Cognition 243 (2024) 105669

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

People's thinking plans adapt to the problem they're trying to solve

Joan Danielle K. Ongchoco [*], Joshua Knobe, Julian Jara-Ettinger

Department of Psychology, Yale University, New Haven, USA

ARTICLE INFO

Keywords:
Computational modeling
Thinking
Decision-making

ABSTRACT

Much of our thinking focuses on deciding what to do in situations where the space of possible options is too large to evaluate exhaustively. Previous work has found that people do this by learning the general value of different behaviors, and prioritizing thinking about high-value options in new situations. Is this *good-action bias* always the best strategy, or can thinking about low-value options sometimes become more beneficial? Can people adapt their thinking accordingly based on the situation? And how do we know what to think about in novel events? Here, we developed a block-puzzle paradigm that enabled us to measure people's thinking plans and compare them to a computational model of rational thought. We used two distinct response methods to explore what people think about—a self-report method, in which we asked people explicitly to report what they thought about, and an implicit response time method, in which we used people's decision-making times to reveal what they thought about. Our results suggest that people can quickly estimate the apparent value of different options and use this to decide what to think about. Critically, we find that people can flexibly prioritize whether to think about high-value options (Experiments 1 and 2) or low-value options (Experiments 3, 4, and 5), depending on the problem. Through computational modeling, we show that these thinking strategies are broadly rational, enabling people to maximize the value of long-term decisions. Our results suggest that thinking plans are flexible: What we think about depends on the structure of the problems we are trying to solve.

1. Introduction

We often have to make decisions involving a wide array of options: actions we can take, things we can say, or people we can interact with. Intuitively, much of our thinking involves considering different options in anticipation of a decision, as when we think about what we want to eat for dinner, or what we want to say before a meeting. But we don't have the time and computational resources to consider every option before we have to make a choice (see Lieder & Griffiths, 2017, 2020; Vul, Goodman, Griffiths, & Tenenbaum, 2014). So we face a complex selection problem: Given that we can only think a few thoughts at a time, what should we think about?

A growing body of computational and empirical work suggests that people solve this problem through a 'good-action bias' (Bear, Bensinger, Jara-Ettinger, Knobe, & Cushman, 2020; Icard, Kominsky, & Knobe, 2017; Mattar & Daw, 2018; Morris, Phillips, Huang, & Cushman, 2021): Over time, people learn the values associated with different options or behaviors in a broad class of situations. When people face a specific situation and can only consider a small number of options about how to behave within it, they tend to consider the options that they have determined to generally have high value in that broader class of situations. This process then leads to increasingly more accurate representations of the high-value options (e.g., Braun, Wimmer, & Shohamy, 2018; Gelly & Silver, 2011; Icard, Cushman, & Knobe, 2018).

Despite the power and usefulness of a good-action bias, this strategy is only useful in situations where people have already computed and stored representations of the values of the different possible options (i.e., "model-free", as opposed to "model-based" values; Gläscher, Daw, Dayan, & O'Doherty, 2010). Yet, many important decisions often come in the context of novel situations, where relevant past experiences can be scarce, or carry little value information for the context that we're in. Consider, for instance, when you move to a new country: You might have some knowledge of how things generally work, but you might not yet have stored values for different options (e.g., of phone plans, restaurants, etc.). In these situations, what do people tend to think about?

1.1. What do we think about?

There are at least two possibilities for what happens when people face new situations. A first possibility is that the good-action bias reflects

* Corresponding author.
E-mail address: joan.ongchoco@psych.ubc.ca (J.D.K. Ongchoco).

https://doi.org/10.1016/j.cognition.2023.105669
Received 27 May 2023; Received in revised form 13 November 2023; Accepted 16 November 2023
Available online 30 November 2023
0010-0277/

a thinking strategy that people use across the board. When people encounter a new type of situation, they might quickly build an initial estimate of value based on superficial cues and focus on the options they estimate to have high value. For example, suppose a person is confronted with a new problem and different possible options. One strategy might be to obtain broad value estimates (possibly based on low-level features), and then prioritize thinking about whichever option is more likely to be high value. In doing so, she would of course think about both the good and bad outcomes that could arise from those options (Ito & Cacioppo, 2005; Lieder, Griffiths, & Hsu, 2018; Rozin & Royzman, 2001; Tversky & Kahneman, 1991), but she would tend not to devote her limited computational resources to thinking at all about the options she estimates to be of lower value.

