to successfully complete comprehension checks to proceed to the main experiment. Otherwise, they were redirected to read the instructions again. The experiment consisted of four blocks, each featuring a different gridworld environment (see the ground truth in Figure 1 as an example).

In the *learning* phase, participants completed 15 *experience* trials, guiding the agent from a starting location to a goal tile. Their goal was to construct efficient paths that minimized encounters with quicksand. During each *experience* trial, participants were shown the agent at the starting location and instructed to construct a path by clicking on successive tiles until reaching the goal. Participants watched the agent follow the path, revealing whether each tile that it stepped on was sand or quicksand. The agent moved through quicksand tiles considerably slower than through safe tiles. Participants were rewarded $0.04 for each experience trial, with a deduction of $0.01 per quicksand tile encountered, down to a minimum of $0.00.

Participants were assigned to one of three between-subjects conditions. In the *experience-only* condition, participants proceeded directly to the next experience trial after receiving their feedback and bonus. In the *experience + hypothetical* condition, participants engaged in an additional *hypothetical* trial after each experience trial. They were instructed to "engage in a thought experiment and consider what to do

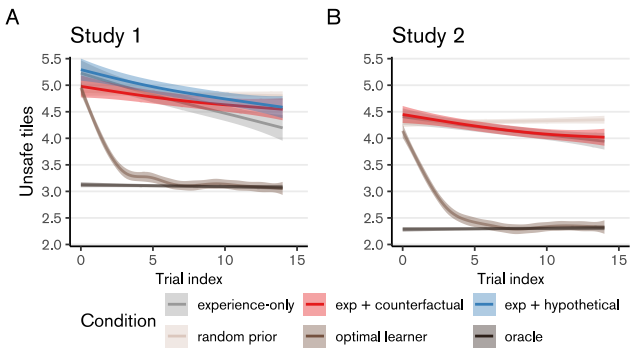

Figure 2: **Experience trial performance**. **A** Study 1 performance (lower values indicate better performance). The x-axis shows the trial index within each block. Participants encountered fewer unsafe tiles as they progressed through the trials in the block. We did not find sufficient evidence to conclude any differences in performance across our experimental conditions. Human performance is contrasted with three model baselines: the performance of an oracle who plans optimally with respect to the ground truth, one with random priors, and one who learns through experience with Bayesian updating. **B** Study 2 performance. Participants similarly encountered fewer unsafe tiles as they gained more experience. Asking participants to simulate counterfactuals did not lead to better performance.

if the agent started at other points." They did this by planning a new path to the goal, imagining it was a new day (i.e., that tiles that were quicksand in the previous experience trial would not necessarily be quicksand in the hypothetical trial again). The start location was randomly chosen from the set of available tiles a distance d from the goal. In the *experience + counterfactual* condition, participants instead completed a *counterfactual* trial after each experience trial. In counterfactual trials, they were prompted to "take some time to reflect on the plan you made and consider better alternatives." Participants started at a location on their original path at a distance d from the goal, where the distances were mirrored across both hypothetical and counterfactual conditions. Unlike in the hypothetical trials, participants imagined the scenario as going back in time on the *same* day, meaning that any tiles revealed in the experience trial remained revealed. Importantly, in both the hypothetical and counterfactual condition, participants received no feedback about their simulated paths. The agent did not walk on these paths, so they did not learn what would happen in the hypothetical condition, or what would have happened in the counterfactual condition.

At the end of the experiment, participants completed a post-experiment questionnaire, where we asked for demographic information, details about their input device, subjective effort and difficulty ratings, and open-ended feedback.

Results

Performance on experienced paths Figure 2 shows that participants' performance improved with experience. Over successive trials within the same gridworld, they encountered fewer unsafe tiles, suggesting that they used past observations to their plan. For comparison, we also plot the performance of three models that make optimal plans based on different beliefs about which tiles are unsafe. The *random prior* model assumes that each tile has approximately 50% chance of being unsafe. The *oracle* model has perfect knowledge of the environmentś probabilistic structure. The *optimal learner* model updates its beliefs over trials using Bayes' rule.

