Computational basis of hierarchical and counterfactual information processing

Mahdi Ramadan[1*], Cheng Tang[1*], Nicholas Watters[1], Mehrdad Jazayeri[1]
* These authors contributed equally to this work as co-first authors, they are listed by alphabetical order of their surname
[1] Department of Brain & Cognitive Sciences, McGovern Institute for Brain Research, MIT
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
Correspondence: mjaz@mit.edu

Author contributions
M.R. and M.J. conceived the study. M.R. collected the human psychophysics data and performed all the analyses. C.T., N.W., M.R. and M.J. conceived the RNN architecture. C.T. and N.W. implemented the RNN code. C.T. performed all RNN analyses. M.R., C.T., N.W. and M.J. wrote the manuscript. M.J. supervised the project.

Competing interest
The authors declare no competing interest.

Data availability
The data used to generate the associated figures will be made available on a public repository after peer-review publication.

Code availability
The code used to generate the associated figures will be made available on a public repository after peer-review publication.

Acknowledgments
M.R. is supported by Lisa K. Yang ICoN fellowship. N. W. is supported by the National Science Foundation Graduate Research Fellowship Program. M.J. is supported by the Simons Foundation, the McKnight Foundation, and the McGovern Institute.

Abstract

Cognitive theories attribute humans' unparalleled capacity in solving complex multistage decision problems to distinctive hierarchical and counterfactual reasoning strategies. Here, we used a combination of human psychophysics and behaviorally-constrained neural network modeling to understand the computational basis of these cognitive strategies. We first developed a multi-stage decision-making task that humans solve using a combination of hierarchical and counterfactual processing. We then used a series of hypothesis-driven behavioral experiments to systematically dissect the potential computational constraints that underlie these strategies. One experiment revealed that humans have limited capacity for parallel processing. Another indicated that counterfactuals do not fully compensate for this limitation because of working memory limits. A third experiment revealed that the degree to which humans use counterfactuals depends on the fidelity of their working memory. Next, we asked whether the strategies humans adopt are computationally rational; i.e., optimal under these constraints. To do so, we analyzed the behavior of a battery of task-optimized recurrent neural networks (RNNs) that were subjected to one or more of these constraints. Remarkably, only RNNs that were subjected to all these constraints behaved similarly to humans. Further analysis of the RNNs revealed that what cognitive theories posit as distinctive strategies such as hierarchical and counterfactual are subdivisions in a continuum of computationally rational solutions that includes optimal, counterfactual, postdictive, and hierarchical.

Introduction

The human brain finds solutions to complex multistage decision problems that are far more flexible than those learned by artificial systems. Cognitive theories attribute this flexibility to specific algorithms such as hierarchical information processing and counterfactual reasoning (Huys et al., 2015; Lake et al., 2017; Tenenbaum et al., 2011; Zylberberg, 2021). Counterfactual reasoning is an important building block of our mental landscape enabling us to imagine alternative accounts of our prior experiences (Gerstenberg et al., 2020; Pearl, 2009; Van Hoeck et al., 2015). A familiar scenario is when we feel the need to revisit our assumptions due to an unexpected turn of events, for example, an unexpected fork in the road while driving, an abrupt emotional reaction from a friend, or a twist in the storyline of a book or movie. These situations compel us to review past events and assumptions and look for alternative interpretations that could provide a plausible explanation for unexpected observations. Given the centrality of counterfactuals in human cognition, it behooves us to understand the computational underpinnings of this strategy.

Humans commonly rely on counterfactuals when facing decision trees with a hierarchy of if-then scenarios leading to different outcomes (Buchsbaum et al., 2012; Gerstenberg, 2022; Lucas & Kemp, 2015; Sloman & Hagmayer, 2006; Zylberberg et al., 2018). However, what is puzzling is that decision trees do not inherently require computing counterfactuals. On the contrary, the optimal strategy for solving a decision tree is to use each observation to update the posterior belief over all states within the tree (Kuperwajs & Ma, 2021; Purcell & Kiani, 2016; Zheng et al., 2022). However, for sufficiently complex decision trees, an optimal posterior updating strategy can become extremely demanding and sometimes intractable (Gershman et al., 2015; Van Opheusden et al., 2017). Curiously, these are also the conditions for which humans typically use counterfactuals. These considerations raise several interrelated questions about counterfactual reasoning. Is this an *ex-post* strategy that humans use as a crutch for updating the posterior belief in complex decision trees? Do counterfactuals introduce sub-optimalities in behavior? If so, what are they, and do humans adopt a computationally rational approach to use counterfactuals effectively?

To answer these questions we developed a moderately challenging multi-stage decision-making task for evaluating the computational basis of cognitive strategies in humans. Briefly, participants had to infer the position of an invisible moving ball within an H-shaped maze using partial and uncertain cues about its position (Fig. 1a). Comparing participants' behavior to inference models implementing various cognitive algorithms, we found that humans use a hierarchical strategy to solve the task sequentially, and when uncertain, revise their decisions by considering counterfactuals. Additionally, participants' reaction time profiles and eye movements provided evidence that they relied on counterfactuals to revise their decisions.

Next, we used a set of experiments to reverse engineer the computational constraints from which these algorithms derive. Experiment 1 indicated that the hierarchical strategy results from a bottleneck associated with processing parallel streams of evidence. Experiment 2 indicated that compensatory counterfactuals are imperfect due to working memory limits. Experiment 3 indicated that humans are computationally rational in that the degree to which they rely on counterfactuals depends on the fidelity of their working memory.

Next, we used a modeling approach to test the importance of these computational constraints on adopting a counterfactual strategy. We trained multiple artificial recurrent neural network (RNN) models to perform

the same task as humans and subjected each to different subsets of those constraints. Unconstrained models adopted an optimal strategy that deviated significantly from the human participants' behavior. The addition of constraints altered the model's behavior. For example, a processing bottleneck constraint shifted the model's responses toward a hierarchical strategy and working memory limits impact the degree to which the model relied on counterfactuals. Remarkably, the model that most accurately emulated humans' behavioral responses patterns was the one that was subjected to all the constraints inferred from human's behavior. Finally, parametric analysis of the models revealed that distinct cognitive algorithms such as optimal, counterfactual, postdictive, and hierarchical may be more accurately characterized as subdivisions in a continuum that neural systems may adopt depending on the task and computational constraints.

Results

The H-Maze Task

On each trial, a ball approaches the horizontal segment of an H-shaped maze from above. Upon reaching the maze, the ball becomes invisible and moves along either the left or right horizontal arm. Upon reaching the corresponding vertical segment, the ball turns upward or downward along one of the vertical arms, and stops after reaching the corresponding exit. After the ball stops, participants must report the exit point. Participants must infer the ball's trajectory using partial and ambiguous information provided by three brief auditory clicks. The first click occurs when the ball turns into one of the horizontal arms, the second, when it enters one of the vertical arms, and the third, when it reaches the exit. The time between the first and second clicks (t_h) provides information about the horizontal arm the ball enters, and the time between the second and third clicks (t_v), provides information about the subsequent vertical arm. Participants must measure these two time intervals and choose the path whose horizontal and vertical arms were compatible with those measurements. In summary, participants have to infer the correct path from the relative timing of the auditory clicks. Crucially, we varied the lengths of both horizontal and vertical arms on a trial-by-trial basis making it essential for participants to use a flexible dynamic inference strategy (Fig. 1b). Throughout the manuscript, we will denote the length of the horizontal arms by *L* (left arm) and *R* (right arm), and the four vertical arms by *LU* (left-up), *LD* (left-down), *RU* (right-up), *RD* (right-down). We may also express arm lengths in units of time reflecting how long it would take for the moving ball to traverse that arm. When doing so, we will use the corresponding labels t_L (left arm), t_R (right arm), t_{LU} (left-up), t_{LD} (left-down), t_{RU} (right-up), and t_{RD} (right-down).

Analysis of behavior indicated that participants learned the task and were able to use the timing cues to infer the exit point. Average performance across participants improved as a function of the difference between both the horizontal and vertical arms (Fig. 1c). We therefore proceeded with a model-based analysis of behavior to infer the participant's cognitive strategy.