At the same time, novel situations can be risky when bad choices can produce particularly bad unforeseen outcomes. In these cases, it may be beneficial to also think about potentially low-value options. This raises a second possibility: The good-action bias might be limited in scope, and people may be able to generate and use different thinking strategies based on the context that they're in. In other words, in the same way we can create ad-hoc action plans in a novel physical task (as when we figure out which path to take based on the layout of a new room), people may also be able to create ad-hoc "thinking plans" in a novel thinking task. Thus, people might be able to execute thinking plans that are not necessarily aligned with the good-action bias, depending on the problem they are attempting to solve.

The present study aims to explore these possibilities. How do people know what to think about in novel situations? Do their strategies reflect a generalized form of the good-action bias? Or can people flexibly adapt to the right strategy based on the structure of the problem at hand?

1.2. The present studies

To answer these questions, we developed an experimental paradigm that allows us to determine what people have (and have not) thought about when faced with an impending novel situation. Participants saw novel incomplete block-puzzles and an array of six different puzzle-piece options (Fig. 1a), some of which had potentially high values, while others had potentially low values. Puzzle pieces varied in their surface features (e.g., their size), which provided information about their apparent value. Participants could then solve the puzzle by mentally rotating the various puzzle pieces to see which pieces fit and which did not, which in turn revealed their value. Participants always first went through this *thinking phase*, where they could freely think about different puzzle piece options (via mental rotation; Fig. 1a).

From here, we used two distinct response methods. First, this paradigm allowed us to ask people in a self-report method what puzzle pieces they thought about. We designed an interactive layout in which participants could click different puzzle pieces that they had mentally rotated in the thinking phase. Second, to rule out potential task demands, we also used a more implicit measure of what people thought about. In a *decision phase*, participants were explicitly prompted to select the best piece from a subset of pieces (Fig. 1b). Because mental rotation is a relatively slow cognitive operation (Cooper, 1975; Shepard & Metzler, 1971), this paradigm also enabled us to probe whether participants