Did counterfactual simulation improve participants' ability to navigate the environment? To assess whether participants in the counterfactual condition avoided more unsafe tiles, we fit a Bayesian mixed-effects model with fixed effects for condition, trial number (within block), and their interaction. We included random intercepts for grid worlds and participants, and random slopes for trial number within participants. There was no credible difference between participants' performance in the counterfactual and the experience condition ($\beta = -0.01$, 95% CrI $[-0.066, 0.043]$). Moreover, the interaction between the counterfactual condition and trial number was similarly not credible ($\beta = 0.009$, 95% CrI $[-0.019, 0.001]$), suggesting learning trajectories were consistent across conditions.

We also examined whether any potential effects of counterfactual simulation could be distinguished from the general benefits of additional reasoning. Participants in the hypothetical condition completed the same task structure as those in the counterfactual condition but without conditioning on prior observations (see task procedure). We tested whether the experience trial paths made by participants assigned to the counterfactual condition contained fewer unsafe tiles than those assigned to the hypothetical condition using a similar Bayesian mixed-effects model using the same fixed and random effects structure, except that the condition parameter contained only the hypothetical and counterfactual conditions. The results similarly indicated no evidence suggesting that paths from the counterfactual condition were safer than those from the hypothetical condition ($\beta = -0.06$, 95% CrI $[-0.13, 0.01]$). In particular, participants assigned to the hypothetical condition encountered an estimated 4.82 unsafe tiles (95% CrI $[3.13, 6.41]$) in their experience trial paths, while those assigned to the counterfactual condition encountered an estimate of 4.67 unsafe tiles (95% CrI $[3.17, 6.36]$).

Performance on simulated paths One possibility why participants may not have benefited from simulating alternatives is that they didn't take that aspect of the task seriously. However, this was not the case. Figure 3a shows the number of unsafe tiles that the agent encountered in the simulated trials in the hypothetical and counterfactual condition. To compare participants' performance, we ran a Bayesian mixed-effects model (using a Poisson linking function for count data) with random intercepts for gridworlds and random intercepts and slopes for participants. Counterfactual paths contained fewer

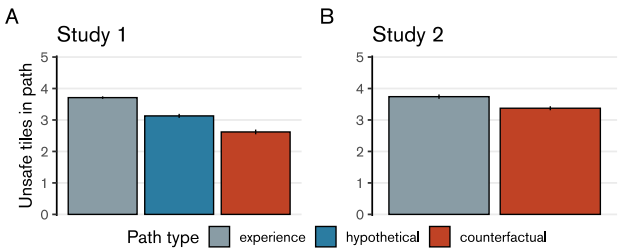

Figure 3: **Simulation trial performance**. **A** Study 1 results. Counterfactual paths had fewer unsafe tiles than hypothetical paths. These two measures are comparable because the distance of the starting location to the goal in the simulation trials was matched in both conditions. Experience trials are expected to have more unsafe tiles because they are longer than simulation trials in Study 1. **B** Study 2 results. Because experience trial and counterfactual paths share start and goal positions, they are directly comparable. Participants' counterfactual paths contain fewer unsafe tiles than their experienced paths. *Note*: Error bars in all figures show 95% bootstrapped confidence intervals.

unsafe tiles than hypothetical paths ($\beta = 0.178$, 95% CrI $[0.129, 0.225]$). To put this into perspective, a model that plans optimally but has a prior belief that all tiles have a 0.5 probability of being unsafe would encounter an estimated 3.49 unsafe tiles, relative to participants' estimated 3.33 tiles in the hypothetical condition.

Discussion

Taken together, these findings suggest that counterfactual (or hypothetical) simulation alone may not be sufficient to yield substantial improvements in planning outcomes within this task environment. To better distinguish between learning and acting, we conducted a follow-up study that directly assessed participants' knowledge of the environment.

Study 2: Directly probing learning

In Study 1, we indirectly assessed learning through participants' *performance*. While better learning should generally translate to better performance, it is possible that counterfactual simulation enhances knowledge in ways that do not immediately manifest in navigation performance – for example, by helping participants identify which paths are worth *exploring* in future trials.