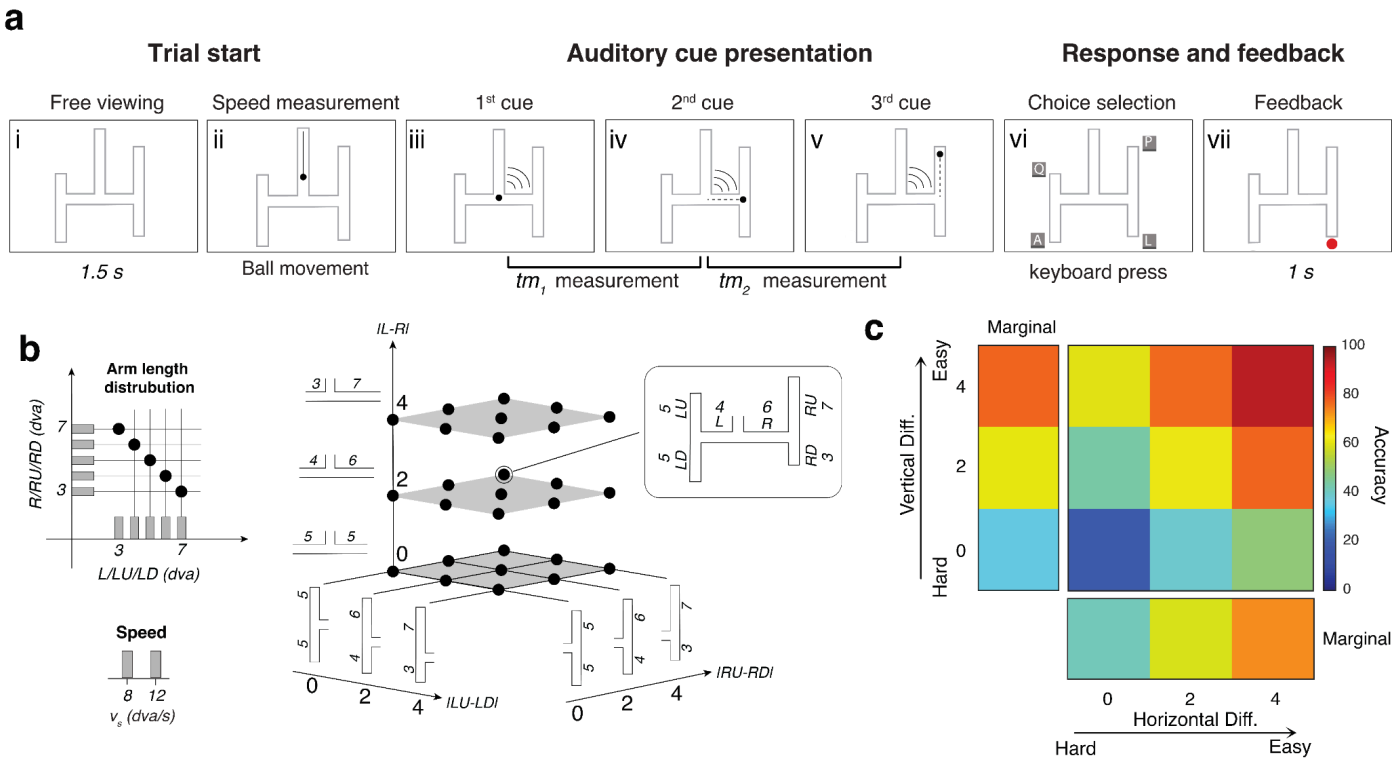


Figure 1. The H-Maze task and performance. (a) Task. (i) Participant is presented with an H-shaped maze and four possible choices. (ii) A ball moves visibly toward the maze with constant speed. (iii) An auditory cue signals when the ball reaches the horizontal segment at which time the ball is made invisible and continues to move leftward or rightward. (iv) A second auditory cue signals when the ball reaches the vertical segment and the ball changes direction upward or downward (tm_1: the time between the first and second cues). (v) A third auditory cue signaled when the ball reached the endpoint (tm_2: the time between the second and third cues). (vi) Participant presses one of four keyboard buttons to indicate their belief about the final exit point. (vii) Participant receives binary visual feedback only at the location of their choice. **(b)** H-maze arm lengths, denoted by *L* (left), *R* (right), *LU* (left-up), *LD* (left-down), *RU* (right-up), and *RD* (right-down), were sampled from a discrete uniform distribution with values 3, 4, 5, 6, and 7 degree visual angle (dva) subject to the constraint that the sum of all arm pairs (*L+R, LU+LD, RU+RD*) was 10 dva (top left). Accordingly, the absolute difference between each arm pair (|*L - R*|, |*LU - LD*|, |*RU - RD*|) could be either 0 (5, 5), 2 (6, 4), or 4 (7, 3) dva (right). The H-maze in the box shows an example H-maze subjected to these constraints. The speed of the ball was either 8 or 12 dva/s, chosen randomly (bottom left). **(c)** Performance averaged across participants. Percentage of correct responses as a function of absolute length difference between the two horizontal arms (|L-R|) and the two vertical arms associated with the correct horizontal arm (|LU-RD| for left and |RU-RD| for right).

Models of different cognitive computational strategies

To quantitatively investigate the strategy participants used for solving the task, we developed a range of models implementing different cognitive strategies with different degrees of optimality. Our first model, which we refer to as the "optimal" model, chooses the exit point based using a maximum-likelihood strategy; it chooses the path whose composition of horizontal and vertical arms is most consistent with the joint distribution of t_h and t_v (Fig. 2a, Top-Left). Our second model implements a hierarchical strategy by breaking down the problem into two hierarchically organized sequential decisions (Fig. 2a, Bottom-Left). It first uses t_h to choose between the horizontal arms, and then uses t_v to choose between the corresponding vertical arms. This strategy is suboptimal because it does not allow the decision about the first arm to benefit from information about the second arm and vice versa. For example, the model may commit to the left horizontal arm based on t_h even if t_v is not consistent with either of the two left vertical arms. Our third model implements a postdictive inference strategy (Fig. 2a, Top-Right). This model functions hierarchically but postpones its first decision until after it has received information about the second arm. More specifically, the model's decision about the horizontal arm is conditioned not only on t_h but also on the degree to which t_v is consistent with the corresponding vertical arms. Note that this differs from the optimal model because the postdictive model makes a horizontal arm choice based on the sum of the likelihoods of the two endpoints for each horizontal arm, hence may be suboptimal when the maximum-likelihood endpoint is on the same horizontal branch as the minimum-likelihood endpoint. Fourth, we developed a hierarchical model that would additionally consider counterfactuals (Fig. 2a, Bottom-Right). This model starts with the two-stage hierarchical strategy. However, if the likelihood of the vertical arms in the second stage is below a certain threshold, it revises its first decision and chooses between the other two vertical arms. In general, counterfactual reasoning is suboptimal because recalling information from memory is imperfect (Ratcliff, 1978; Shadlen & Shohamy, 2016). Note that revisions in counterfactual models differ qualitatively from vacillations that occur during simple decision-making (Peixoto et al., 2021; Resulaj et al., 2009). Vacillations arise from fluctuation of evidence, whereas counterfactuals arise from a deliberate process that uses late evidence to revise earlier commitments.

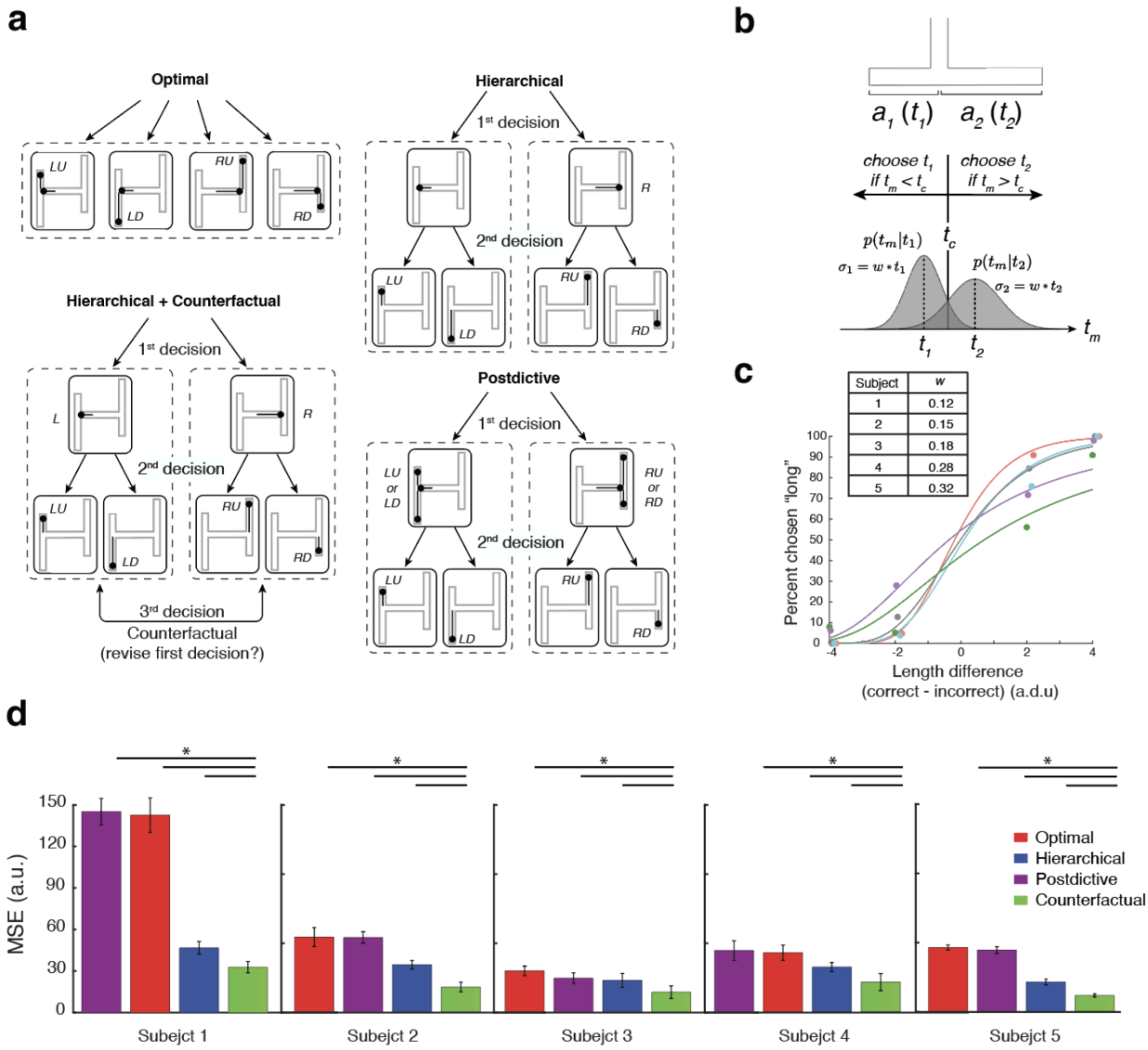


Figure 2. Model-based analysis of behavior. (a) Schematics showing models implementing different computational strategies to solve the H-maze task. Arrows indicate potential decisions and dashed boxes enclose the hypotheses being considered at each stage. The black lines with a ball attached (inside H-maze) indicate which arm segments are being evaluated. Top-Left: The optimal model, which infers the exit based on the joint likelihood of two time intervals measured between consecutive clicks. Top-Right: The hierarchical model, which infers the exit by first choosing between the two horizontal arms using the first and second clicks, and then between the corresponding vertical arms using the second and third clicks. Bottom-Right: The postdictive model, which behaves similarly to the hierarchical model with the key difference that it chooses the horizontal arm postdictively using both time intervals. Bottom-Left: The counterfactual model. This model behaves similarly to the hierarchical model but when the likelihoods for both vertical arms under consideration are below a threshold, the model reevaluates the evidence in the context of the alternative horizontal arm, which may lead to a revision of the decision about the horizontal arm. **(b)** The decision model. Left: The ball takes one of two arms with lengths a_1 and a_2. Since the ball moves at a constant speed, the two arm lengths correspond to two specific time intervals, t_1 and t_2. Right: The participant makes a noisy measurement, t_m, and uses the two conditional probabilities, $p(t_m|t_1)$ and $p(t_m|t_2)$, to choose between the two arms. We modeled the measurement noise as a zero-mean Gaussian distribution whose standard deviation scales with the base interval (t_1

and t_2), with the constant of proportionality w ($\sigma=wt$). The two distributions cross over at the criterion, t_c. The model chooses t_1 when t_m is smaller than t_c and t_2, otherwise. We used this model to measure each animal's timing variability (weber fraction, w). **(c)** Performance and model fit for the T-maze task. Participants' performance is plotted in terms of the percent of trials the longer arm was chosen as a function of the length difference between the correct and incorrect arms (colored circles). Model fits for each participant are shown as lines. Table: fitted w values for each of the five participants. **(d)** The mean squared error computed between models' and participants' choice probability (see choice analysis in Methods), cross-validated five-fold (errorbars: SEM). Pairwise comparisons (lines on the top) indicated that the MSE of the counterfactual model was significantly lower than other models in capturing participants' choice patterns (Left-Tailed binomial test, $p<0.05$).