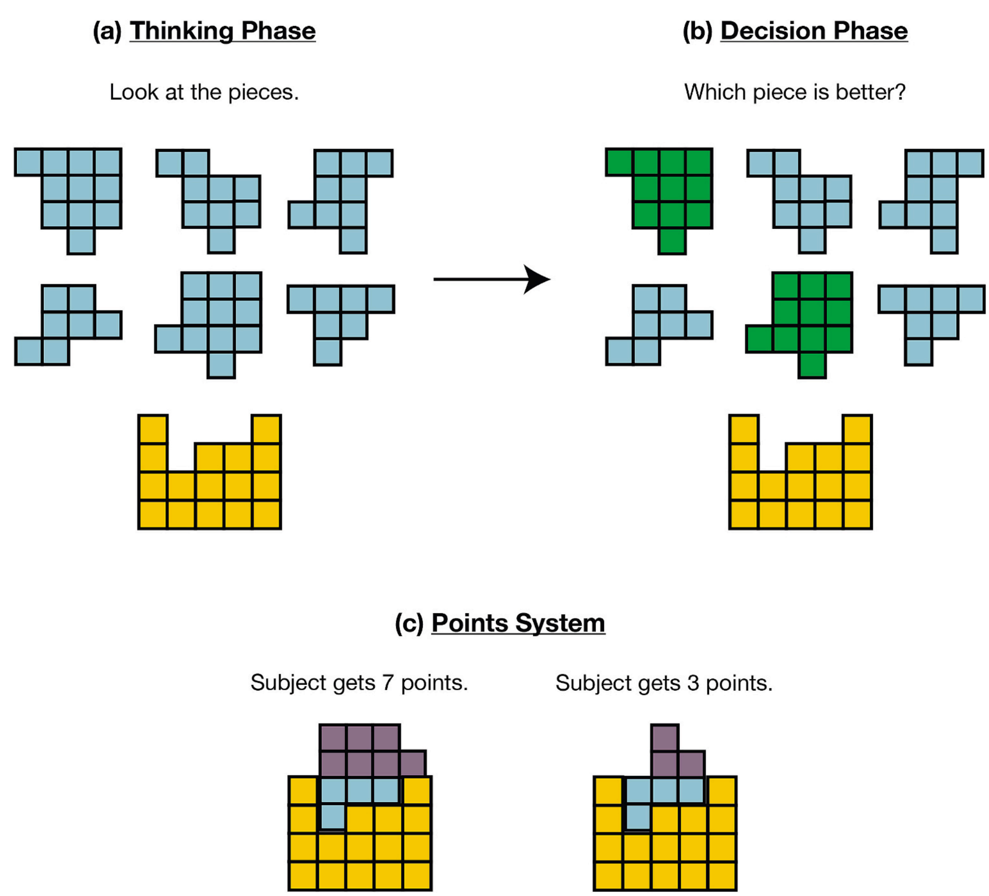

Fig. 1. The two-part design of the implicit response time method experiments. Participants first went through a (a) thinking phase where they were allowed to freely evaluate how well each of the six pieces fit into the base (bottom piece) (b) Decision phase where participants were asked to choose between two pieces. (c) Points were calculated based on the additional squares (colored purple) above the puzzle, when the puzzle piece was slotted in. On the left side, the subject gets 7 points corresponding to the 7 squares above the puzzle. On the right side, the subject gets 3 points corresponding to the 3 squares above the puzzle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

had previously thought about a puzzle piece by measuring their response times in a subsequent decision phase (Fig. 1b). A slow response time would suggest that the participant is evaluating the puzzle piece (i.e., performing a mental rotation and computing its value) only after being asked about it, while a fast response time would suggest that the participant had already mentally rotated the piece and computed its value before being asked. Therefore, participants' response times can reveal whether they had already considered and cached the values of the pieces before they were presented to them in the decision phase: Participants should be faster when deciding between two pieces that they had already evaluated, and slower when they had not thought about some (or all) of the presented pieces.

This paradigm enabled us to test what people think about in novel situations, as a function of the context that they're in. In Experiments 1 and 2, we first considered a situation where observable cues contained information about the potential value of different pieces, and we tested whether these cues influenced what people chose to think about first. Participants were told that only puzzle pieces that fit into the main structure would give them points, and that the object's final size (see Fig. 1c), after attaching the puzzle piece, would determine the number of points that they could obtain. Therefore, participants could estimate the potential value of a piece based on its size, but had to think about (and mentally rotate) it to discover its exact value.

To see if people don't just default to potentially high-value options, we also needed to identify situations where a better strategy would be to think of potentially low-value options (therefore acting against a good-action bias). (And we will note that there may be some situations in everyday life in which it could be helpful to do so, such as in delicate social situations — where ultimately the space of possible options can change and be significantly reduced based on what another person says, and so we need to know not just which options to act on, but also which options to *avoid* so we do not make a situation worse.)

To explore this possibility, we formalized our block-puzzle paradigm in a computational framework and varied different task parameters to explore the space of decision problems. This allowed us to identify contexts in which the best strategy would be to think about potentially low-value options first, which we then tested in Experiments 3 and 4.

Finally, given the evidence we find that people deploy different thinking strategies depending on the problem, Experiment 5 explored whether these different strategies emerge at the individual level, testing if people adjust their thinking strategy as the problem structure changes. Altogether, these experiments allowed us to probe what people would think in a range of decision problems, varying across different dimensions: a case where there was an equal number of high and low-value options versus a case where there were more low-value options; a case where there was only one high-value option (needle in the stack) versus a case where there was a really low-value option (snake in the stack); and a case where the critical low-value option (the “snake”) could be more or less difficult to find.

2. Computational framework

To formalize the problem of which options would be best to think about, we used partially-observable Markov decision processes (POMDP; Cassandra, 1998; Sutton, Barto, et al., 1998). POMDPs originally rose to prominence in robotics and AI to model action-planning in complex spatial environments under partial information, but more recent work has shown that this framework can be used not only to model decisions between different physical actions but also decisions between different thought processes (see Callaway et al., 2021; Chen, Chang, & Howes, 2021; Griffiths et al., 2019; Lieder & Griffiths, 2020). Here we adapted the framework to implement a space of thinking actions rather than physical actions. Before introducing our model, we will begin by briefly introducing POMDPs in the context in which they are classically used.

2.1. Classical uses of POMDPs

To illustrate the logic of POMDPs, consider a simple situation where an agent's goal is to take an object in a house (also see Fig. 2a, b). To achieve this, the agent must first find the object and then retrieve it. To formalize this problem, POMDPs define a state space S consisting of all possible physical states of the world. In a case like this one, the state space would include the combination of any position where the agent and the object might be at any given time point. Given this state space, the agent's goals can be represented as a reward function R that assigns a numerical reward to combinations of states and actions. In this example, the goal can be encoded as a reward function that returns a high positive value in states where the agent is holding the object.

