Research Article

Psychological Science
2023, Vol. 34(11) 1281–1292

DOI: 10.1177/09567976231200547
www.psychologicalscience.org/PS

Rational Simplification and Rigidity in Human Planning

Mark K. Ho[1,2], Jonathan D. Cohen[3], and Thomas L. Griffiths[1,2]

[1]Department of Psychology, Princeton University; [2]Department of Computer Science, Princeton University; and [3]Princeton Neuroscience Institute, Princeton University

Abstract

Planning underpins the impressive flexibility of goal-directed behavior. However, even when planning, people can display surprising rigidity in how they think about problems (e.g., "functional fixedness") that lead them astray. How can our capacity for behavioral flexibility be reconciled with our susceptibility to conceptual inflexibility? We propose that these tendencies reflect avoidance of two cognitive costs: the cost of representing task details and the cost of switching between representations. To test this hypothesis, we developed a novel paradigm that affords participants opportunities to choose different families of simplified representations to plan. In two preregistered, online studies (*N*s = 377 and 294 adults), we found that participants' optimal behavior, suboptimal behavior, and reaction time were explained by a computational model that formalized people's avoidance of representational complexity and switching. These results demonstrate how the selection of simplified, rigid representations leads to the otherwise puzzling combination of flexibility and inflexibility observed in problem solving.

Keywords

planning, problem solving, causal reasoning, functional fixedness, task switching, open data, open materials, preregistered

Received 1/19/23; Revision accepted 7/20/23

Flexible, goal-directed behavior rests on the human capacity to plan (Daw et al., 2005; Mattar & Lengyel, 2022; Newell & Simon, 1972). For example, if your normal driving route to work is suddenly blocked by construction, you could plan a new route by creating a model of the situation (e.g., forming a "cognitive map" of nearby roads, traffic, and other places with construction) and mentally simulating possible solutions (i.e., driving routes that could lead to work). Planned behaviors contrast with habits, which are inflexible behavioral repertoires that have been automatized or learned through trial and error (Wood & Rünger, 2016) and persist even in the face of new information (e.g., reflexively taking your normal route to work even though you may have seen on the news that there is construction).

In contrast to the recognition that planning is a cornerstone of behavioral flexibility, an equally long-standing tradition in psychology emphasizes how rigidity in people's planning processes can lead to systematic suboptimalities or impasses in problem solving. For instance, in the classic candle problem (Duncker, 1945), participants are presented with a box of tacks, a book of matches, and a candle and are instructed to mount the candle on a wall and light it. Motivated participants routinely fail to identify the solution to the task (empty the box of tacks, tack the box to the wall, place the candle in the box and light it). This failure to identify a good plan is often attributed to *functional fixedness*, in which participants are "fixed" on certain object affordances (how the box can be used as a container) and overlook other affordances (how the box can be used as a support).

Phenomena such as functional fixedness are particularly intriguing because they are not easily explained

Corresponding Author:
Mark K. Ho, Department of Psychology, Princeton University
Email: mho@princeton.edu

by classical accounts of planning. For instance, in models based on heuristic search (Kaplan & Simon, 1990; Newell & Simon, 1972), the problem solver is assumed to have a fixed representation of a task (e.g., in chess, a representation of the board, pieces, and how they move) and then perform computations over that representation (i.e., simulating sequences of moves and countermoves). Systematically overlooking an obvious solution (such as using a box to support a candle) should not happen if people are searching through a fully specified task representation. This observation has led researchers to conclude that problem solving and planning depends crucially on identifying the appropriate *problem representation* (Knoblich et al., 1999; Ohlsson, 1984). However, although it has been argued that choosing the right representation plays a role in how people flexibly solve new problems, the general principles and/or mechanisms underlying such processes remain unclear (Kaplan & Simon, 1990; Newell & Simon, 1972; Ohlsson, 2012).

Here, we aim to shed light on the long-standing question of how people can be actively engaged in goal-directed planning yet fail to use the best representation for a problem, as exemplified by classic findings on functional fixedness. We propose that such cognitive biases are a consequence of how people manage two cognitive costs inherent to planning: *complexity costs* and *switch costs*. Complexity costs correspond to the amount of detail that is represented for the purposes of planning a solution—for example, the number of chess pieces attended to when planning a move. Recent work has demonstrated that people manage representational complexity by intelligently constructing simplified models of problems called *task construals* (Ho et al., 2022). Crucially, complexity costs are a function of the limited cognitive resources (e.g., attention) available to solve the immediate problem at hand.

In contrast, switch costs emerge from changing the family of representations used across problems. Evidence for switch costs primarily comes from work on task switching in simple decision-making paradigms. For example, when classifying a digit on the basis of its parity (odd/even) or its magnitude (high/low), people display slower reaction times when switching between the tasks than when continuing to perform the same task (Arrington & Logan, 2004). The cost of switching which features of a stimulus to attend to in order to guide actions—often called the *task set*—has been attributed to a number of different cognitive mechanisms, including reconfiguration of the new task set as well as persistence of the previous task set (Grange & Houghton, 2014; Monsell, 2003; Vandierendonck et al., 2010). One way to view the current work is as an extension of ideas such as switching and task sets from deciding on what single action to take to deciding on what sequence of actions to take—that is, to multistep planning.

Statement of Relevance

Even when planning ahead, people exhibit surprising cognitive biases. One classic example is *functional fixedness*, the tendency to overlook relevant but unfamiliar ways of using an object to solve a problem despite actively searching for a solution. Functional fixedness is not only puzzling but also pervasive: Thinking about situations in rigid and unhelpful ways has consequences in virtually every domain of human life, from overcoming personal challenges to tackling complex problems in science, business, and politics. Here, we tested a new theory of functional fixedness based on the avoidance of two kinds of cognitive effort: the effort required to represent details of a problem and the effort required to switch to a new way of representing a problem. Our findings reveal how cognitive biases such as functional fixedness, despite appearing superficially irrational, can reflect a deeper rationality grounded in the strategic allocation of cognitive effort.

However, whereas existing work on task switching has mainly addressed the particular mechanisms of switch costs, our goal was to determine how the avoidance of switch costs interacts with the avoidance of complexity and achievement of goals to produce functional fixedness. This investigation required two components that we developed in the present research. First, we needed an experimental paradigm that could systematically elicit functional fixedness (as in the classic studies) and was also amenable to trial-by-trial analysis (as in standard task-switching studies and contemporary research on decision-making more broadly; Wilson & Collins, 2019). Second, we needed a computational framework for interpreting the intricate dynamics of planning, task representations, and cognitive costs both within and across problems. These two building blocks are the core methodological contributions of this research.