Humans use hierarchical reasoning and counterfactuals to solve the task

Next we compare participants' performance against the models across different H-Maze geometries. If we assume that models can measure t_h and t_v perfectly, their performance would be trivially at 100%. To make the models' more comparable to humans, we assumed that time measurements are noisy. Following the scalar property of timing in humans (Malapani & Fairhurst, 2002), we modeled noise in timing with a Gaussian distribution whose standard deviation scales with the base interval with a scale factor known as the weber fraction, w. We fit w for each participant on an independent T-Maze control task (Fig. 2b, Top) in which participants had to discriminate between two horizontal arms associated with two time intervals (t_1 and t_2). Under the assumption of scalar Gaussian noise (Fig. 2b, Middle), we computed the maximum likelihood estimate of w based on the participant's performance in the T-maze (Fig. 2b, Bottom). We used the fitted w values for each participant (Fig. 2c, inset) for all subsequent model comparisons. Note that we estimated w for each participant using an independent measure based on the performance in the T-maze task. This approach enabled us to compare participants' behavior in the H-maze task to cognitive models (optimal, hierarchical, etc.) without the need to additionally estimate each participant's timing accuracy, providing a more rigorous approach to model comparison.

To compare the behavior of the models and participants, we proceeded with a more granular analysis that moved beyond a simple accuracy measure, and instead, examined the full pattern of participants' choices (see Methods). We encoded response patterns in terms of four probabilities, one for each possible exit, denoted by p_{LU}, p_{LD}, p_{RU}, and p_{RD}. We measured these probabilities for participants and models across subsets of mazes with the same set of arm length differences and measured the Euclidean distance for choice patterns between models and participants. Finally, we compared the average mean squared error (MSE) between participants and models across all geometries. Results of this comparison indicated that the participants' behavior was most accurately captured by the counterfactual model; that is, the full pattern of participants' choices was most similar to that of the counterfactual model. For all participants, the counterfactual model had the smallest cross-validated MSE compared to the other models (Fig. 2d).

Note that with perfect memory, the counterfactual model behaves identically to the optimal model with the only difference that it evaluates the two sides of the maze sequentially. In reality, counterfactual reasoning may deviate from optimality because of the capacity limitations of working memory needed to recall and process past evidence. Accordingly, our counterfactual model had an additional parameter associated with working memory noise to account for how well participants recall past information. To account for the additional complexity of the counterfactual model, we performed all model comparisons using five-fold cross-validation.

We performed additional analyses to further test whether participants computed counterfactuals. One prominent feature of the counterfactual model is the presence of a third decision process associated with deciding whether to run a counterfactual, and if so subsequently evaluating the alternative arm segments before making a final choice (Fig. 2a). Since each decision increases processing time, an analysis of reaction times could provide additional evidence as to whether and when participants used a counterfactual strategy. An analysis of the counterfactual model indicated that counterfactuals are relied upon more often when the first decision is more difficult; i.e., the difference between the two horizontal arms is small. Accordingly, the hypothesis that participants compute counterfactuals predicts that reaction times would increase systematically with the degree of difficulty of the first decision. Consistent with this prediction, participants' average reaction times increased in a graded fashion with the difficulty of the first decision (Supplementary Fig. 1). This finding provides further evidence that, when uncertain, participants computed counterfactuals.

We also analyzed participants' eye movements to examine the strategy they used to perform the task. Previous studies have shown that the pattern of eye movements during spatial reasoning tasks can reveal behavioral strategy (Gerstenberg et al., 2017; Hayhoe & Ballard, 2005; Levy et al., 2009; Li et al., 2023; Liversedge & Findlay, 2000; Spering, 2022). Accordingly, we reasoned that shifts in gaze position over the H-maze would serve as another clue in deciphering participants' strategy. Specifically, shifts of the eye toward the left or right early in the trial (e.g., after the 1st time interval) would serve as evidence for a hierarchical strategy and subsequent revisions of gaze position from one side to the other would serve as evidence for considering counterfactuals. Eye movements showed evidence for both hierarchical and counterfactual processes (Supplementary Fig. 2). Notably, the frequency of counterfactual eye movements was highest in trials with a difficult 1st decision in which computing counterfactuals was expected to occur most frequently (Supplementary Fig 3). Together, the model-based analysis of choices, the pattern of reaction times across conditions, and gaze positions provide strong converging evidence that participants' strategy for solving the H-maze task is consistent with a combination of hierarchical and counterfactual decisions.

Computational constraints on decision-making

Participants' decision strategy raises important questions about the underlying computational constraints. First, why do participants not process the alternatives in parallel? Second, why are the counterfactual decisions suboptimal? Finally, is the participants' reliance on suboptimal counterfactuals rational (i.e., do participants know when to rely on counterfactuals)? We conducted three additional experiments to address these questions and elucidate the constraints that underlie the participants' strategy.

Experiment 1. We reasoned that hierarchical information processing may be due to an inability to process multiple streams of evidence in parallel, in line with previous findings (Davis & Marcus, 2016; Ludwin-Peery et al., 2021; Smith et al., n.d.; Van Opheusden et al., 2017). To test this possibility, we developed a variant of the task that could only be solved by parallel processing. Participants were presented with four separate arms serving as conduits for four separate balls (Fig. 3a). Upon presentation of an auditory cue, the balls entered the four arms moving at constant speed. If a ball reached an end, it reversed direction and continued to move at the same speed. All balls kept moving until the presentation of a second auditory cue that coincided with one randomly chosen ball reaching an arm's end. At that time, participants were offered

four choices associated with the four arms and had to report the arm in which the ball was at an exit. Importantly, all task parameters including ball speed, arm lengths, and movement times were identical to the H-maze. Moreover, the introduction of direction reversion dissociated the balls' position from time and forced a strategy in which all balls had to be tracked simultaneously, replicating the computational demands of solving the H-maze task using an optimal decision strategy. Participants' performance on this task was slightly higher than chance level, but significantly lower than the performance of an ideal observer model solving experiment 1 using fitted *w* values measured for each participant separately on a control task (Fig. 3a, right; see T-Maze experiment in methods). The relatively low performance associated with parallel processing suggests that participants' choice of a hierarchical strategy is due to a bottleneck associated with processing parallel streams of evidence.

Experiment 2. An ideal observer that relies on counterfactuals should, in principle, be able to attain optimal performance since counterfactuals can be used to evaluate all possibilities sequentially. However, to attain optimal performance, counterfactual processing should be noise-free, or else the performance would drop. Our analysis of participants' behavior indicated that they relied on counterfactuals, but their performance was suboptimal, which led us to hypothesize that computing counterfactuals may incur additional noise. To test this hypothesis directly, we developed a variant of the task using a simpler T-maze geometry with two conditions (Fig. 3b). Condition 1 is analogous to the original H-maze task: the ball enters the maze through a vertical hallway and moves invisibly through the maze while auditory cues indicate the turning points and the time the ball reaches the one of the two exits, and participants are asked to report the exit point. In condition 2, we reversed the order of stimulus presentations such that the T-maze was presented *after* the presentation of the auditory cues. Importantly, in condition 2, participants must solve the task using counterfactuals; i.e., review past evidence from the auditory cues to evaluate alternatives offered by the subsequent presentation of the T-maze. The hypothesis that counterfactual processing is noisy predicts that the performance should be lower in condition 2 compared to matching T-maze geometries in condition 1. Results were consistent with this prediction (Fig. 3b, right). Across participants and T-maze geometries, there was a significant drop in performance when the T-maze was presented after the auditory cues (Left-Tailed T-Test: $p=0.044$). Notably, this difference was dependent on T-maze geometry in a difficulty-dependent manner. In relatively difficult trials in which the difference between the arm lengths was small, there was no significant difference in performance between the two conditions. This is expected since the performance bottleneck in these conditions is dictated by timing uncertainty. In contrast, when the difference in arm lengths was larger, and performance was far above chance, there was a large drop in performance in condition 2 compared to condition 1 (Left-Tailed T-Test: $p = \{0.87, 0.94, 0.47, 0.017, 0.028\}$) revealing the impact of counterfactual processing noise on performance.

Experiment 3. Next, we examined the bounded rationality of participants while using counterfactuals. According to the bounded rationality hypothesis, participants should take into account the degrading effect of the counterfactual processing noise and titrate their reliance on counterfactuals accordingly. Specifically, this hypothesis predicts more reliance on counterfactuals for lower counterfactual processing noise and vice versa. To test this hypothesis, we leveraged the performance variance across participants in experiment 2, which varied widely between ~58% to ~74%. We asked the same participants to perform a variant of the H-maze task in which they could choose whether and when to rely on counterfactuals (Fig. 3c). In this variant, participants were presented with one side of the H-maze (left or right) during the

presentation of the auditory cues. Afterward, participants were given an option to choose between two decision paths. They could either choose one of the two visible exit points as their final answer or ask for the other half of the H-maze to be revealed so that they can counterfactually evaluate the other exit points. Remarkably, the proportion of trials in which participants asked for the other half to be revealed in experiment 3 was strongly correlated with their counterfactual processing noise indexed by performance in experiment 2 (r^2= 0.92, p = 9e-3). Consistent with the bounded rationality hypothesis, this result suggests that participants rely on counterfactuals more when they can process them effectively (Fig. 3c, right).