To obtain these rewards, the agent can take sequential actions (from a set of actions A) that change the state of the world. The relationship between actions and states is captured by the transition function T, where $T(s,a,s^0)$ represents the probability that the world will change from state s to s^0 when the agent takes action a. For instance, an agent taking the action ‘walk north’ in a state should assign a high probability to the state where the agent is now one spot north of where they used to be.

At their core, POMDPs assume that agents can have partial or incomplete knowledge about the world. For instance, the agent may know their position in space but not know where the object is located. To achieve this, POMDPs introduce belief representations, expressed as probability distributions over the state space. To model how the agent's beliefs change as they move in space, POMDPs introduce an observation function O which determines what information is made available to the agent in different states, where $O(i,s,a)$ is the probability that the agent receives information i when taking action a in state s. For instance, a simple observation function might encode that the agent can see whether the object is present or absent in any room as soon as she enters it. Formally, this is achieved by defining a space of observations, and specifying which observations are associated with each state through an observation function (which can include probabilistic components).

Given the six-tuple defined above—a state space, an action space, an observation space, a reward function, a transition function, and an observation function—it is possible to compute the series of actions that, given an agent's knowledge, maximize the long-term rewards that the agent obtains (requiring one additional parameter, λ, that specifies how rewards are discounted over time). Computing the exact solution to a POMDP is computationally demanding and often intractable in practice, particularly in problems with large state spaces. Nonetheless, research in the past two decades has led to the development of multiple algorithms that provide approximate solutions to POMDPs (such as by not computing the actions that would be associated with implausible belief states that the agent could have), making them a useful practical framework for determining rational action under imperfect information (Hsu, Lee, & Rong, 2008; Kurniawati, Hsu, & Lee, 2008; Ng & Jordan, 2000).

2.2. Modeling thinking through POMDPs

Although POMDPs are most typically used to model choices between physical actions, they can also be used to model choices between different *thoughts* (see Fig. 2c). For simplicity, we explain this model structure in the context of Experiment 1a (a more comprehensive presentation is available in the Supplemental Materials). Here, participants were presented with a puzzle like the one shown in Fig. 1a, and they learned that pieces that do not fit into the puzzle have no value, and pieces that fit into the puzzle have larger values whenever they increase the overall number of blocks in the structure. After being given time to think about whichever pieces they liked (i.e., rotating them mentally to see if they fit in the puzzle during the thinking phase), participants were asked to quickly determine which of two puzzle pieces had a higher value (decision phase). Although the full problem is ultimately encoded