To address this limitation, we probed participants' learned representations more directly this time. After completing the learning phase, participants engaged in an *exam* where they reported which tiles they believed were safe or unsafe.

Methods

Participants and design Participants were recruited via Prolific. A total of 98 participants (*age*: median = 36, range = 21-72; *gender*: 56 female, 39 male, 3 non-binary; *race*: 12 black/African American, 9 Asian, 7 multiracial, 68

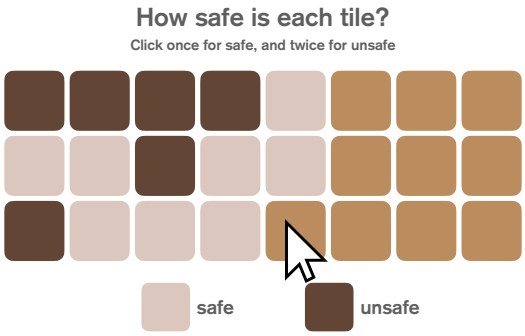

Figure 4: **Exam trial**. In Study 2, after completing the learning phase, participants were asked to report how safe they believed each tile in the grid was. Participants did this by clicking on each tile in the grid either once (for safe) or twice (for unsafe), or vice versa.

white) completed the experiment, divided evenly among two between-subjects conditions: *experience-only* and *experience + counterfactual*.

All participants were native English speakers residing in the US. Participants were paid $3.5 for an estimated 15 minutes to complete the study (*mean completion time*: 24.9 mins. Additionally, participants were awarded a bonus of up to $1.20 depending on their performance on the task (*mean bonus*: $0.39).

Task procedure The experiment consisted of two blocks. Within each block, participants completed a *learning* phase followed by an *exam* phase. In the *learning* phase, participants completed 15 experience trials, guiding an agent from start to goal location. The experience and counterfactual simulation trials were identical to those in Study 1, with two modifications: (1) participants encountered only one road block per trial instead of two, and (2) in the *counterfactual* condition, participants started from the same location as in the experience trial rather than from a designated spot on their previous path. We did not include a hypothetical condition in this study.

Figure 4 shows an *exam* trial. Participants viewed a grid of the environment and had to classify each tile as safe or unsafe by clicking on it. They had to click all tiles to proceed, and could click multiple times to change the colors back and forth. This exam trial directly tests participants' understanding of the environment.

Results

Performance on experienced paths To assess whether participants in the counterfactual condition improved their ability to avoid hazards, we fit a Bayesian mixed-effects model with a Poisson linking function to the number of unsafe tiles encountered in experience trials. The model included fixed effects for condition (*counterfactual* = 1, *observation* = -1), trial number (within block), and their interaction. Random intercepts for gridworlds and participants, as well as random slopes for trial number within participants, were also included. The results provided no credible evidence that counterfactual reflection improved planning: there was no effect of condition ($\beta = -0.00$, 95% CrI $[-0.05, 0.04]$). Performance increased across trials, as participant encountered fewer hazardous tiles ($\beta = -0.01$, 95% CrI $[-0.01, -0.005]$). The interaction between condition and trial number was not credible ($\beta = 0.00$, 95% CrI $[-0.01, 0.007]$).

Performance on simulated paths Participants in the counterfactual condition engaged meaningfully in counterfactual reflection: counterfactual paths contained fewer unsafe tiles than those taken in experience trials (see Figure 3).

A Bayesian mixed-effects model using a Poisson link function to model the distribution of unsafe tiles, with random intercepts for gridworlds and participants, as well as random slopes for counterfactual paths within participants revealed a credible reduction in the number of unsafe tiles counterfactual paths would have encountered ($\beta = 0.13$, 95% CrI $[0.10, 0.17]$). This indicates that participants engaged with counterfactual reflection meaningfully.

Performance on exam To test whether counterfactual reflection led to a more accurate understanding of the environment, we analyzed participants' exam trial accuracy, which measured how well participants identified tiles as safe or unsafe (see Figure 5). A Bayesian mixed-effects model was fit to the number of tiles correctly labeled as safe or unsafe, with fixed effects for condition and random intercepts for gridworlds and participants. We found no credible main effect of condition ($\beta = 0.24$ tiles, 95% CrI $[-1.01, 1.48]$), suggesting that participants in the counterfactual condition did not demonstrate greater accuracy than those in the observation condition. In other words, there is little evidence that counterfactual simulation enhanced participants' understanding of the gridworld environments.