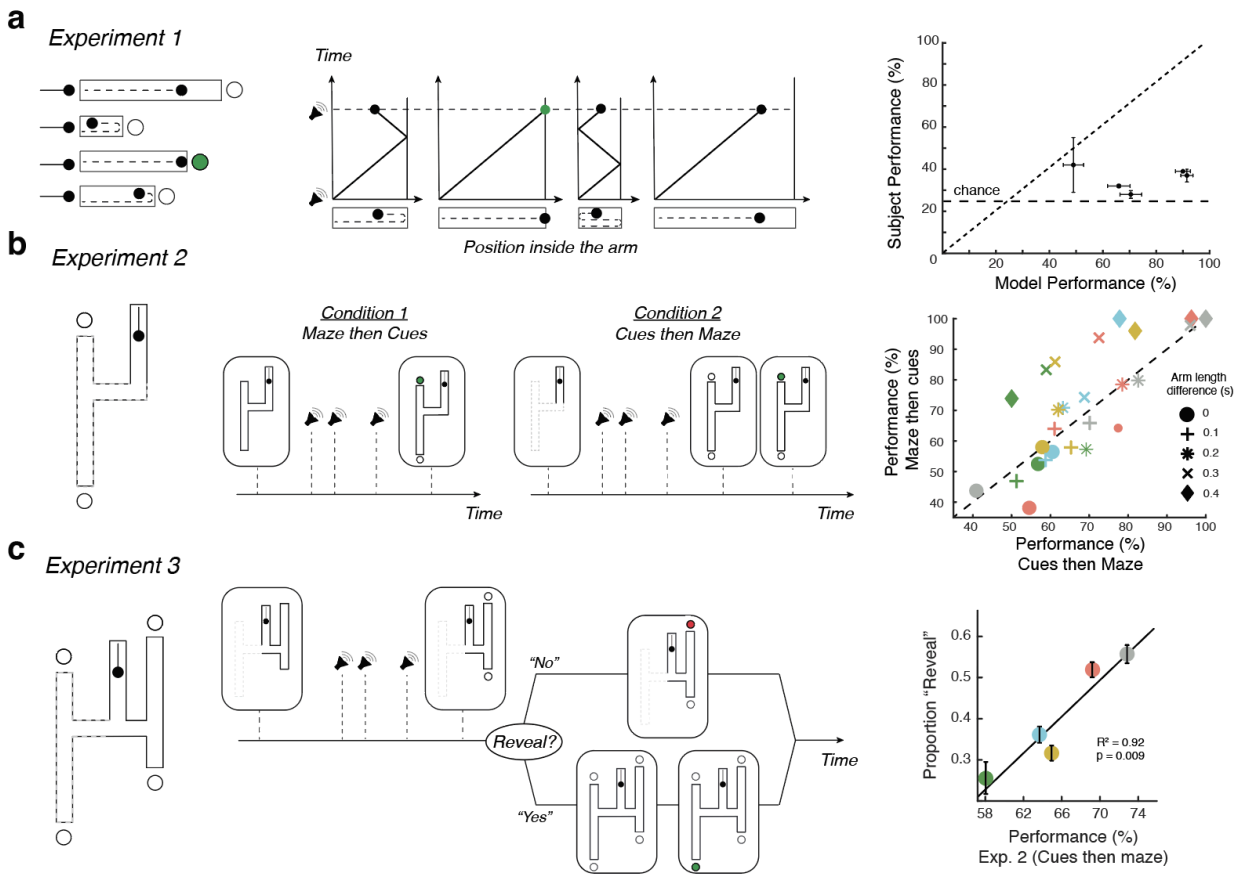


Figure 3. Task variants. (a) Experiment 1. (Left) A variation in H-maze with independent arms. four balls moving at constant and identical speed enter four independent arms simultaneously. Upon entry, an auditory cue is played, and the balls become invisible. The balls continue to move inside their respective arm and reverse direction every time they reach an end. A second auditory cue is played when one of the balls, chosen at random, is at one end of its respective arm (left or right). Subsequently, participants are presented with four options and must choose the arm in which the ball is at its endpoint. Participants receive a binary visual feedback at the location of their choice. (Middle) Balls move back and forth inside their respective arms until the time of the second auditory cue. In the example shown, the correct answer is the second arm from the bottom. (Right) Performance for individual participants compared to a ideal observer models solving experiment 1 using fitted *w* values measured separately for each participant on a control task (horizontal black dashed line: chance-level performance; diagonal black dashed line: unity; vertical error bars: participant S.E.M over sessions; horizontal error bars: model S.E.M). Performance was below the level expected by an ideal observer (Two-Tailed T-Test: $p<0.01$ for all subjects) . **(b)** Experiment 2. (Left) A variation of H-maze with only the left half. (Middle) The experiment includes two conditions. In condition 1, the time course of the trial is identical to the original H-maze: stimulus, then the three auditory cues, followed by decision and feedback. In condition 2, participants first hear the three cues and then see the left half of the H-maze, followed by decision and feedback. (Right) Performance in condition 2 compared to condition 1 (colors: different subjects, symbols: maze geometries; diagonal: unity line). Performance drops for higher working memory load in condition 2 compared to condition 1 for mazes where the difference in arm lengths was large (Left-Tailed T-Test: p = {0.87, 0.94, 0.47, 0.017, 0.028} for arm length differences {0, 0.1, 0.2, 0.3, 0.4, 0.5}). **(c)** Experiment 3. (Left) A variation on H-maze in which half of the H-maze is visible during the presentation of the auditory cues, and participants may choose to reveal the other half afterward. (Middle) After the presentation of one-half of the H-maze (left or right, chosen randomly) and the three auditory cues, participants may either choose between the two available exit points or ask to have the other half revealed and subsequently choose among the four exit points. (Right) The proportion of

trials that participants asked for the other half of the H-maze to be revealed as a function of performance in Condition 2 of Experiment 2 (Cues then maze) across participants (colors). Error bars indicate S.E.M. and the solid black line is the best linear fit to the data ($R^2 = 0.92$, $p = 0.009$).

A neural network model of counterfactual behavior

Behavioral responses in H-maze task variants revealed three characteristics of human decision-making: 1) an attentional bottleneck that limits processing multiple streams of information in parallel, 2) counterfactual processing noise that causes suboptimal performance, and 3) a judicious use of counterfactuals that takes this processing noise into account. Next, we used a modeling approach to ask whether these constraints provide an explanation for the participants' cognitive strategy. Specifically, we asked whether imposing these constraints on recurrent neural network (RNN) models optimized to perform the H-maze task would lead to behavioral response patterns that are compatible with humans.

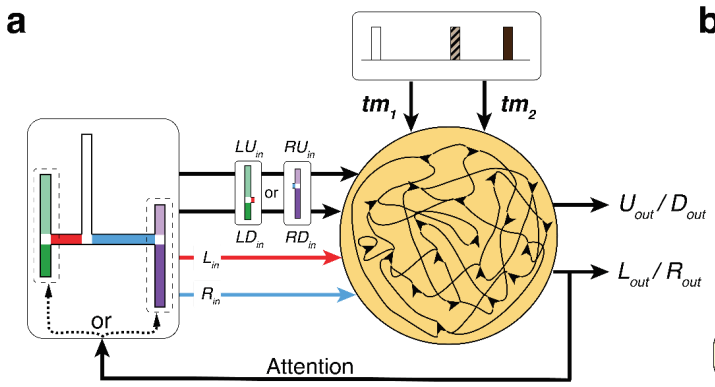
Architecturally, the model receives two types of inputs, has 128 recurrent units, and generates two outputs (Fig. 4a). The first type of input is a 6D vector with values associated with the geometry of the H-maze (L_{in}, R_{in}, LU_{in}, LD_{in}, RU_{in}, and RD_{in}). The second type of input, I_{time}, is a 2D vector specifying the time intervals demarcated by the three auditory cues (tm_1, tm_2). The first output specified the model's horizontal choice (L_{out} versus R_{out}), and the second output, the vertical choice (U_{out} versus D_{out}).

Using this architecture, we developed various task-optimized RNNs subject to different sets of constraints. The base model was trained to choose the correct exit without any additional constraints. In the absence of any timing noise, this model could solve the H-maze perfectly without any errors. However, a noise-free model is unrealistic, as it does not account for the scalar variability that influences human timing behavior. We thus altered I_{time} by adding scalar noise to tm_1 and tm_2 with Weber fraction, w=0.15, which is within the range observed across our participants (Fig. 2c, inset). As expected, introducing timing noise caused the model to make errors. In addition to the base model, we developed various task-optimized RNNs subject to one or more computational constraints inspired by the three experiments on participants' decision-making strategy (Fig. 3). The constraints we considered were:

1. Attentional bottleneck: This constraint modifies training such that the RNN first makes a left/right choice (L_{out} versus R_{out}) and then uses this choice to *attend* to either the left or the right pair of vertical arms. We implement this constraint using a soft attentional gate that biases the input associated with the vertical arms to the left or right side of the H-maze (see Methods). In effect, this constraint forces the RNNs to solve the task hierarchically.
2. Counterfactual processing noise: This constraint modifies training such that the timing information the RNN relies on becomes progressively less reliable for each counterfactual revision. We implement this constraint by adding noise to tm_1 and tm_2 upon each revision (see Methods). This modification is analogous to assuming that the system is subject to counterfactual processing noise.
3. Rationality: This constraint modifies training such that the RNN learns to choose the exit that is most consistent with its noisy time interval measurements (as opposed to the correct exit). We implement this constraint using a cost function that enforces a maximum-likelihood decision strategy (see Methods). This is, in effect, a self-consistency constraint because decisions are made only based on information that is available to the model. Note that the base model violated this rationality constraint since it was optimized with perfect labels that were not available to the learner.

We compared the pattern of responses of each RNN to the pattern expected from the counterfactual model that best captured the behavior of human participants using a state-space analysis (Fig. 4b, see

Methods). The response pattern of the base RNN was most dissimilar to humans. Next was the RNN with the rationality constraint, which fared better than the base RNN but was still quite dissimilar to the counterfactual model. Third was the RNN that was subjected to both the rationality and the attention-bottleneck constraints. Finally, the RNN whose response pattern was most similar to humans was the one that was subjected to all three constraints. We will refer to this RNN subjected to all constraints as RNN_{best}.



b

Attention constraint	Rationality constraint	Counterfactual noise	Latel model fit MSE
x	x	x	$P \overset{**}{<} O \overset{**}{<} C \overset{**}{<} H$
x	√	x	$O \overset{**}{<} P \overset{**}{<} C \overset{**}{<} H$
√	x	x	$P \overset{**}{<} O \overset{**}{<} C \overset{**}{<} H$
√	√	x	$O \sim C \overset{**}{<} P \overset{**}{<} H$
√	x	√	$H \overset{**}{<} C \overset{**}{<} O \overset{**}{<} P$
RNN_{best} √	√	√	$\mathbf{C} \overset{**}{<} H \overset{**}{<} O \overset{**}{<} P$

Figure 4. RNN with different constraints implement different decision strategies. (a). RNN architecture. The RNN receives two types of inputs, a 6D vector specifying the arm lengths of the H-maze (L_{in}, R_{in}, LU_{in}, LD_{in}, RU_{in}, and RD_{in}) and a 2D vector specifying the noisy timing information for interval between the consecutive auditory cues (tm_1 and tm_2). The RNN produces two outputs. The first output specified the model's horizontal choice (L_{out} versus R_{out}), and the second output, the vertical choice (U_{out} versus D_{out}). The horizontal output also serves as an attentional bottleneck forcing the RNN to choose between the vertical arms on one side of the maze, which changes the input it receives. **(b)** MSE computed on the choice probabilities between RNN variants and four latent models using a state-space analysis (see Methods). We trained variants of the RNN using different combinations of attention-bottleneck, rationality and counterfactual noise constraints. The table shows a subset of RNNs trained with ('✓') or without ('✘') certain constraints. Then for each of the RNN variants, we compared its choice pattern with the four latent behavioral models and ranked the models by their similarity to the RNN. Asterisks(**) indicate 'significantly less than' ($p<0.05$, one-tailed t-test, statistics calculated on 50 simulations). The yellow band highlights the RNN with all three constraints present, which is also the only RNN variant fitted best exclusively by the counterfactual model. We refer to this variant as RNN_{best}.

Next, we took a closer look at the RNN_{best} to assess its behavior more carefully (Fig. 5). We first verified that the overall performance of RNN_{best} was similar to human participants and that its responses were sensitive to both horizontal and vertical arm-length differences (Fig. 5a). Next, we analyzed the "attention" signal in RNN_{best} within single trials to assess whether the network relied on a counterfactual strategy. A purely hierarchical strategy would predict that RNN_{best} would choose left or right after tm_1 and one of the corresponding vertical arms after tm_2. In other words, under the hierarchical strategy, RNN_{best} would never switch its attention signal to the alternate horizontal arm after tm_2. In contrast, under the counterfactual strategy, RNN_{best} would occasionally switch its left/right attention depending on arm lengths and tm_2.