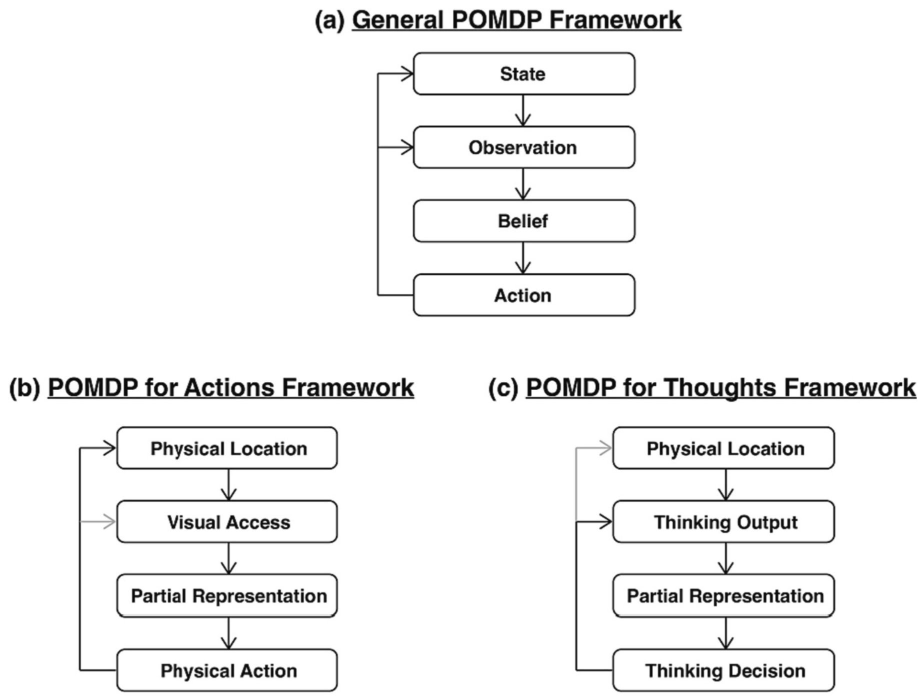

Fig. 2. High-level schematic of how we use POMDPs in our approach. (a) classic POMDP graphical model. States produce observations, which affect an agent's beliefs, the agent then chooses subsequent actions based on their beliefs, which causally affect the state of the world and the observations they receive. (b) One example of a standard use of POMDPs. The state is the physical location of the agent, the observations are determined by the agent's visual access (e.g., receiving information that is in their line of sight). These observations update the agent's beliefs (e.g., seeing a locked door updates their representation of the door), and the agent then takes a new physical action (such as walking in a particular direction), which changes the physical location. The link between physical movement and visual access is not necessarily critical, as this information is often already encoded in the representation of the physical location. (c) Simplified formulation of our approach. Here, the space of physical actions is replaced with a set of thinking decisions. These decisions therefore do not affect the state of the world, but have a direct impact on the thinking output. Under this formulation, the observations are the result of thinking, rather than the result of a purely external physical state. Note that, in reality, actions affect both states and observations and these are simplified representations to clarify the use of POMDPs for thinking.

in a single POMDP, we first explain the components at use during the thinking phase, and then turn to the ones used in the decision phase.

In this context, the potential value of a puzzle piece can be determined visually: The larger the piece, the more likely it is to be valuable. The true value of a piece, however, can only be revealed by thinking about it (i.e., mentally rotating it to test if it fits in the puzzle or not). Thus, our model represents each piece in terms of its *potential* value, determined by its size, and its *true* value, which equals the potential value when the piece fits into the puzzle, and 0 when it does not. Given the two hypotheses about each piece (the true value either equals the potential value or 0), we defined the state space as every possible setting over which pieces' true value matches the potential value and which do not (i.e., in the case with six pieces, the state space consists of $2^6 = 64$ states).

To model the thinking phase, we gave our model the ability to execute *thinking* actions, which revealed the true value of whichever piece the agent chose to think about (through an observation). While the thinking actions do not have any causal impact on the state space, there was always a small probability that, at any given point, the state space may switch to a decision phase (capturing the idea that the participant knew that at any point they might get asked to choose a piece among a subset).

To model the decision phase, we included an additional set of states, where each state encoded a forced choice between two possible pieces selected from a uniform distribution (as in Fig. 1a). To select a piece, the agent could take a 'selection' action, obtaining a reward depending on which piece they picked. Thinking actions could be performed both during the thinking and the decision phases, and were always costly (see Supplemental Information for details).

These specifications enabled us to use the POMDP framework to compute what sequence of thinking actions was best suited to the decision problem, with the goal of maximizing the agent's expected reward in the decision phase. Under this formulation, the model is pressured to optimize its thinking plan due to two forces. The first is that thinking is costly (which pressures the model to minimize its thinking actions when possible), and the second is that rewards are temporally discounted (an intrinsic feature of POMDPs) such that, upon entering the decision phase, the model prefers responding sooner rather than later.

3. Experiment 1: a good-action bias in explicit thought reports

We first sought to test what people think about in a novel situation where the potential value of different choices can be estimated based on superficial observable cues. Participants saw a display like the one shown in Fig. 1a. The value of each piece was given by the final size of the completed puzzle, after the piece was attached (and 0 value for pieces that did not fit the base puzzle). Thus, each piece's size gave a superficial cue about its potential value, but participants needed to mentally rotate each piece to test if it indeed fit the puzzle. We first describe the model and simulations, confirming that, in a context like this one, our model prioritizes thinking about potentially high-value pieces. We then validate the model's parameters and its predictions in a self-report paradigm, probing *how many* pieces people can mentally simulate in the given period of time, as well as *which* pieces they mentally simulate.