While we found little evidence that counterfactual simulation improved participants' explicit knowledge of which tiles were safe, this does not necessarily mean that participants in the counterfactual condition did not acquire a more *useful* model of the environment. Specifically, they may have learned a representation that was better adapted for locations relevant to navigation. To the extent that participants were bottlenecked by memory limitations, one might ask whether counterfactual simulation plays a *focusing* role on people's mental representations – heightening attention to performance-critical tiles and reinforcing learning where it matters most.

To test this, we computed the average number of unsafe tiles a rational planner would have encountered under the probability distribution specified by participants' exam responses in the trials they experienced within the same block (Figure 5b). This characterizes participants' exam responses in a task-relevant manner because it weighs the practical consequences of errors: inaccuracies on frequently traveled tiles would have greater impact on navigation performance than errors on less relevant tiles. We fit a Bayesian mixed-effects model with a Gamma linking function to the average number

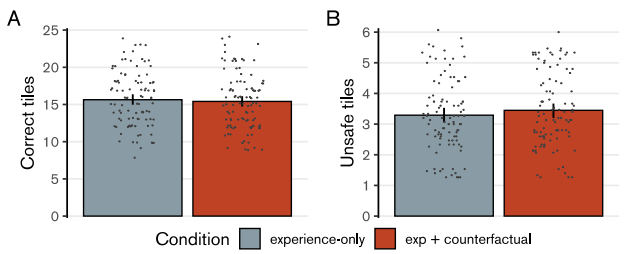

Figure 5: **Exam trial results**. **A** The number of tiles participants guessed correctly in the exam trial. Participants did not guess more tiles correct in the counterfactual condition than in the experience condition. **B** How many unsafe tiles a rational planner would encounter when planning with the participants' exam responses. We did not find evidence that the paths a rational planner would take when using exam responses from the counterfactual condition are safer than those using responses from the experience condition.

of simulated unsafe tiles, with fixed effects for condition and random intercepts for gridworlds and participants. We found no credible main effect of condition ($\beta = -0.04$, 95% CrI $[-0.13, 0.05]$), suggesting that counterfactual simulation did not lead to more task-relevant representations of the environment compared to the experience-only condition.

Discussion

While participants in both conditions improved their navigation performance over time, counterfactual reflection did not provide a measurable advantage in participants' ability to avoid unsafe tiles (Figure 2b). Taken together, these findings suggest that counterfactual simulation alone may not be sufficient to yield substantial improvements in planning outcomes or learning rates within this task environment.

General discussion

The present study examined whether counterfactual simulation enhances learning. We explored this in a navigation task and hypothesized that engaging in counterfactual simulations would lead to improved planning and a more accurate mental representation of the environment. However, our findings did not support these hypotheses. Compared to participants in the 'experience-only' condition, participants in the 'experience + counterfactual simulation' condition did not perform better on navigation trials, nor did they learn the environment better. This was the case even though participants took the simulation part seriously, as evidenced by their performance on the simulated trials.

Why did counterfactual simulation not improve learning? One possible explanation for why counterfactual simulation was not beneficial for learning is that *forcing* participants to engage in mental simulation may have introduced cognitive costs that outweighed its potential benefits. Prior work suggests that people regulate their use of mental simulation strategically, adjusting their cognitive effort based on task demands and expected utility (Hamrick et al., 2015). In other words, counterfactual simulation may indeed facilitate learning in some contexts, but our task may not have made it sufficiently valuable relative to its cognitive cost. Our implementation of counterfactual simulation was highly *explicit* and scaffolded by the task structure – participants had direct visual access to prior experiences rather than relying on their memories of the past. Future work could investigate whether counterfactual simulation is more beneficial in contexts where it arises spontaneously rather than being externally imposed. Moreover, it may be that counterfactual simulations rely on offline processes, rather than through explicit actions. That is, it could be that participants are implicitly engaging in counterfactual simulation in *both* experience-only and counterfactual conditions. In future work, we may consider tracking participants' eye movements to test this possibility (Gerstenberg et al., 2017).