We found that RNN_{best} single trial behavior was indeed consistent with the counterfactual strategy. When the difference between horizontal arm lengths was large, RNN_{best} adopted a hierarchical strategy without any revisions (Fig.5b, left and middle). In these conditions, the final decision was usually correct when the difference between the corresponding vertical arms was large (Fig.5b, left), and sometimes wrong when the vertical arms were more similar (Fig.5b, left). Remarkably, when the difference between the horizontal arm lengths was smaller, RNN_{best} would sometimes revise its first left/right decision (as inferred from the attention signal) depending on tm_2 (Fig. 5b, right). This attention-switching behavior increased systematically for smaller differences between the horizontal arms for which the first decision was more uncertain (Fig. 5c). These behavioral characteristics provide direct evidence that RNN_{best} solves the task using a combination of hierarchical and counterfactual strategies.

Finally, we probed the parametric influence of counterfactual processing noise on the behavior of RNN_{best}. Similar to our Experiment 3 in humans, (Fig. 3c), we predicted that the frequency with which RNN_{best} would rely on counterfactuals would decrease for higher levels of counterfactual processing noise. Results were consistent with this prediction; the proportion of attentional switches dropped systematically under higher levels of counterfactual processing noise. One insight gleaned from this analysis was that the network's inferred strategy was qualitatively different for small and large amounts of counterfactual processing noise (σ_{noise}). On one end of the spectrum, when noise levels were small ($\sigma_{noise}<0.18$), the network could rely on counterfactuals without any cost, and thus its behavior was indistinguishable from the optimal model (Fig. 5d). On the other end of the spectrum, when noise levels made counterfactual processing too costly ($\sigma_{noise}>0.8$), the network rarely relied on counterfactuals, and thus its behavior was indistinguishable from the hierarchical model. In other words, computations RNN_{best} were adaptive and exhibited a graceful transition from optimal to counterfactual to hierarchical with the counterfactual strategy emerging as the best solution for intermediate levels of noise ($0.18<\sigma_{noise}<0.8$; $p<0.05$, one-tailed t-test comparing counterfactual to the next best model).

Together, these results indicate that human participants' strategy is consistent with the predictions of the bounded rationality hypothesis governed by an attentional-bottleneck and the magnitude of counterfactual processing noise. Moreover, the continuous transition between strategies implies that the distinct cognitive algorithms could be mere subdivisions in a strategy continuum under the same objective function.

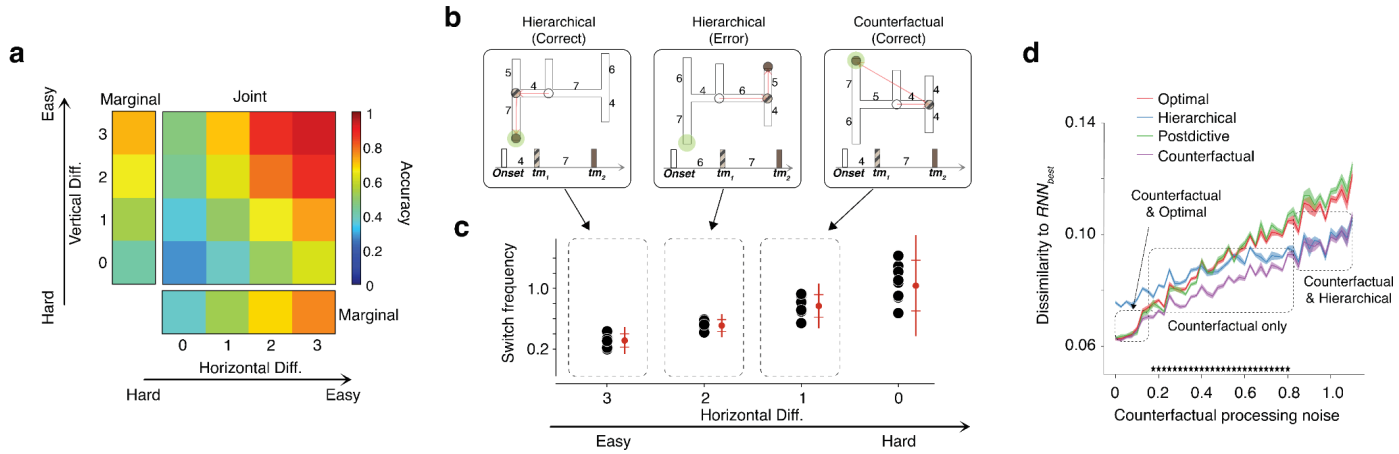


Figure 5. The counterfactual network (RNN_{best}). (a) Behavioral performance. The accuracy of the RNN_{best} improved systematically as a function of the difference between competing horizontal and vertical arms. **(b)** Single-trial behavior of RNN_{best} for three geometries associated with different arm lengths and different measured time intervals (tm_1 and tm_2). Left: an easy trial where tm_1 and tm_2 match the left bottom exit. Middle: tm_1 and tm_2 are inconsistent with all exits; the RNN first chooses rightward based on tm_1 and then upward based on tm_2. Right: tm_1 and tm_2 are inconsistent with all exits; the RNN first chooses rightward based on tm_1, but upon seeing tm_2, switches its attention to left and chooses the left-up as its final choice. The two red arrows trace RNN_{best} decisions after tm_1 and tm_2. The green circle shows the correct answer. See Methods for details. **(c)** The probability of attentional switches as a function of the first choice difficulty (i.e., the difference between the horizontal arms is smaller). Results are sampled from 10 networks. The three trials in **(b)** are samples of the three horizontal differences highlighted in **(c)** (guided by arrows). **(d)** Dissimilarity between the response pattern of RNN_{best} and various cognitive models as a function of counterfactual processing noise (σ_{noise}). The shading shows the standard deviation across 50 simulations. Asterisks indicate the conditions in which RNN_{best} is most similar to the counterfactual model. The dashed blocks show regions with low, intermediate and high σ_{noise}, which specify the cognitive model to which RNN_{best} is most similar to.

Discussion

Humans' cognitive capacity limitations make them unable to find optimal solutions when facing moderately complex problems. Yet, we are quite efficient at finding reasonably good solutions. In cognitive sciences, this ability is viewed through the lens of computationally rational bounded optimality, which posits that humans rely on strategies that are optimal within the bounds of their computational capacity. When facing large decision-trees, one of the most common strategies humans use is to think through a hierarchy of if-then scenarios and, when needed, consider counterfactuals. While numerous studies have found evidence for these strategies (Huys et al., 2015; Lake et al., 2017; Tenenbaum et al., 2011; Zylberberg, 2021), less is known about the computational constraints that motivate these strategies. Here, we used a moderately complex task for which we could generate precise models to predict behavior under optimal, bounded-optimal and suboptimal strategies. Comparing model predictions to behavior, we were able to verify that humans used a combination of hierarchical and counterfactual strategies.

Next, we performed a series of complementary experiments to gain insight into the potential constraints that underlie these strategies. In a first experiment, we verified that parallel processing degrades performance, consistent with previous findings (Davis & Marcus, 2016; Ludwin-Peery et al., 2021; Smith et al., n.d.). This result is consistent with previous work showing humans' inability to process multiple streams of information simultaneously, and the importance of an attentional bottleneck to address this limitation (Dasgupta et al., 2020; Duncan, 2013; Gershman & Burke, 2022; Ramírez-Ruiz & Moreno-Bote, 2022; Wolfe et al., 2006; Zylberberg, 2021). In a second experiment, we quantified performance degradation due to counterfactual processing, which is likely due to capacity limitations of memory-based computations (Beck et al., 2012; Egger & Jazayeri, 2018; Panichello et al., 2019; Ratcliff, 1978; Remington et al., 2018; Sarafyazd & Jazayeri, 2019; Shadlen & Shohamy, 2016; Yu et al., 2020; Zylberberg, 2021). In a third experiment, we made the intriguing observations of lower likelihood of counterfactual reasoning for subjects for whom counterfactuals incurred a larger performance cost. This finding suggests that human's use of counterfactuals is not optimal but computationally rational (Gershman et al., 2015; Piloto et al., 2022). Together, the results of these experiments provided a clear hypothesis for when humans rely on hierarchical and counterfactual processing: hierarchical processing may be used to compensate for performance degradation due to parallel processing and counterfactual processing may be used judiciously when it can improve performance.

However, human experiments alone do not provide a strong test of these hypotheses because we cannot experimentally manipulate cognitive capacity limitations and computational constraints in participants. To address this shortcoming, we developed task-optimized RNNs that were subjected to different computational constraints, and analyzed their behavioral patterns to see whether there might be a direct link between computational constraints and cognitive strategies. This modeling effort proved highly fruitful. It provided clear evidence that the models that were subjected to constraints identified in humans found solutions to the H-Maze task that were consistent with hierarchical and counterfactual processing, similar to humans. Moreover, the parametric sweeps of the model revealed a direct link between working memory limitation and counterfactual processing. Finally, we found that a single model with a fixed architecture may generate behavior consistent with optimal strategy, hierarchical strategy, or counterfactual strategy, depending on the computational constraints and noise-levels. It is therefore conceivable that strategies that are typically considered as distinct within cognitive sciences may be part of a continuum within neural

systems (Donoso et al., 2014). This finding may open new avenues of exchange between adjacent studies of the brain in systems neuroscience and the mind in cognitive sciences.

Materials and Methods

Human participants

Fifteen participants (age: 18 to 65 y, six male and five female) participated in the experiments after giving informed consent. All participants were naive to the purpose of the study, had normal or corrected-to-normal vision, and were paid for their participation. All experiments were approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

Five participants completed the baseline H-Maze task, five participants completed task variant experiment one, and five participants completed task variant experiments two & three. The testing sequence for the two tasks was counterbalanced across participants. In each session, a participant was seated in a dark quiet room and asked to perform the task of interest for ~60 min. For both tasks, stimuli and behavioral contingencies were controlled by an open-source software (MWorks; mworks-project.org/) running on an Apple Macintosh platform.

Tasks

<u>Baseline experiment and naming conventions.</u> The main stimulus in the baseline experiment was a ball moving inside a maze shown from above at a constant speed. The ball entered the maze from the top through a vertical entry hallway, continued moving through the maze at a constant speed, and stopped at one of many possible exit points. In the majority of experiments, the maze was shaped like the letter H, hence the name H-maze. In the H-maze experiment, the vertical entry hallway branched into 2 horizontal arms each of which branched into 2 vertical arms. We will use these subscripts to refer to the length of specific arms (a_L, a_R, a_{LU}, a_{LD}, a_{RU}, a_{RD}) as well as the time it would take the ball to move along those arms (t_L, t_R, t_{LU}, t_{LD}, t_{RU}, t_{RD}). The initial hallways was 7 degree visual angle (dva), and the H-maze arms lengths were sampled from a discrete uniform distribution with values 3, 4, 5, 6, and 7 dva subject to the constraint that the sum of all arm pairs (L+R, LU+LD, RU+RD) was 10 dva (bottom left). Accordingly, the absolute difference between each arm pair (|L - R|, |LU - LD|, |RU - RD|) could be either 0 (5, 5), 2 (6, 4), or 4 (7, 3) dva (right). The speed of the ball was 8 or 12 dva/sec. With these parameters, the shortest and longest arms lengths (3 and 7 dva) were associated with 0.25 and 0.875 sec, respectively. The width of all arms was 0.5 dva.