3.1. Model simulations and results

To determine what people should think about in this situation, we implemented this puzzle in our computational framework (section 2.2). Because people can sometimes make reasoning errors, our model included a small probability that a mental rotation would lead to an incorrect conclusion (i.e., believing that a piece fits when it does not, or believing that a piece does not fit when it does). To achieve this, we included a certainty parameter, which we varied from 50% (i.e., a 50% chance of making a reasoning error) to 100% (i.e., full confidence of no reasoning error) in intervals chosen to match the empirical error rate from our studies (see Supplementary Materials). We set the probability of switching to the decision phase as 0.37 (a parameter we validated from the explicit self-report experiment we ran below). We used a future discount parameter of 0.95. These parameters were all set prior to model evaluation.

After solving each POMDP (i.e., computing the optimal policies), we obtained an expected thinking plan by running 100 simulations under each certainty parameter with a random set of puzzle pieces (set to always have two large pieces, two medium pieces, and two small pieces to match the experiment). For each of the 100 simulations, model predictions were computed as the average behavior across all certainty values. To ensure that all simulations revealed the full thinking plan, we modified the state transition dynamics to ensure that the model would not switch to the decision phase until after the model had the chance to think of all six pieces. That is, the model's solution reflected the belief that the agent might be prompted to decide before having had the opportunity to think about all six pieces, but the simulations were modified to stay in the thinking phase long enough for us to observe the model's full thinking pattern.

Fig. 3 shows the results from this simulation. As this figure shows, our model always uses the first 3–4 time steps to think about the two pieces with the highest probability of being valuable. Afterwards, the model shifts to thinking about medium-value pieces in steps 5 and 6. This reflects the performance of the models that believe reasoning errors are likely. In these cases, our model believes it is better to double-check potentially high-value pieces to confirm their true value. Finally, the model only begins considering the medium and maybe, the lowest-value pieces at the very end of the task.

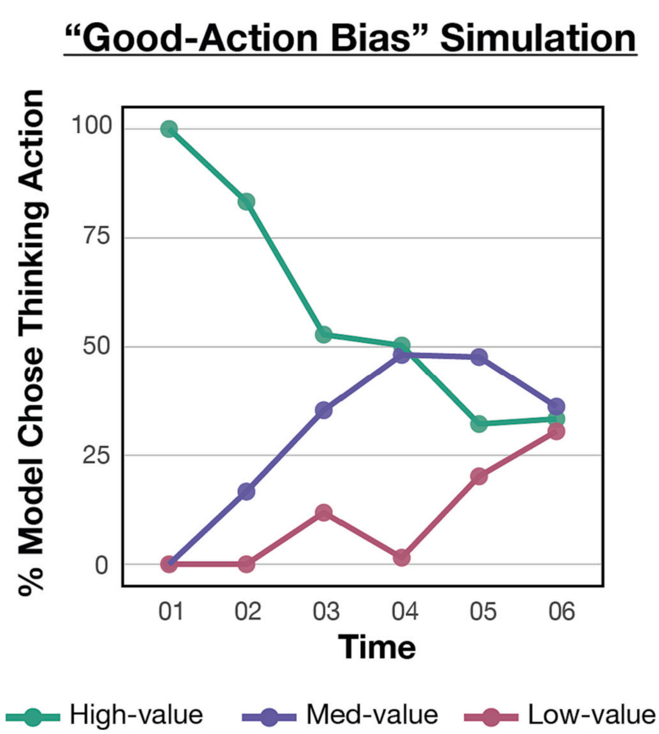

Fig. 3. Model predictions about what pieces to think about as a function of time step in Experiment 1. The model could think about the potentially high-value pieces (green), the medium-value pieces (purple), or the low-value pieces (red). Our rational model suggests that the best strategy is to think about the high-value pieces. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. *Explicit self-report method*

Here, we tested the situation we had set up in the model, where the value of pieces was correlated with size: Larger pieces that fit into the puzzle were more valuable than smaller pieces that fit into the puzzle. However, pieces that did not fit into the puzzle had no value, regardless of their size. In this initial experiment, we used a self-report method in which participants were explicitly asked which options they had thought about, and we used the average number of pieces that participants thought about to align the model's expected number of pieces it could evaluate.

3.2.1. *Method*

All methods and analyses were pre-registered (https://aspredicted.org/sh65h.pdf). Data and code for all experiments reported here are available on: https://osf.io/n8em7.

Participants. 150 participants were recruited on the online Prolific platform and completed a 5-min single-trial experiment for monetary compensation. The saomple size was determined based on a power analysis run on pilot data.