Another possibility is that counterfactual simulation was simply the wrong form of "thought intervention" for this task. Perhaps *simulation* is only helpful for model learning to the extent that it makes participants spend more cognitive resources towards encoding the experienced information (Lefebvre et al., 2024). Future work could directly test this by using *explicit attention manipulations* (e.g., prompting participants to recall specific tile probabilities).

Another possibility is that the task was structured in a way that limited the potential benefits of counterfactual reasoning. While participants had to estimate latent environmental parameters, the optimal action policy once unsafe tiles were identified – was relatively simple and available even without engaging in counterfactual reasoning. That is, participants could improve their performance simply by avoiding previously encountered unsafe tiles, without needing to mentally simulate alternatives. This raises the possibility that counterfactual simulation may primarily support learning when it helps refine complex action policies or generalize beyond directly observed experiences.

Relatedly, prior work suggests that counterfactual reasoning may be most effective when learners already have a reasonably well-calibrated model of the world. Indeed, computational work has shown that attempting to learn a policy using simulated experiences from a model that is mismatched with the actual environment can lead to poor generalization and systematic failures (Jiang & Li, 2016; Talvitie, 2014). If counterfactual simulation serves as a stand-in for real experience, as some work suggests (Anderson, 1983; Kappes & Morewedge, 2016), then applying it in a setting where participants' initial models are highly inaccurate may do more harm than good—introducing noise rather than facilitating learning. Future research could investigate whether the effectiveness of counterfactual simulation depends on the learner's prior knowledge, and whether its benefits emerge only when their mental model is already somewhat aligned with reality.

References

Anderson, C. A. (1983). Imagination and expectation: The effect of imagining behavioral scripts on personal influences. *Journal of personality and social psychology*, *45*(2), 293.

Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 135–157.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145–182.

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, *43*, e28.

De Leeuw, J. R. (2015). Jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*, 1–12.

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

Gendler, T. S. (1998). Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science*, *49*(3), 397–424.

Gerstenberg, T. (2022). What would have happened? counterfactuals, hypotheticals and causal judgements. *Philosophical Transactions of the Royal Society B*, *377*(1866), 20210339.

Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*, *28*(10), 924–936.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological science*, *28*(12), 1731–1744.

Hamrick, J. B., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? the amount of mental simulation tracks uncertainty in the outcome. *CogSci*.

Jiang, N., & Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. *International conference on machine learning*, 652–661.

Kappes, H. B., & Morewedge, C. K. (2016). Mental simulation as substitute for experience. *Social and Personality Psychology Compass*, *10*(7), 405–420.

Lefebvre, G., Esmaily, J., Rezazadeh, Z., & Bahrami, B. (2024). The temporal dynamics of attentional allocation during counterfactual learning [https://osf.io/preprints/psyarxiv/2y4nv$_v$1]. *PsyArXiv*.

Lombrozo, T. (2024). Learning by thinking in natural and artificial minds. *Trends in Cognitive Sciences*.

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.

Norton, J. (1991). Thought experiments in einstein's work. *Thought experiments in science and philosophy*, *129*.

Parikh, N., De Brigard, F., & LaBar, K. S. (2022). The efficacy of downward counterfactual thinking for regulating emotional memories in anxious individuals. *Frontiers in psychology*, *12*, 712066.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, *121*(1), 133–148.

Roese, N. J. (1994). The functional basis of counterfactual thinking. *Journal of personality and Social Psychology*, *66*(5), 805.

Roese, N. J., & Olson, J. M. (1993). The structure of counterfactual thought. *Personality and social psychology bulletin*, *19*(3), 312–319.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science*, *26*(5), 521–562.

Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 116.

Smallman, R., & McCulloch, K. C. (2012). Learning from yesterday's mistakes to fix tomorrow's problems: When functional counterfactual thinking and psychological distance collide. *European Journal of Social Psychology*, *42*(3), 383–390.

Talvitie, E. (2014). Model regularization for stable sample rollouts. *UAI*, 780–789.