In the H-maze, the ball (1) started moving downward along the entry hallway until reaching the horizontal segment, (2) turned left or right and continued at the same speed until reaching one of the vertical segments, (3) turned up or down and continued at the same speed until reaching one exit point where it stopped. Throughout the manuscript, we use subscripts *L* (left) and *R* (right) to refer to the left and right arms of the horizontal segment. Moreover, we will use subscripts *LU* (left-up), *LD* (left-down), *RU* (right-up), and *RD* (right-down) to refer to the 4 corresponding vertical arms.

<u>H-maze experiment.</u> The trial structure in the H-maze experiment was as follows: (1) An H-maze was presented for 1.5 s. Together with the maze, four circles marking the four exit points were presented (color: white; diameter = 1 deg) (2) a ball visibly traveled through the initial hallway at one of two randomly sampled speeds. (3) Three identical auditory cues were presented (cue duration: 80ms). The first cue was presented when the ball reached the horizontal segment, the second cue when the ball reached the

corresponding vertical segment, and the third cue when the ball reached the end of the maze. We will refer to the interval between the first and second cues as the first sample interval, denoted by ts_1, and the interval between the second and third cues as the second sample interval, denoted by ts_2. (4) Participants had to make a keyboard choice (q for LU, a for LD, p for RU, and l for RD) to indicate the exit point corresponding to the end position of the ball. The trial would abort if no answer was made within 2500 ms. (5) The circle corresponding to the participant's choice changed color to provide feedback (green if correct, purple if incorrect). (6) Trials were separated by an inter-trial interval of 1.5 sec.

T-maze experiment. The T-maze experiment is identical to the H-maze experiment except that there are vertical arms and the exit points are at the end of the two horizontal arms (*L* and *R*). Accordingly, participants had to report their binary decision about the exit by choosing one of two circles presented at the end of the two horizontal arms. The two arm lengths (a_L, a_R) were sampled from the same distribution as the H-maze. The speed of the ball was also the same as the H-maze task. Therefore the distribution of the time intervals associated with the two arms (t_L, t_R) was also the same as the H-maze task. We performed this experiment to estimate each participant's timing variability parameterized by a weber fraction, *w* (see below).

Task variant 1. Task variant 1 was designed to force participants to implement the optimal model, where the participant updates their beliefs about all four possible ball trajectories simultaneously. To do so, (i) participants were presented with a variant of H-maze where the four possible ball pathways are flattened into four parallel options, and the timing statistics (e.g. total time of occluded ball movement) were kept identical between the H-maze task and task variant 1 ; (ii) after a 1.5 second viewing period, four balls initialized in the same location moved visibly toward the maze with identical and constant speed; (iii) an auditory cue signaled when the balls reached the horizontal arms at which time the balls were made invisible and continued to move rightward. If a ball reaches the left or right end of its arm, the ball reverses direction. Crucially, the 1) nonlinearity introduced by the ball direction reversal dissociated the balls' position from time and thus forcing participants to separately keep track of four ball trajectories separately, and 2) no cue was provided at the time of any ball direction reversal, thus participants cannot solve this task hierarchically or sequentially; (iv) a second auditory cue signaled when one the balls reached the endpoint of its arm. Crucially, the arm segments and the number of direction reversals were chosen so that the three other balls were not at any arm segment endpoint and were still traveling through their respective arm segments. The endpoint of the ball could be on the right or left end of the arm and after a variable number of direction reversals; (v) Participants had to click one of four circles that appeared next to four possible end points of the ball, and participants had to indicate at which arm end pint (vi) participants received a binary visual feedback at the location of choice.

Task variant 2. Task variant 2 is similar to the H-maze experiment except that only one half of the maze is shown to the participants (a_L, a_{LU}, a_{LD} or $a_{R,}$ a_{RU}, a_{RD}), and thus there are only two exit points similar to the T-maze experiment. The occluded ball traveled through the horizontal arm shown and one of the two vertical arms shown. Accordingly, participants had to report their binary decision about the exit by choosing one of two circles presented at the end the vertical arms. All arm lengths were sampled from the same distribution as the H-maze. The speed of the ball was also the same as the H-maze task. Therefore the distribution of the time intervals associated with the two arms (t_L, t_R) was also the same as the H-maze task.

In condition one, the participants were shown the half maze and then the auditory temporal cues related to ball movements, just as in the baseline task. In condition two however, the order of stimulus presentations were reversed: the half maze was presented after the auditory temporal cues related to ball movements, forcing participants to maintain sensory evidence in working memory and use it afterwards to test different hypotheses about the subsequently shown half maze (i.e., counterfactual reasoning). All arm lengths were sampled from the same distribution as the H-maze. The speed of the ball was also the same as the H-maze task.

Task variant 3. Task variant 3 is similar to the task variant 2 where initially only one half of the maze is shown to the participants (a_L, a_{LU}, a_{LD} or a_R, a_{RU}, a_{RD}), however the true generative process of the ball's movements were identical to the baseline H-maze task (the ball could travel through the revealed horizontal arm and one of the revealed vertical arms, or through the hidden horizontal arm and one of the two hidden vertical arms). The participants were shown the half maze and then the auditory temporal cues related to ball movements. Participants decided the exit point among the two available options in the half maze, and subsequently were given the option to reveal the hidden half of the maze, which would allow them to revise their decision and choose between the two new exit ports (i.e., revision via counterfactual reasoning). All arm lengths were sampled from the same distribution as the H-maze. The speed of the ball was also the same as the H-maze task.

Decision model. Every time the ball reached a bifurcation point, it took one of two alternative arms. We developed a simple decision model based on signal detection theory (Green & Swets, 1966) to characterize how participants discriminated between the two arms using the time between the corresponding two cues. Let us denote the two arm lengths by a_1 and a_2. Since the ball moves at a constant speed, the arm lengths correspond to the two time intervals, t_1 and t_2. Depending on which arm the ball takes, the interval between the two cues, denoted t_s, will either match t_1 or t_2. The participant has to use a noisy measure of the actual interval, denoted t_m, to decide which of two arms the ball has taken. Consistent with scalar variability of timing (Malapani & Fairhurst, 2002), we modeled noise in the measurement of time as a sample from a zero-mean Gaussian distribution whose standard deviation scales with the base interval (t_1 or t_2), with the constant of proportionality w. According to this model, the conditional probability of t_m for the two arms can be written as:

$$p(t_m|t_1, w_m) = \frac{1}{2\pi w_m t_1} e^{\frac{-(t_m - t_1)^2}{2w_m^2 t_1^2}}$$

$$p(t_m|t_2, w_m) = \frac{1}{2\pi w_m t_2} e^{-\frac{(t_m - t_2)^2)}{2w_m^2 t_2^2}}$$

Using signal detection theory (Green & Swets, 1966), the optimal decision is to choose the arm with the higher conditional probability. This is equivalent to choosing the side on which t_m falls relative to a criterion, t_c, at the crossing point of the two conditional probability distributions. We refer to this model as the core decision module.

Estimating each participant's Weber fraction. We estimated w for each participant by fitting the core decision module to the behavior in the T-maze task. To accurately estimate w, we included t_c as a model parameter to account for idiosyncratic biases. The model also included an extra lapse rate parameter, Γ, to account for participants' lapses in performance. During lapse trials, the model chooses randomly between the two arms. With these additional parameters, the conditional probability for each arm can be written as follows:

$$p(\text{choose } t_1|tm) = 0.5\Gamma + (1-\Gamma)\int_{-\inf}^{t_c} p(t_m|t_1, w_m)dt_m$$

$$p(\text{choose } t_2|tm) = 0.5\Gamma + (1-\Gamma)\int_{t_c}^{\inf} p(t_m|t_2, w_m)dt_m$$

We computed the maximum likelihood estimate of w_m (as well as the other two parameters) for each participant by fitting this model to behavior during the T-maze task.

Optimal model. This model chooses the most likely exit by comparing the joint conditional distribution of tm_1 and tm_2 for the four alternatives. The four conditional probabilities are can be written as $p(tm_1,tm_2|t_L,t_{LD})$, $p(tm_1,tm_2|t_L,t_{LU})$, $p(tm_1,tm_2|t_R,t_{RU})$, and $p(tm_1,tm_2|t_R,t_{RD})$.

We computed these joint conditional probabilities by the product of corresponding marginals assuming that the two measurements are conditionally independent. Note that this model has no additional parameters other than w that was estimated from the T-maze experiment. Therefore, we did not fit this model to the participants' behavior; instead, we evaluate its behavior predictively.

Hierarchical Model. The hierarchical model solves the H-maze task by making two decisions, first between the horizontal arms and then between the corresponding vertical arms. The first decision chooses between the two horizontal arms by comparing the corresponding conditional probabilities, $p(tm_1|t_L)$ and $p(tm_1|t_R)$. The model chooses the more likely arm and then uses tm_2 to choose between the two vertical arms that branch off of the chosen horizontal arm. For example, if the model chooses the left arm in the first stage, it would then compare $p(tm_2|t_{LD})$ and $p(tm_2|t_{LU})$ to choose between the left-up and left-down vertical arms. This model also has no additional parameter other than w, which was estimated from the T-maze experiment. Therefore, the behavior of this model was also evaluated predictively.

Postdictive Model. The postdictive model is similar to the hierarchical model in that it solves the task by making two hierarchical decisions, first for the horizontal arms and then for the corresponding vertical arms. However, in this model, the first decision for the two horizontal arms is made postdictively by incorporating additional information about the tm_2 and the vertical arms. Specifically, the model makes its first decision by comparing $p(tm_1,tm_2|t_L,t_{LD})+p(tm_1,tm_2|t_L,t_{LU})$ to $p(tm_1,tm_2|t_R,t_{RU})+p(tm_1,tm_2|t_R,t_{RD})$. Next, it uses tm_2 to choose between the two vertical arms that branch off of the chosen horizontal arm, which is identical to the second

decision in the hierarchical model. Similar to the optimal and hierarchical model, this model has no additional parameter (w_m is derived from the T-maze experiment).

Counterfactual Model. The counterfactual model is identical to the hierarchical model but has the flexibility to revise its decisions when a certain measure of expected accuracy (X) is lower than a certain threshold (θ) that was fit as a free parameter.