Apparatus. After agreeing to participate, subjects were redirected to a website where stimulus presentation and data collection were controlled via custom software written using a combination of HTML, CSS, JavaScript, PHP, and jsPsych libraries (de Leeuw, Gilbert, Petrov, & Luchterhandt, 2023). Subjects completed the experiment in fullscreen mode on either a laptop or desktop computer.

Stimuli. Each trial consisted of a yellow puzzle base and six blue puzzle pieces (see Fig. 1a for an example). Each participant was presented with a randomly generated puzzle base, which was always a rectangle with three to four blocks missing at the top. The goal was simply to figure out whether the puzzle pieces would fit, such that the puzzle piece, when rotated and positioned into the puzzle, would form a shape without any holes within the 4 $\times$ 5 puzzle. Each of the six puzzle pieces options consisted of the three/four blocks that were missing, along with additional blocks that determined the value of the piece. For example, a piece with a value of 7 points is one that locks into the puzzle base after being rotated and has 7 additional blocks that go beyond the puzzle's 5 $\times$ 4 rectangular shape (e.g., the upper leftmost piece in Fig. 1a; the left side of Fig. 1c). Critically, some puzzle pieces did not have the correct shape to lock into the puzzle and therefore had value 0 (e.g., the lower middle piece in Fig. 1a, which has a potential value of 7, but its true value is 0 because it does not fit into the puzzle). The six puzzle pieces were randomly generated but always consisted of three pairs of potential values: two potentially high-value options (potential value: 7), two potentially medium-value options (potential value: 5), and two potentially low-value options (potential value: 3). In each pair, one piece would always fit (true value equals its potential value), and the other would not (true value equals 0).

Procedure and Design. Participants first read a brief set of task instructions where they learned the logic of the task and they were told that their goal was to earn as many points as possible. A ‘Total Points’ counter was visible on the top-left of the screen throughout the entire experiment to give participants the sense of “earning” points across the experiment (though the critical test trial only occurred at the very end after the instructions and comprehension questions were completed). Participants were then shown a sample block-puzzle, and were asked questions about these different options to test their understanding of the task and the point system. During these comprehension questions, participants obtained points for correct answers, which we used as an exclusion criterion for people who did not understand the task. To ensure that participants would not just look for puzzle pieces whose bottom part resembled the structure of the missing section of the puzzle base, they were told that the pieces could only be rotated but never simply flipped. After these instructions, the critical single trial of the experiment began. Participants were shown a new block-puzzle. At the start of the thinking phase, participants were told that six pieces would now appear above the puzzle, in two rows of 3 pieces each (as in Fig. 1a). Participants were told that they did not have to do anything but just look and study the pieces, and that they would simply be asked questions about the pieces afterwards. To incentivize participants to really think about the options during the thinking phase, they were also told that they would get bonus points for responding (i.e., "Try to answer as fast as you can. You'll get a bonus if you answer fast"), but this was not tied to any actual monetary reward. Once the puzzle pieces appeared, participants were given five seconds to study them, with a countdown timer shown above the puzzle. After five seconds, participants were asked to click on the pieces that they thought about. When they had clicked on all the pieces they had thought about, they pressed a key to complete the study.

3.2.2. Exclusions

Per the preregistered criteria, we excluded people who reported (in a debriefing phase) having an attention level below 70% ($n = 10$); whose total completion time was more than 2 standard deviations from the grand population mean ($n = 11$); and whose performance in the practice trials was below chance ($n = 49$).

3.2.3. Results

People reported thinking of an average of 2.66 pieces (as in Fig. 4a). Fig. 4b shows that people were more likely to think about a piece if it was a potentially high-value option than if it were *not* a potentially high-value option (permutation test over 10,000 permutations, $p = .004$).

3.2.4. Discussion

The results of this initial experiment are critical in two ways. First, participants' responses matched what the model predicts: participants were significantly more inclined to think of the potentially high-value pieces than medium or low-value pieces. This serves as at least some initial evidence in support of a good-action bias in play.

Second, we can use the average number of pieces people selected to set the model's probability of switching. Given that people could think of approximately 2.68 pieces within the 5-s period, we can compute a likelihood that the model will switch to a decision phase. This problem is mathematically equivalent to considering the expected number of consecutive heads on a biased coin until getting tails, with heads representing another 'thinking chance' and tails representing the switch to the decision phase. In this formulation, the expected number of flips is $n = 1/p$, where p is the probability of getting tails. In our case, we therefore used $p = 1/n$ with n set to 2.68, resulting in $p = .37$. This ensured that the model's policy reflects the rational thinking plan under the thinking constraints that participants faced.