To understand our paradigm, consider the 2D mazes shown in Figure 1, where the blue circle marks the start location, and the yellow square marks the goal. Black tiles represent walls that prevent movement, and the blue tiles correspond to blocks that generally prevent movement (darker blue) but have smaller notches that can be traversed (lighter blue). These mazes can be represented in one of two ways: Either in terms of both

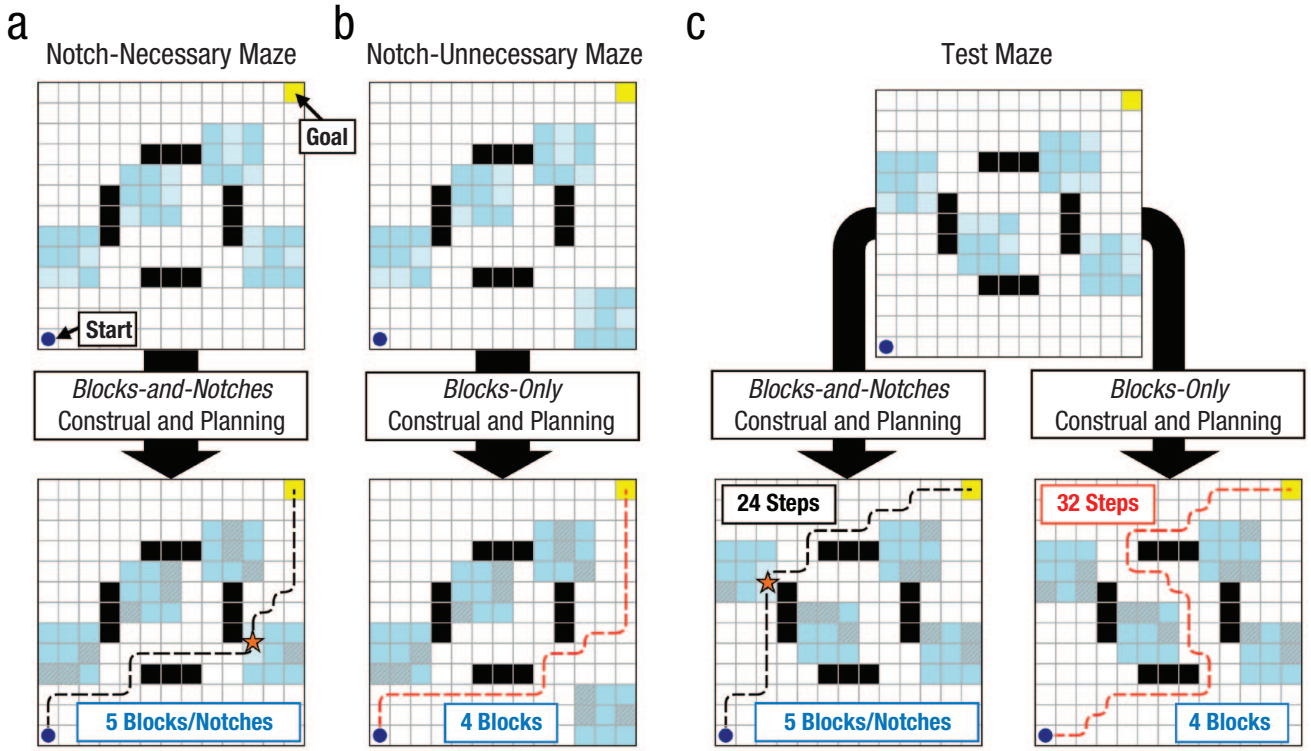

Fig. 1. Blocks-and-notches mazes in Experiment 1. (a) Participants move through the maze by controlling the blue circle. Mazes were composed of walls (black tiles), a goal (yellow tile), and blocks (blue 3 × 3 objects). The walls and blocks prevented movement, except for the lighter notches in the blocks, through which movement was allowed. Our account predicted that people would attend only to the notch relevant to identifying an optimal path (marked by an orange star) and ignore the other notches (indicated with gray stripes). A *notch-necessary* maze is shown, in which the construal that optimizes representational simplicity and behavioral utility requires attending to a notch. (b) On *notch-unnecessary* mazes, an optimal path can be found by forming a blocks-only construal rather than a more complex blocks-and-notches construal. (c) *Test mazes* produce different planned behaviors depending on whether the optimal blocks-only versus optimal blocks-and-notches construal is used, allowing us to diagnose which construal strategy they are using. In particular, only participants who used a blocks-and-notches strategy would visit a *critical notch*, marked here with an orange star.

blocks and notches or only in terms of blocks (Figs. 1a and 1b). These two families of representations induce distinct construal strategies that participants can apply to a given problem. We predicted that because the blocks-only construal strategy had a lower complexity than the blocks-and-notches strategy, people would tend to prefer the former, when possible. Furthermore, we predicted that when participants were solving mazes in sequence, the cost of switching between representational strategies would bias them to reuse their previous strategy even when it would lead to taking a suboptimal path (e.g., overlooking an opportunity to take a shorter route via a notch). That is, we predicted that functional fixedness would arise because people would attempt to avoid complexity and switching during planning.

To precisely characterize and test this hypothesis, we built on the computational framework of value-guided construal (Ho et al., 2022), which formulates the problem of selecting a task construal in terms of an approximately optimal trade-off between cognitive costs and task performance. In this sense, value-guided construal approaches planning representations from a normative, resource-rational perspective (Anderson, 1990; Gershman et al., 2015; Griffiths et al., 2015; Lewis et al., 2014; Marr, 1982; Shenhav et al., 2013). To apply these ideas to functional fixedness, we treated construal selection as a hierarchical process consisting of construal strategy selection (e.g., deciding whether to attend to only blocks or both blocks and notches) and construal selection (e.g., attending to a specific combination of blocks and/or notches). As we will discuss in more depth in the analysis of our first experiment, this modeling approach provides an invaluable analytical tool for assessing the paradoxical combination of behavioral flexibility and conceptual rigidity observed in human problem solving.

Open Practices Statement

Preregistrations, data, code, and materials for Experiments 1 and 2 have been made publicly available via OSF (https://osf.io/yuta6).