Noise in the counterfactual model. To compute the likelihood of a counterfactual possibility, participants must perform an offline mental computation that involves inferring the values of tm_1 and tm_2 from memory to test a new hypothesis (the likelihood that the exit is down the other horizontal arm). Since the process of inferring time intervals from memory is noisy, we formulated the counterfactual model such that the recalled values of tm_1 and tm_2 (values used during the revision process) were participant to additional noise. Similar to the noisy optimal, hierarchical, and postdictive models, the noise in the counterfactual model was sampled from a Gaussian distribution with mean β and standard deviation that is proportional to the measured interval (tm_1 or tm_2) with a fixed constant of proportionality α (in accordance with scalar variability of time). The parameters α and β were fitted using maximum likelihood estimation.

Choice state-space analysis. For the state-space analysis, we encoded the pattern of choices in terms of four probabilities, one for each possible exit, denoted by p_{LU}, p_{LD}, p_{RU}, and p_{RD}. We first applied isometric transformations (rotation and reflection) to all H-mazes on each trial such that the left-up exit was the correct answer for all trials (the correct exit point of the ball was always *LU)*. p_{LU} is the probability of the correct exit point which is the most likely option (for analysis the H part of the maze on each trial was reflected horizontally and/or vertically so that the correct exit point of the ball was always *LU*), p_{LD} is the probability of choosing the correct horizontal arm but incorrect vertical arm, and p_{RU} and p_{RD} are the probabilities of choosing the incorrect horizontal and vertical arms. Next, we grouped unique trial conditions into a smaller number of maze sets. Each maze set contained all the mazes with the same arm length differences. We then quantified the dissimilarity between the behavior of each model and the participants/RNNs for each maze set by measuring the distance between the corresponding 3D vectors composed of [p_{LU} p_{LD} p_{RU} + p_{RD}]. Finally, we averaged the distance across maze sets to compute an overall measure of dissimilarity between choice patterns of the participants/RNNs and each model.

The RNN model. To construct the trials for the RNN, we sample the length of each of the six maze segments [L, R, LU , LD, RU, RD] from the set of lengths {4, 5, 6, 7} uniformly. Thus there are 4^6 = 4,096 possible maze geometries. We randomly split the geometries into 3,072 (75%) for model training and 1,024 (25%) for model testing. When RNN performs the task, we apply noise to the inter-click intervals. We convert each ground truth interclick interval ts_i to a perceived inter-click interval $tm_i = ts_i \cdot (1 + \sigma_{noise})$, where ϵ ~ N (0, 1) for σ_{noise} = 0.15. This noise is independently sampled for each of the two inter-click intervals per stimulus. For each maze geometry, we consider each of the four possible ball paths and for each of the conditions sample 10 perceived inter-click interval pairs (tm_1, tm_2). Thus we generate 40 trials for each maze geometry in both the training and testing sets.

To train RNNs on the H-maze task, we discretize time into 32 timesteps per trial. We let the ball begin 5 length units above the horizontal maze arms and travel at 1 unit per timestep, so on every trial the ball

reaches the first T-junction at timestep 6. For auditory input, we construct a horizontal and vertical auditory timing variable [a_1, a_2]. We let a_1 be 0 until the ball reaches the first T-junction, then ramp linearly from 0 to the first perceived inter-click interval tm_1 in $int(tm_1)$ (discretize to nearest integer) timesteps and remain constant thereafter. Similarly, we let a_2 be 0 until a_1 reaches tm_1, upon which a_2 ramps linearly to the second perceived inter-click interval tm_2 and in $int(tm_2)$ timesteps then remains constant. For visual input, we provide models with scalar values of the arm lengths [L, R, LU , LD, RU , RD]. The RNN provides two outputs, a horizontal choice and a vertical choice ([h, v]) that jointly specify the chosen exit.

Our counterfactual RNN model consists of a recurrent state $\boldsymbol{r}$ of N = 128 hidden units, updated each time step according to

$$\boldsymbol{r}_{t+1} = tanh\ (\boldsymbol{W}_{rec}\, \boldsymbol{r}_t + \boldsymbol{W}_{in}\, \boldsymbol{I}_{t+1}) \qquad (1)$$

where $\boldsymbol{W}_{rec}$ is an $N \times N$ matrix of trainable recurrent weights, $\boldsymbol{W}_{in}$ is a $N \times 8$ matrix of trainable input weights, and $\boldsymbol{I}$ is a 8-dimensional input vector [a_{1t}, a_{2t}, L, R, LU, LD, RU, RD] at each timestep.
The model produces two choices (horizontal and vertical) at each timestep representing the vertical and horizontal choices:

$$\boldsymbol{h}_t, \boldsymbol{v}_t = \sigma\ (\boldsymbol{W}_{out} \cdot \boldsymbol{r}_t\ ) \qquad (2)$$

where $\boldsymbol{W}_{out}$ is an $N \times 2$ matrix of learnable weights and σ is the sigmoid function $\sigma\ (x) = 1/(1 + e^{-x}\ )$. We use a threshold of 0.5 to binarize the choice to evaluate behavioral outcomes.

For the basic version of the RNN, we use an external-supervising signal to train the RNN. For each trial, we minimize the MSE between the RNN output [h_t, v_t] and the ground truth output. The loss is summed over the timesteps after tm_2 finishes ramping.

<u>Attention constraint.</u> To add the attentional bottleneck to the RNN, we change the original 8-dimensional input [a_{1t}, a_{2t}, L, R, LU, LD, RU, RD] into a 6-dimensional input [a_{1t}, a_{2t}, L, R, U_t, D_t], where U_t and D_t are the lengths of vertical arms gated by an attention module. Mechanistically, the attention module produces outputs U_t and D_t with the horizontal choice h_t at each timestep t as follows:

$$U_t = h_t \cdot LU + (1 - h_t\ ) \cdot RU$$
$$D_t = h_t \cdot LD + (1 - h_t\ ) \cdot RD$$

Accordingly, the input weight matrix becomes $N \times 6$ with the attention constraint.

<u>Rationality constraint.</u> Instead of supervised learning, we train the RNN with the objective of maximizing the likelihood of the chosen exit given the noisy timing information, which is the rational choice under uncertainty. Given a maze geometry and perceived inter-click intervals tm_1 and tm_2, an rational policy on the H-maze task chooses an exit-point corresponding to a path in the maze that maximizes the joint log likelihood log ($p\ (tm_1|ts_1)$) + log ($p\ (tm_2|ts_2)$). Because tm_i is sampled from a Gaussian distribution centered at the ground truth ts_i, maximizing the log likelihood is equivalent to minimizing the loss function

$$\mathcal{L} = \left(\frac{tm_1 - ts_1}{ts_1}\right)^2 + \left(\frac{tm_2 - ts_2}{ts_2}\right)^2$$

Considering the output logits [h_t, v_t] as the model's choice of which side of the maze the ball exits (left/right and up/down), the loss function becomes

$$\mathcal{L}(t) = h_t \left(\frac{tm_1 - L}{L}\right)^2 + (1 - h_t)\left(\frac{tm_1 - R}{R}\right)^2 + v_t \left(\frac{tm_2 - U_t}{U_t}\right)^2 + (1 - v_t)\left(\frac{tm_2 - D_t}{D_t}\right)^2 \qquad (3)$$

On the right-hand-side of the equation, the first two terms are the loss coming from the horizontal arm choice h_t and the last two terms are the loss coming from the vertical arm choice v_t given the horizontal arm choice. When training the models, this loss function is summed over the timesteps after tm_2 finishes ramping.

<u>Counterfactual noise constraint.</u> In order to discourage the model from switching its attention from side to side every trial, we impose an implicit cost for switching: after the second auditory click (namely after a_2 finishes ramping), each time $\hbar$ crosses 0.5 we add noise to the perceived inter-click intervals [tm_1, tm_2]. Specifically, for each timestep t where $\hbar$ crosses 0.5, we update tm as

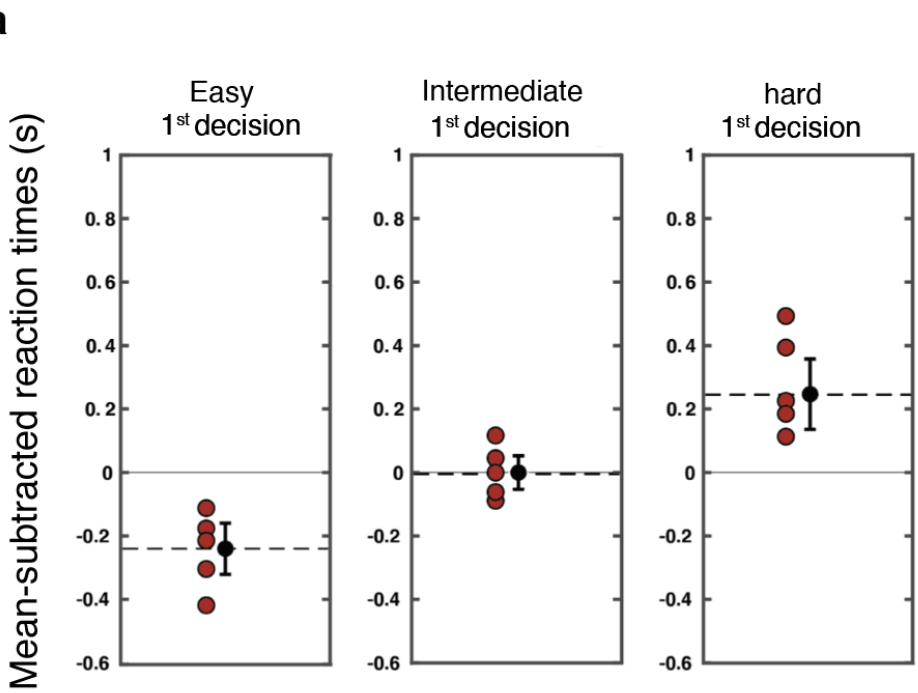
$$tm_i \leftarrow tm_i\,(1 + \sigma_{noise}\epsilon),\ \epsilon \sim N(0, 1),\ i = 1, 2$$

and use this noised tm for all subsequent timesteps in the trial. Here σ_{noise} is a hyperparameter controlling the magnitude of this switching penalty. We set σ_{noise} = 0.5 for the counterfactual noise constraint and σ_{noise} = 0 without this constraint.