Experiment 1: Functional Fixedness in Maze Navigation

Our first experiment tested whether people form simplified and rigid task representations when planning in maze tasks such as those discussed above. Participants first completed one of two sets of *training mazes* that differed in whether they could be solved by representing mazes in terms of blocks only (notch unnecessary) or both blocks and notches (notch necessary). This was followed by a set of *test mazes* in which the blocks-only solution was longer than the blocks-and-notches solution (Fig. 1c). If people avoid both complexity and switching, then those trained on the notch-unnecessary trials will adopt a blocks-only construal strategy and persist in using it on the test trials. In other words, we predicted functional fixedness: They would overlook opportunities to use the notches to solve the task, even when doing so would have led to a better solution.

Method

Four hundred nineteen participants from the Prolific experimental platform with U.S. and UK Internet protocol (IP) addresses completed our experiment. This number was based on a power analysis with pilot data. Each participant was paid a $1.34 base wage and could additionally receive a $1 to $2 performance-based bonus. All experiments were approved by the Princeton University Institutional Revew Board (Protocol ID: 10859, protocol title: Computational Cognitive Science).

The experiment consisted of pretask instructions, the main task, and posttask questions (free response and age). After providing consent and before the main part of the experiment, participants were given instructions and practice trials that introduced them to the dynamics of the mazes. Each maze was a 13×13 grid consisting of central black walls (fixed across all mazes), empty white regions, a blue circle that could be controlled with the arrow keys, a yellow goal location, and four 3×3 blue blocks. The body of the 3×3 blocks prevented movement, but they could have smaller 1×1 or 1×2 notches that permitted movement through that part. The body of the blocks were rendered in the browser with a red, green, blue (RGB) value of 173, 216, 230, and the notches were indicated by a lighter shade (alpha transparency value was reduced to .5). At the beginning of each trial, a green square (slighly smaller than 1 × 1) was shown overlaying the goal to help draw attention to its location and to encourage planning at the beginning of the trial by indicating a deadline (i.e., time pressure) for moving once actions were begun. The green square remained fixed in size until the first move was taken, after which it progressively shrank for 1,000 ms until the next move was taken or it disappeared. That is, after each move, the square returned to its original size but immediately began to shrink again until the next move was taken.

The first two practice trials included a trial with blocks that did not have any notches, as well as a trial with notched blocks that required navigating through a notch to reach the goal. This set of practice trials was crucial, as it ensured that all participants were explicitly introduced to the idea of a notch and directly experienced traversing it to navigate to the goal. The second two practice trials demonstrated the bonus structure of the mazes: On each trial, participants began with 100 points, each step led to a reduction of 1 point, and each bump into a wall resulted in a reduction of 10 points. Points on a trial could not go below 0. If the green square disappeared, 0 points were received on that trial. Points were converted to bonus payments at a rate of 10 cents per 100 points. On these two practice trials, but not on the main trials, the points left at each time step were shown at the top of the screen. Finally, before starting the main trial sequence, participants were shown a review of the rules and needed to answer five comprehension questions correctly within two tries to continue (questions incorrectly answered on the first try were marked in red).

In the main task, participants were placed into two groups, each of which was assigned to one of the two training conditions: Participants in the *notch-unnecessary* training condition solved 12 mazes for which the shortest path did not require going through a notch and therefore could be solved by construing mazes only in terms of blocks. The mazes were pseudorandomly sampled from 128 distinct mazes that were generated from eight base mazes that were randomly transformed according to the eight symmetries of a square and having their start/goal locations randomly swapped. Participants in the *notch-necessary* training condition solved 12 mazes sampled from another set of 128 mazes in which the shortest path required passing through a notch. This second set was similarly generated from eight base mazes, but these mazes were designed to have a shortest path that was a similar sequence of actions as the notch-unnecessary training mazes. Thus, the training conditions used mazes that were matched on optimal behavior, but solving the mazes in the notch-necessary condition required using the more cognitively more expensive strategy of considering both blocks and notches.

After the 12 training mazes, participants from both conditions were given eight test mazes that had two types of solutions: a short, direct path to the goal that passed through a notch, and a long, roundabout path to the goal that did not pass through any notches

(Fig. 1c). Crucially, if participants planned by adopting a blocks-only construal strategy, they should have taken the long route. Conversely, if they planned by adopting a blocks-and-notches construal strategy, they should have taken the short route. The eight test mazes were pseudorandomly sampled from a set of 64 mazes constructed from four distinct base mazes that were transformed in the same way as the training mazes (see the Supplemental Material available online for base mazes).

Results

Behavioral results. We analyzed data for test trials that met preregistered exclusion criteria—not more than 25,000 ms or 1,000 ms spent at the initial state or a noninitial state, respectively. In addition, we excluded participants who had more than half of their test trials excluded, which resulted in a total of 377 for the behavioral analyses (212 male, 160 female, four nonbinary, one gave no response; age: $M = 38$ years, range = 18–75).

The main behavioral results support the hypothesis that people adopt simplified but rigid construal strategies, as summarized in Figure 2. In particular, Figure 2a shows that participants trained on notch-unnecessary mazes were more likely to take longer, notch-free paths on test mazes; this indicates that they initially adopted the blocks-only construal strategy to plan (consistent with avoiding complexity) but then persisted in using this even when it led to a less optimal solution in the test mazes (consistent with avoiding switching). Additionally, Figure 2b shows that across both conditions, taking a longer, notch-free path was associated with faster first-move reaction times, consistent with the assumption of our account that the blocks-only construal strategy is less cognitively effortful (i.e., costly) than the blocks-and-notches strategy.

To test the statistical significance of these observed effects, we conducted two sets of analyses. We first compared whether participants in the two conditions visited the *critical notch* on a maze, defined as the unique notch that lies along the optimal path when considering blocks and notches but not when only considering blocks. Four logistic regression models with critical notch visitation as a binary dependent variable were fitted by starting with an intercept-only model and then progressively adding the following four regressors: training condition (sum coded: notch necessary = .5, notch unnecessary = −.5), test-trial number, and an interaction term. Adding training condition significantly increased fit according to a log likelihood-ratio test, $\chi^2(1) = 313.83$, $p = 3.2 \times 10^{-70}$, as did including test-trial number, $\chi^2(1) = 41.72$, $p = 1.1 \times 10^{-10}$. The interaction term was not significant after Bonferroni correction, $\chi^2(1) = 4.59$, $p = .032$, uncorrected. The estimated weights on the final model were consistent with less visitation after notch-unnecessary training ($\beta = 1.90$, $SE = 0.17$), increasing notch visitation over time ($\beta = 0.11$, $SE = 0.02$), and a sharper increase in visitation following notch-unnecessary training mazes ($\beta = -0.09$, $SE = 0.04$). These results suggest that participants in the notch-unnecessary condition adopted a blocks-only construal strategy and persisted in this strategy even when it led to suboptimal task performance (longer paths).

We also examined first-move reaction times on test mazes as a proxy for time spent planning and cognitive effort. In an intercept-only model, adding whether a critical notch was visited on a trial as a regressor (sum coded: visited = .5, not visited = −.5) led to a significant increase in fit, $\chi^2(1) = 96.22$, $p = 1 \times 10^{-22}$. Including training condition as a regressor did not significantly increase fit, $\chi^2(1) = 2.19$, $p = .14$, but including trial number did, $\chi^2(1) = 12.68$, $p = .00037$, uncorrected. Finally, including critical-notch visitation and trial number as regressors indicated that visiting a notch predicted slower reaction times ($\beta = 0.35$, $SE = 0.03$), whereas responses became faster over time ($\beta = -0.02$, $SE = 0.01$). These results suggest that planning a shorter path by using the optimal blocks-and-notch construal required more time ($e^{7.77+0.5\times0.35} - e^{7.77-0.5\times0.35} \approx 833$ ms more based on the estimated coefficients), consistent with requiring greater cognitive effort, as predicted by our account.

Model-based results. Value-guided construal provides an account of how complexity costs within mazes and switch costs between mazes interact to shape measured behavior. To more directly assess the interactions posited by the theory, we compared a series of computational models fitted to participant movements across the experiment. These analyses revealed that low-level movements executed by participants are best explained by a version of the model that includes both complexity and switch costs rather than either one alone. Additionally, we found that the maze-specific construal complexity costs estimated by the model can be used to predict first-move reaction times on trials, despite being fitted only to maze movements. By explicitly formalizing how complexity and switch costs affect decision-making, these results provide a quantitative account of functional fixedness in terms of people's adoption of simplified but rigid task representations during planning.

We began by formalizing complexity and switch costs and relate these quantities to planning and action selection. Value-guided construal (Ho et al., 2022) provides an account of how people form task construals, simplified task representations used during planning. For example, in the maze task, different construals correspond to attending to different sets of blocks and notches, and planning with a construal involves finding the optimal path assuming only the blocks and notches

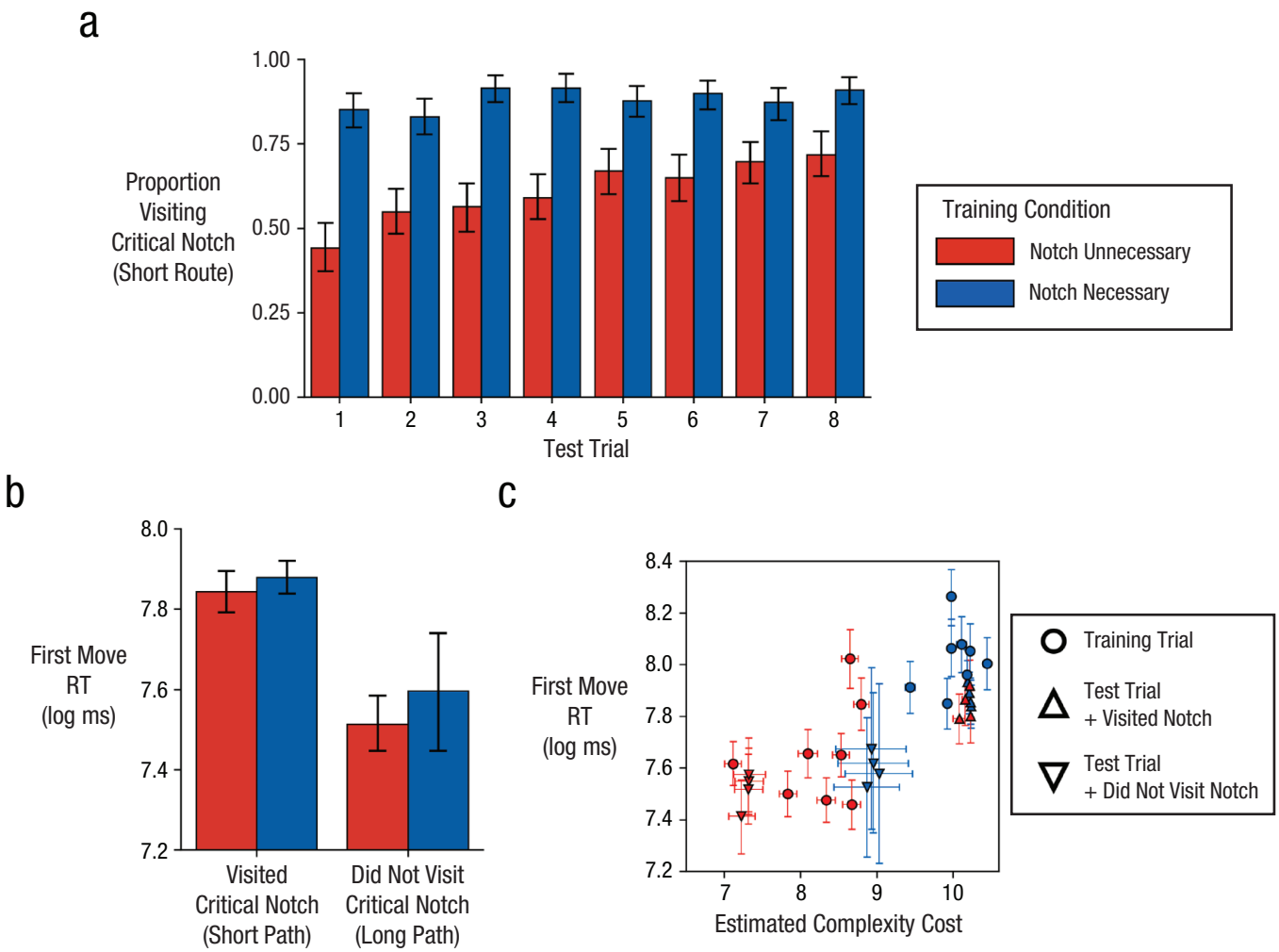

Fig. 2. Experiment 1 results. (a) Participants (N = 377) received 12 training mazes whose shortest paths required traversing a notch (notch necessary) or not traversing a notch (notch unnecessary). Both groups then received eight test mazes that could be solved without traversing any notches or by taking a shortcut through a critical notch. Participants in the notch-unnecessary group were less likely to take shortcuts, consistent with our account. (b) When participants took shortcuts on test trials, they were slower to begin moving, regardless of training condition. If first-move reaction times (RTs) reflect planning costs, this result is consistent with a higher cost for more complex construals (i.e., attending to more blocks and/or notches to plan). (c) By fitting complexity and switching cost parameters (see Model-Based Results in the text), we were able to estimate latent planning processes. The plot shows that complexity costs estimated only from movements predicted reaction times on individual training mazes (circles) and test mazes (triangles; direction indicates whether a shortcut was taken). See the text for more detailed analyses. Error bars are bootstrapped 95% confidence intervals.