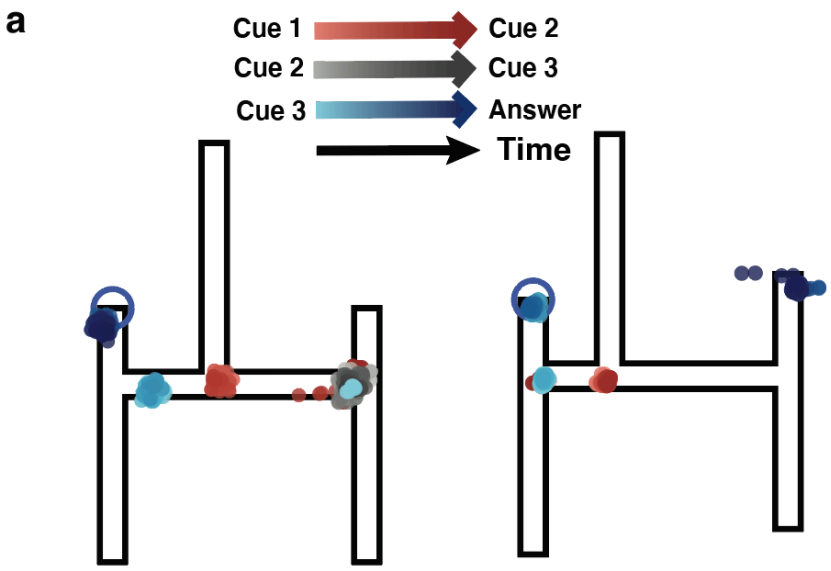
<u>RNN variants and model comparison</u> (Fig. 4b). There are in total eight possible RNN variants with three constraints that are binary (attention, rationality, and counterfactual noise). However, we ruled out two illogical variants: since counterfactual noise is triggered by the reverse of attention, it does not make sense to have counterfactual noise without the attention constraint. When comparing each RNN variant with the counterfactual model, we need a maximum likelihood estimation to find the best noise level of the counterfactual model. For simplicity, we simulated the counterfactual model three times with different noise levels: σ_{noise}=0 (no noise), σ_{noise}=0.5 (medium noise), σ_{noise}=1 (high noise). We chose the noise level for the counterfactual model that produces the least MSE to the RNN.

<u>RNN saccade</u> (Fig. 5b). Although the RNN generates continuous outputs [h_t, v_t] throughout the trial, we plot the saccade at two discrete time points for clarity (the time steps when tm_1 and tm_2 finishes ramping). For the first saccade (right after tm_1), we only consider h_t and ignore v_t because the RNN cannot make a vertical decision without tm_2. For the second saccade (right after tm_2), the saccade position reflects both h_t and v_t.

Supplementary Figures

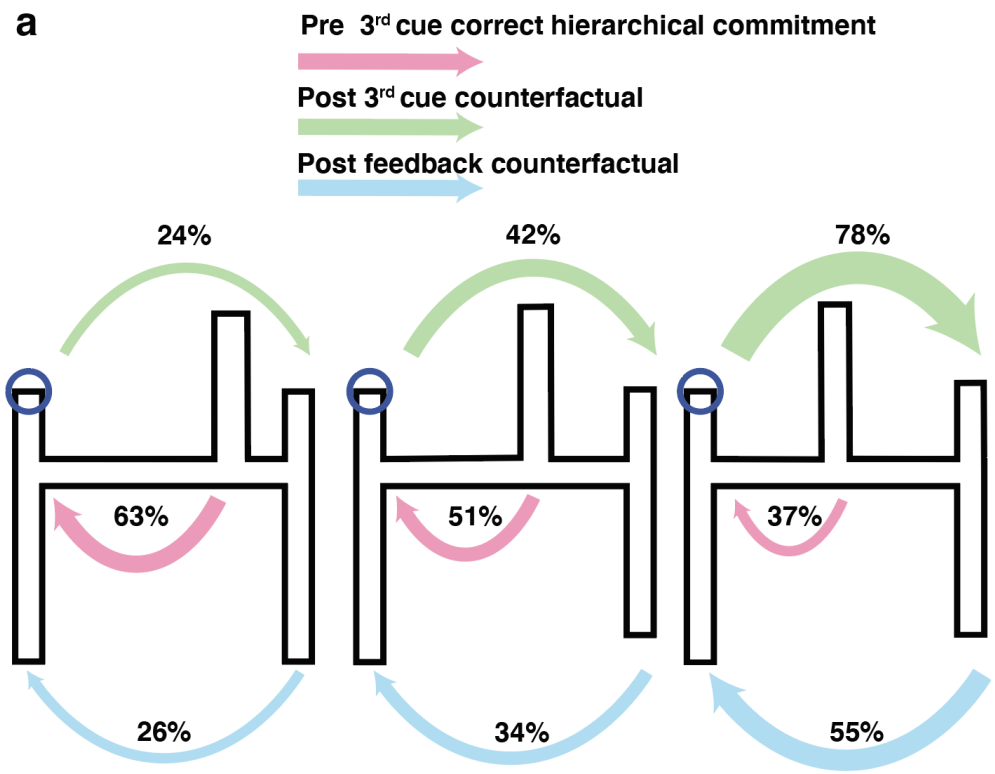


Supplementary Figure 1. Participants' reaction times on Baseline H-Maze. **(a)** Mean subtracted reaction times (computed as the time from flash three to the subject's choice) of human participants (seconds) for 1st decision conditions with low, intermediate, and high difficulty are plotted in red. Average reaction times of participants per condition is plotted in black, with std. indicated by black error bars.



Supplementary Figure 2. Dynamic Eye-Movement Readout. **(a)** Gaze position density throughout the trial is overlaid on the H-maze. In these examples, the correct endpoint is top-left and the participant's answer is indicated by the open blue circle. Gaze position density is color coded based on the task epoch (red: the ball is in the horizontal

segment; gray: the ball is in the vertical segment; blue: the trial has ended and the ball has reached the endpoint). Bright and dark shades represent early and late in each epoch. The pattern of eye movements indicate that participants made their choice hierarchically (early leftward or rightward saccades in the top left panel). They changed their mind occasionally and made a saccade to the other vertical bar either after the second auditory cue (top right), or the third auditory cue (bottom left and right).



Supplementary Figure 3. Condition dependent eye-movements. **(a)** A directed probabilistic graph of the transition probabilities between hierarchical and counterfactual saccades is shown for the three horizontal arm difficulty conditions, plotted separately here on three maze schematics representing each condition. The pink arrows represent correct hierarchical saccades made to the one of the two horizontal arms before the third cue is presented. The green arrow represents counterfactual saccades made after the third cue was provided to the arms of the maze alternative to the initial decision commitment. The blue arrow represents saccades to the arms made after choice feedback to the arms of the maze alternative to the initial decision commitment. The width of arrows is proportional to probability, and the percentage of trials on which the transition occurred is shown.

References

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39.

Buchsbaum, D., Bridgers, S., Skolnick Weisberg, D., & Gopnik, A. (2012). The power of possibility: causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1599), 2202–2212.

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412–441.

Davis, E., & Marcus, G. (2016). The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, *233*, 60–72.

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–1486.

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron*, *80*(1), 35–50.

Egger, S. W., & Jazayeri, M. (2018). A nonlinear updating algorithm captures suboptimal inference in the presence of signal-dependent noise. *Scientific Reports*, *8*(1), 12597.

Gershman, S. J., & Burke, T. (2022). Mental control of uncertainty. *Cognitive, Affective & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-022-01034-8

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gerstenberg, T. (2022). What would have happened? Counterfactuals, hypotheticals and causal judgements. *Philosophical Transactions of the Royal Society of London. Series B,*

Biological Sciences, *377*(1866), 20210339.

Gerstenberg, T., Goodman, N. D., Lagnado, D., & Tenenbaum, J. (2020). A counterfactual simulation model of causal judgments for physical events. In *https://psyarxiv.com › ...https://psyarxiv.com › ...* https://doi.org/10.31234/osf.io/7zj94

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. *455*. https://psycnet.apa.org/fulltext/1967-02286-000.pdf

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3098–3103.

Kuperwajs, I., & Ma, W. J. (2021). Planning to plan: a Bayesian model for optimizing the depth of decision tree search. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/3td607f3

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *The Behavioral and Brain Sciences*, *40*, e253.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21086–21090.

Li, J., Watters, N., Sohn, H., & Jazayeri, M. (2023). Modeling Human Eye Movements with Neural Networks in a Maze-Solving Task. In I. Lourentzou, J. Wu, S. Kashyap, A. Karargyris, L. A. Celi, B. Kawas, & S. Talathi (Eds.), *Proceedings of The 1st Gaze Meets*

ML workshop (Vol. 210, pp. 98–112). PMLR.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14.

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*(4), 700–734.

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on Simulation Approaches in Intuitive Physics. In *https://psyarxiv.com › xhzuchttps://psyarxiv.com › xhzuc*. https://doi.org/10.31234/osf.io/xhzuc

Malapani, C., & Fairhurst, S. (2002). Scalar Timing in Animals and Humans. *Learning and Motivation*, *33*(1), 156–176.

Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Communications*, *10*(1), 3366.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peixoto, D., Verhein, J. R., Kiani, R., Kao, J. C., Nuyujukian, P., Chandrasekaran, C., Brown, J., Fong, S., Ryu, S. I., Shenoy, K. V., & Newsome, W. T. (2021). Decoding and perturbing decision states in real time. *Nature*, *591*(7851), 604–609.

Piloto, L. S., Weinstein, A., Battaglia, P., & Botvinick, M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature Human Behaviour*, 1–11.

Purcell, B. A., & Kiani, R. (2016). Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(31), E4531–E4540.

Ramírez-Ruiz, J., & Moreno-Bote, R. (2022). Optimal allocation of finite sampling capacity in accumulator models of multialternative decision making. *Cognitive Science*, *46*(5), e13143.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Remington, E. D., Parks, T. V., & Jazayeri, M. (2018). Late Bayesian inference in mental transformations. *Nature Communications*, *9*(1), 4419.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266.

Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, *364*(6441). https://doi.org/10.1126/science.aav8911

Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron*, *90*(5), 927–939.

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, *10*(9), 407–412.

Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (n.d.). *Physical predictions over time*. Retrieved July 28, 2023, from https://escholarship.org/content/qt9m0197n4/qt9m0197n4_noSplash_bcd635b87e5a33fe34a1ee16b8c141dc.pdf?t=os605k

Spering, M. (2022). Eye Movements as a Window into Decision-Making. *Annual Review of Vision Science*, *8*, 427–448.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Van Hoeck, N., Watson, P. D., & Barbey, A. K. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in Human Neuroscience*, *9*, 420.

Van Opheusden, B., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2017). A computational model for decision tree search. *CogSci*. https://cogsci.mindmodeling.org/2017/papers/0241/paper0241.pdf

Wolfe, J. M., Reinecke, A., & Brawn, P. (2006). Why don't we see changes?: The role of attentional bottlenecks and limited visual memory. *Visual Cognition*, *14*(4-8), 749–780.

Yu, Q., Panichello, M. F., Cai, Y., Postle, B. R., & Buschman, T. J. (2020). Delay-period activity in frontal, parietal, and occipital cortex tracks noise and biases in visual working memory. *PLoS Biology*, *18*(9), e3000854.

Zheng, Z. (sam), Lin, X. (daisy), Topping, J., & Ma, W. J. (2022). Comparing Machine and Human Learning in a Planning Task of Intermediate Complexity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/8wm748d8

Zylberberg, A. (2021). Decision prioritization and causal reasoning in decision hierarchies. *PLoS Computational Biology*, *17*(12), e1009688.

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. *Neuron*. https://doi.org/10.1016/j.neuron.2018.07.035