included in a construal. The value of a construal for a particular task then depends on its behavioral utility—that is, the performance of the plan induced by the construal when applied to the actual problem—and its representational complexity—here, the total number of blocks and notches included in the construal (note that we assume that attending to a notch requires attending to the block of which it is a part). Formally, the value of a construal c on task τ is

$$\text{value}(c, \tau) = \text{utility}(\pi_c, \tau) - \text{complexity}(c), \quad (1)$$

where π_c is the unique optimal stochastic plan for construal c, $\text{utility}(\pi_c, \tau)$ is the behavioral utility of the plan computed by the construal when evaluated on the actual task, and $\text{complexity}(c) = \lambda_{\text{Complexity}} \times \text{Size}(c)$ is the complexity cost, which is the size of the construal multiplied by a constant weight. As in other computational work on sequential decision-making (Dayan & Niv, 2008; Sutton & Barto, 2018), tasks and construed tasks are modeled as Markov decision processes and plans as policies that map states (locations on the grid) to distributions over actions (cardinal directions). For further details, see the Supplemental Material.

In this work, we extend the original value-guided construal formulation by making explicit the idea of a *construal strategy*, which corresponds to how a decision-maker selects a family of construals—for example, the

family of block-only versus the family of block-and-notch construals. Formally, a construal strategy σ is a function from a maze to a set of contruals. Given the construal set generated from a construal strategy, $\sigma(\tau)$, choosing a specific construal on a particular task is modeled as a softmax (Luce, 1959) over the values of construals (Equation 1) in that set, $\pi(c|\tau,\sigma) \propto e^{\text{value}(c,\tau)}\mathbf{1}[c \in \sigma(\tau)]$, where $\mathbf{1}[X]$ is an indicator function that equates to 1 when X is true and 0 otherwise.

For the blocks-and-notches mazes, we considered two construal strategies: block-only construals and block-and-notch construals (Fig. 1). The *intrinsic value* of a construal strategy depends on the expected behavioral utilities and complexity costs of construals consistent with that strategy on a task. Thus, for instance, the block-and-notch construal strategy will tend to have higher complexity costs because attending to notches necessarily requires attending to blocks (notches are parts of blocks). To characterize the dynamics of construal strategy selection across mazes, we calculated the value of a construal strategy on a particular maze as the sum of its intrinsic value applied to that problem and the cost of switching strategies. That is, formally, the value of adopting a construal strategy σ_t on a task τ_t when the previous strategy is σ_{t-1} is

$$\text{value}(\sigma_t, \tau_t, \sigma_{t-1}) = \text{value}(\sigma_t, \tau_t) - \text{switch}(\sigma_{t-1}, \sigma_t), \quad (2)$$

where $\text{value}(\sigma_t, \tau_t) = \sum_{c_t} \pi(c_t \mid \tau_t, \sigma_t)\text{value}(c_t, \tau_t)$ is the intrinsic value of adopting the construal strategy for the new task, and $\text{switch}(\sigma_{t-1}, \sigma_t) = \lambda_{\text{Switch}} \times \mathbf{1}[\sigma_{t-1} \neq \sigma_t]$ is the switch cost, which is determined by whether the new construal strategy is different from the previous one and a constant weight.

Using Equation 2, we defined a *construal strategy transition system*, $P(\sigma_t \mid \tau_t, \sigma_{t-1}) \propto e^{\text{value}(\sigma_t, \tau_t, \sigma_{t-1})}$, which represents the probability of transitioning to a new construal strategy given the old strategy and current task. From a cognitive-modeling perspective, we could then treat estimation of the representational complexity cost weight ($\lambda_{\text{complexity}}$) and construal-strategy switch-cost weight (λ_{switch}) as learning the parameters of a hidden Markov model (Russell & Norvig, 2009). Specifically, the model consisted of a sequence of latent construal strategies (i.e., blocks-only vs. blocks-and-notches strategies) that output a sequence of observed paths on each maze. The transition and output probabilities of this hidden Markov model were highly nonlinear, as they involved construal selection, planning, and execution of a plan in a maze on each trial. Nonetheless, our implementation exactly marginalized out all latent variables using dynamic programming (see the Supplemental Material), allowing us to calculate the maximum marginal likelihood parameters.

By comparing model fits in the absence of a cost (i.e., fixing its coefficient to 0) with costs in its presence (letting its coefficient range from 0 to 10), we could assess whether those costs played a role in shaping decision-making. Thus, we fitted four models in which $\lambda_{\text{complexity}}$ and λ_{switch} were either set to 0 or allowed to range from 0 to 10. To allow additional flexibility in how the dynamics of switching construals relate to low-level movements, we fitted an ε-greedy action policy (Sutton & Barto, 2018) using the optimal value function associated with plans computed by a construal with a $\varepsilon_{\text{move}}$ parameter. All the models were implemented in the *msdm* Python package, and a single set of parameter estimates were calculated by minimizing the negative log likelihood of all participant movements on all mazes (both training and test) using the L-BFGS-B algorithm in *scipy.optimize* (Virtanen et al., 2020).

Estimated parameters for the four models on participant data, shown in Table 1, reveal that the model that included both complexity and switch costs provided a better fit to movement responses across trials than when either or both costs were absent. Additionally, we applied the forward-backward algorithm (Russell & Norvig, 2009) to the dynamics of the model fit with both costs to calculate a marginal probability of the blocks-only versus blocks-and-notches construal strategy being used by each participant on each trial t, $P(\sigma_t)$. This allowed us to calculate the *expected complexity* on each trial, $\sum_{\sigma_t} P(\sigma_t) \sum_{c_t} \pi(c_t \mid \tau_t, \sigma_t)\text{complexity}(c_t)$, which we can compare with first-move reaction times, a measure of time spent planning (Fig. 2c). To assess the relationship between costs in the model and reaction times, we fitted a hierarchical linear regression with by-participant random effects as well as trial number and expected complexity as fixed effects to reaction times. This analysis showed a significant effect of adding both—trial number: $\chi^2(1) = 31.48$, $p = 6.1 \times 10^{-8}$; expected complexity: $\chi^2(1) = 102.37$, $p = 1.4 \times 10^{-23}$—as well as decreasing reaction times over trials ($\beta = -0.007$, $SE = 0.001$) and a positive relationship with expected complexity ($\beta = 0.065$, $SE = 0.006$). Thus, our model fitted to participant movements predicted participant reaction times as a function of expected complexity costs. This provided additional confirmation of the representational complexity costs posited by our account.

Experiment 2: Controlling for Simple Rules

Experiment 1 suggests that people avoid complexity and switching, which leads to functional fixedness. However, this conclusion depends on the contrast with behavior in the notch-necessary condition and assumes

Table 1. Results of Value-Guided Construal Models Fitted to Movements Over the Course of 12 Training and Eight Test Mazes for All Participants in Both Conditions of Experiment 1

Model	$\lambda_{complexity}$	λ_{switch}	ε_{move}	*df*	AIC	△AIC
No complexity or switch cost	0.00	0.00	0.1	1	132,094	2,804
Only complexity cost	5.42	0.00	0.1	2	130,535	1,245
Only switch cost	0.00	8.54	0.1	2	130,057	767
Both complexity and switch cost	1.69	5.33	0.1	3	129,289	0

Note: We compared the fit of models with or without nonzero complexity or switch cost coefficients. Responses were best explained by a model that included both costs, based on lowest Aikaike information criterion (AIC) score, which penalizes for number of parameters. All models were also fitted with separate ε-greedy policies to capture variation in maze movements.

that those participants were engaged in the more costly blocks-and-notches construal strategy. Could mechanisms other than construal explain why they were more likely to visit critical notches? One alternative is that rather than adopting the more complex construal strategy, participants trained on notch-necessary mazes learned to apply a simple rule such as "go to the closest notch" to solve each maze. This possibility would undercut the evidence supporting our hypothesis, so we designed Experiment 2 to rule it out. Specifically, we leveraged the fact that simple procedural heuristics such as "go to the closest notch" are insensitive to whether following them makes sense given the overall structure of a maze, whereas value-guided construal will lead to visiting a notch only if it is useful in the context of other maze objects. Thus, for Experiment 2, we examined participants' behavior on modified variants of the original test mazes that were designed to dissociate the predictions of a construal-based account from following simple rules.

Method

We recruited 300 participants from the Prolific experimental platform. The instructions, payment scheme, practice, and posttask questions were the same as in Experiment 1, but all participants received notch-necessary training sequences followed by one of the three test sequences: the original test mazes in which the shortest path passed through a critical notch (notch optimal), ones in which an equally short path was available that did not pass through a notch (notch optional), and ones in which the shortest path did not pass through any notch (notch suboptimal). All the mazes were designed such that the block containing the critical notch from the original test mazes was unchanged (Fig. 3). This allowed us to separate the effect of the critical notch on its own from its role in the overall context of the other blocks. A simple rule would be insensitive to this distinction because it would produce a plan to satisfy a local criterion (such as "get to the closest notch"), whereas value-guided construal would be sensitive to these modifications because the optimal construal and plan are a function of the global structure of each maze. Additionally, although participants received randomly flipped or rotated versions of the test mazes, we did not include mazes where start/goal states were swapped to ensure that the critical notch in notch-suboptimal mazes would be directly accessible from the start state.

Results

Although the specific motivation for Experiment 2 was to rule out the use of simple rules, we also wanted to assess whether the model could capture movements and reaction times even in this modified setting. To generate quantitative predictions based on our account, we simulated construal selection, planning, and maze behavior on trials using the model described in Experiment 1. Specifically, we used the best-fitting set of parameters from Experiment 1 that included nonzero complexity and switch costs to simulate 20 runs for each of the 294 participants in Experiment 2 (the exclusion criteria were the same as in Experiment 1; 162 male, 131 female, one nonbinary; age: M = 36 years, range = 18–75). For each run, we recorded whether a notch was visited as well as the complexity cost. The results, averaged over all simulations by trial and test condition, are plotted in Figure 4. The analysis of participant notch visitation and first-move reaction times reveals a qualitative correspondence between data and predictions.

Our first statistical analysis examined critical notch visitation in the different test conditions to assess the presence of a simple heuristic strategy in the notch-necessary training mazes. Using a logistic regression model, we found that, consistent with value-guided construal and inconsistent with a simple heuristic strategy, participants were more likely to visit

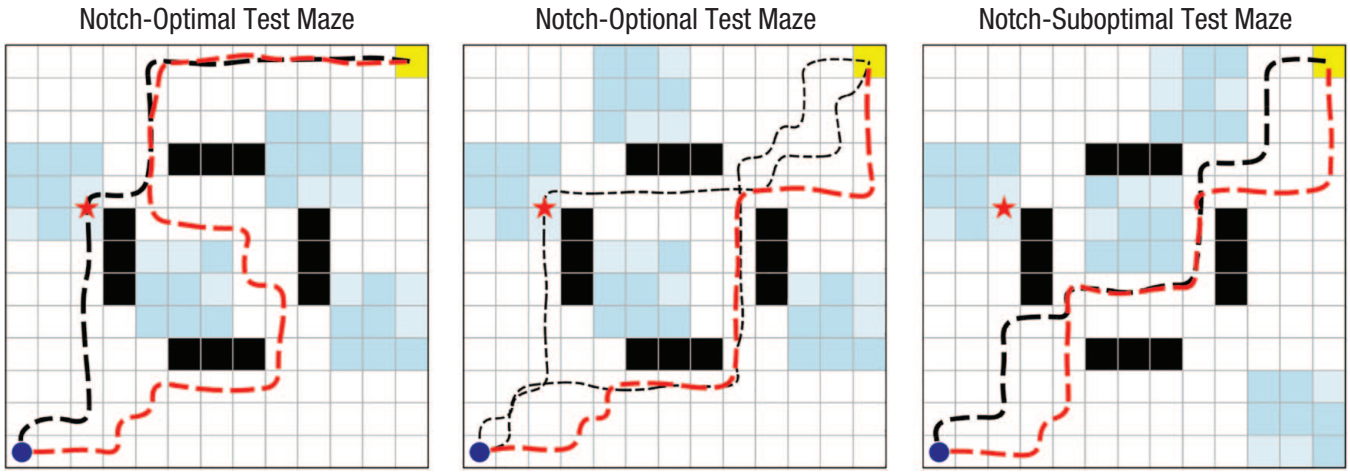

Fig. 3. Test mazes in Experiment 2. After completing the notch-necessary training sequence from Experiment 1, each participant completed one of three sets of eight test mazes: the original test mazes with the critical notches (notch optimal), test mazes modified so that an equally short path could be taken without visiting the critical notch (notch optional), and test mazes modified so that the critical notch was on a path that was longer than the shortest (optimal) path (notch suboptimal). The red star marks the position of the critical notch from the original test maze used for this example and its equivalent notches in the new conditions. Black lines indicate the shortest paths using a blocks-and-notches construal strategy, whereas red lines indicate the shortest paths using a blocks-only construal strategy. Note that in the notch-optional and notch-suboptimal conditions, there are equally good paths without using any notches (critical notch included); however, in the notch-optional condition, there is also an equally good path with the critical notch that can be used to assess any biases toward its use when pathway lengths are equal.

critical notches in notch-optimal compared with notch-optional test mazes—likelihood-ratio test with notch-optimal/notch-optional factor sum-coded as .5 / −.5: $\chi^2(1) = 715.71$, $p = 1.1 \times 10^{-157}$; $\beta = 3.16, SE = 0.14$. Similarly, participants were more likely to visit critical notches in notch-optimal compared with notch-suboptimal test mazes—likelihood-ratio test with notch-optimal/notch-suboptimal factor sum-coded as .5 / −.5: $\chi^2(1) = 1,485.27$, $p < 2.0 \times 10^{-16}$; $\beta = 7.27, SE = 0.51$. These results confirm that following a simple rule such as "go to the closest notch" does not explain behavior in the notch-necessary training condition of Experiment 1. Moreover, as can be seen in Figure 4, the overall qualitative patterns in the data are captured by model simulations.

Our second set of statistical analyses examined log-transformed first-move reaction times as a proxy for planning time and complexity cost. These analyses did not directly address the possibility of simple rule confound but could further corroborate the complexity costs posited by value-guided construal model, as in the analyses for Experiment 1. Additionally, by comparing reaction times between the different test conditions, we could determine whether participants in the notch-optional or notch-suboptimal conditions shifted away from a blocks-and-notches strategy to a blocks-only strategy. In particular, if participants in these conditions had faster first-move reaction times compared with those in the notch-optimal condition, this would indicate a shift away from a blocks-and-notches strategy to a less costly blocks-only construal strategy, which is another distinctive prediction of our account.

Using a linear model, we found that participants spent more time on their first move in the notch-optimal test mazes compared with the notch-optional mazes—likelihood-ratio test with notch-optimal/notch-optional factor sum-coded as .5 / −.5: $\chi^2(1) = 42.96$, $p = 5.6 \times 10^{-11}$; $\beta = 0.25, SE = 0.04$. Similarly, participants spent more time planning in the notch-optimal compared with the notch-suboptimal test mazes—likelihood-ratio test with notch-optimal/notch-suboptimal factor sum-coded as .5 / −.5: $\chi^2(1) = 139.24$, $p = 3.9 \times 10^{-32}$; $\beta = 0.47, SE = 0.04$. These results provide additional evidence that planning over more complex representations is costly (at least in terms of time spent planning).

General Discussion

Here, we evaluated the hypothesis that functional fixedness reflects the avoidance of complexity and switching costs during planning. To do so, we developed a novel paradigm in which participants navigated mazes that could be represented simply as blocks or more complexly as blocks and notches. Experiments revealed that people simplify problems (e.g., by adopting a blocks-only construal strategy if navigating through notches was unnecessary) and that they persist in these strategies (e.g., continuing to ignore notches even when attending to a notch would lead to a better solution).

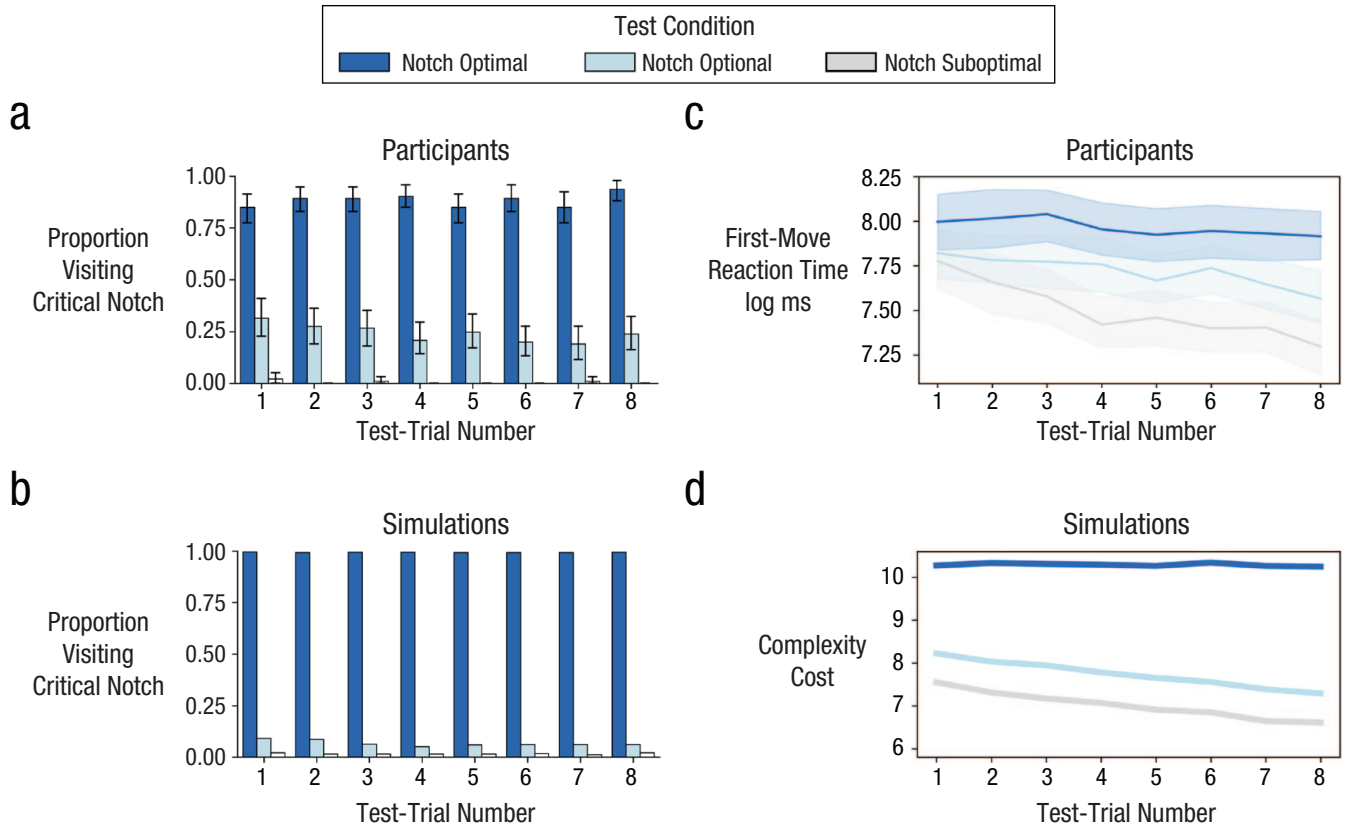

Fig. 4. Experiment 2 simulations and results. (a, b) Each participant (N = 294) received one of three sets of test mazes. Notch-optimal mazes were used in Experiment 1, whereas the notch-optional and notch-suboptimal mazes were modified versions of the original test mazes in which visiting the critical notch became either optional or suboptimal, respectively (see the text). Using the estimated parameters from Experiment 1, we simulated sequences of construals, plans, and actions on the trials that participants received (20 per participant) and recorded the proportion of simulated participants that visited critical notches. Error bars are bootstrap estimated 95% binomial confidence intervals. Simulated critical notch visitation matched qualitative patterns in actual participants behavior. (c, d) The simulations also predict how complexity costs will change over the course of test trials. Note that for the notch-optional and notch-suboptimal conditions, complexity costs decrease because one can switch to the less costly blocks-only strategy and still solve the mazes. Participants' log-transformed first-move reaction times—a proxy for time spent planning movements on a trial—qualitatively parallel predicted complexity costs. Error bands are 95% confidence intervals.

Additionally, our computational analyses using the value-guided construal framework (Ho et al., 2022) confirmed that the avoidance of complexity and switching costs explains observed patterns of optimal behavior, suboptimal behavior, and reaction times under different experimental manipulations. Overall, these results support our proposal and help clarify the computational principles that underlie functional fixedness.

This work integrates ideas from task switching and planning, considering them within a single, formally rigorous framework. It also highlights important, as yet not fully answered questions, such as the mechanistic underpinnings of switch costs. Previous work suggests several possible sources of switch costs, including processing specific to engagement of a new task set and/or interference from the previously active one (Grange & Houghton, 2014; Vandierendonck et al., 2010). Alternatively, switch costs may reflect delayed adaptation of model-free reinforcement-learning mechanisms (Dayan & Niv, 2008; Sutton & Barto, 2018; see Cushman & Morris, 2015, for this idea applied in the context of subgoal choice). We hope that by exploring task switching in the context of construals and planning, the work presented here can facilitate efforts to understand switch costs in other settings.

Additionally, we note several limitations of the current work. In particular, we tested only online participants from the United States and United Kingdom on maze-navigation tasks in which functional fixedness could be experimentally detected at a specific point in the trial sequence. Further studies with other populations, on more varied sequences of tasks (e.g., those that require more switching between construal strategies), and with more varied tasks (e.g., complex physical and social problem-solving tasks) will be essential for testing the generalizability of our account.

The current work also demonstrates a new approach to studying classic problem-solving phenomena such as functional fixedness. From an experimental perspective, our paradigm overcomes several limitations of standard tasks that have been used to study these effects, such as the candle task (Duncker, 1945) or nine-dot problems (Maier, 1930), where functional fixedness effects rely on idiosyncracies of how problems are initially presented to participants (Metcalfe, 1986; Sternberg & Davidson, 1995). In contrast, our blocks-and-notches design permits better experimental control using a trial-based design congruent with approaches used in contemporary learning and decision-making research (Daw et al., 2005; Dayan & Niv, 2008; Wilson & Collins, 2019). From a theoretical perspective, our results demonstrating conceptual rigidity can be contrasted with procedural rigidity, embodied in studies of Einstellung effects (Binz & Schulz, 2023; Luchins & Luchins, 1959), action chunking (Huys et al., 2015), and policy priors (Allen et al., 2020). In cases of conceptual rigidity, the problem-solver limits which causal affordances of a problem they are attending to when planning, whereas in cases of procedural rigidity, they limit themselves to precompiled action routines or a prior plan to solve a problem. A key direction for future work will be characterizing the interaction of these varied manifestations of rigidity. We hope that our account provides a useful perspective on these phenomena and deeper insights into the peculiar combination of flexibility and rigidity that comprise people's problem solving.

Transparency

Action Editor: Lasana Harris
Editor: Patricia J. Bauer

Author Contributions

Mark K. Ho: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

Jonathan D. Cohen: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

Thomas L. Griffiths: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was funded by the National Science Foundation (Grant No. 1545126), John Templeton Foundation (Grant No. 61454), and Air Force Office of Scientific Research (Grant No. FA 9550-18-1-0077).

Open Practices

This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976231200547

References

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences, USA*, *117*(47), 29302–29310.

Anderson, J. R. (1990). *The adaptive character of thought*. Erlbaum.

Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science*, *15*(9), 610–615.

Binz, M., & Schulz, E. (2023). Reconstructing the Einstellung effect. *Computational Brain & Behavior*, *6*, 526–542.

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences, USA*, *112*(45), 13817–13822.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196.

Duncker, K. (1945). On problem-solving (L. S. Lees, Trans.). *Psychological Monographs*, *58*(5), i–113.

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.

Grange, J. A., & Houghton, G. (2014). Task switching and cognitive control: An introduction. In J. A. Grange & G. Houghton (Eds.), *Task switching and cognitive control* (pp. 1–26). Oxford University Press.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature*, *606*, 129–136. https://doi.org/10.1038/s41586-022-04743-9

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, USA*, *112*(10), 3098–3103.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*(3), 374–419.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534–1555.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311.

Luce, R. D. (1959). *Individual choice behavior*. John Wiley.

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung*. University of Oregon Press.

Maier, N. R. F. (1930). Reasoning in humans. I. On direction. *Journal of Comparative Psychology*, *10*(2), 115–143.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.

Mattar, M. G., & Lengyel, M. (2022). Planning in the brain. *Neuron*, *110*(6), 914–934.

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 623–634.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, *25*(2), 117–129.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future of a cognitive research paradigm. *The Journal of Problem Solving*, *5*(1), Article 7. https://doi.org/10.7771/1932-6246.1144

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Sternberg, R. J., & Davidson, J. E. (1995). *The nature of insight*. MIT Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272.

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, Article e49547. https://doi.org/10.7554/eLife.49547

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314.