Though the results suggest that participants prioritized thinking about potentially high-value options during the thinking phase, at the same time, participant responses were noisy, and a substantial proportion of people reported thinking about the medium or low-value pieces. Of course, some part of this may be a function of the self-report measure itself, such that some people may mistakenly report thinking about the other pieces as well, or may have misinterpreted the task (which we address later in Experiment 2). But a different possibility is that the noise in people's responses could reflect something about people's thinking process itself. Given the six puzzle pieces that are readily available (and presented visually and simultaneously together), the process of thinking about the optimal pieces, and not just considering each of the pieces with equal likelihood, may require the exertion of cognitive control (see e.g., Icard, 2018). This process of balancing the value of thinking of the potentially good pieces versus exerting control to *not* think about the other pieces may also contribute to the noisier responses that we see here.

4. Experiment 2: a good-action bias in implicit decision-making times

We validated our model with a task in which we asked people explicitly which pieces they had thought about, but it is possible that these self-reports were not completely accurate. For example, it might be that people did not know which options they were actually thinking of and simply responded by listing the options that they thought they *should* have been thinking about. To verify the results from Experiment 1, Experiment 2 used an implicit method of measuring what people thought about when there wasn't any decision to be made yet. After the thinking phase—where participants could freely focus on different pieces to check if they'd fit into the puzzle—participants were asked to select the best option among either two large pieces or two small pieces. We predicted that, if participants prioritize thinking about potentially high-value options, their response time should be significantly faster when asked to select which is better of two large pieces relative to when asked to select which is better of two small pieces.

4.1. Experiment 2a

4.1.1. Methods

This experiment was identical to Experiment 1, except as noted. Sixty

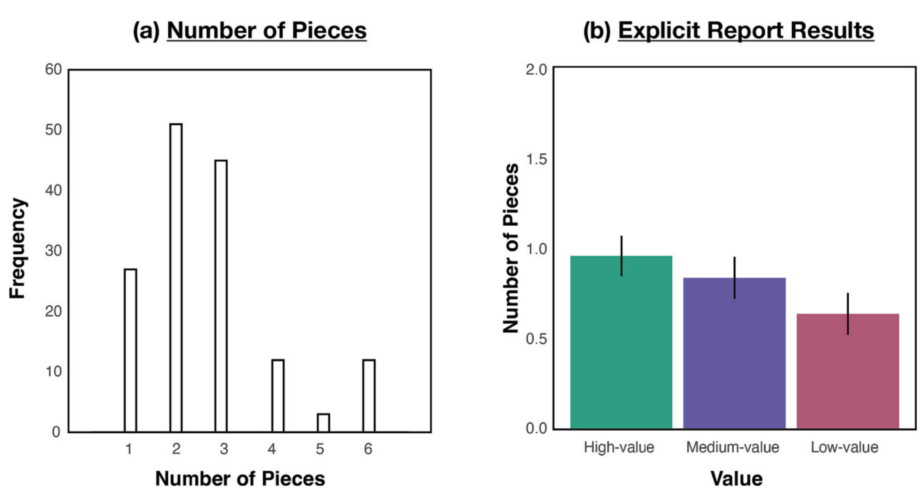

Fig. 4. Results from Experiment 1. (a) Histogram of how many pieces people selected. The x-axis depicts the total number of pieces per participant, with the maximum number of pieces being 6. (b) The average number of pieces people selected of a particular value, either: high, medium, or low. There were two possible pieces for each value. Error bars reflect 95% confidence intervals.

participants from the broader community participated in a single-trial experiment (conducted in the university library with participants who participated voluntarily for candy). The sample size was determined based on a power analysis run on pilot data (see Supplemental Materials for details). Stimuli were presented using custom software written in Python with the PsychoPy libraries (Peirce et al., 2019) and were displayed on a monitor with a 60 Hz refresh rate. Participants completed the study on a 13-in. MacBook Air with a 1440 × 900 resolution. This time, instead of having participants click on the pieces that they had mentally rotated, they instead switched to a decision phase, after the countdown timer for the thinking phase reached 0. Then, two of the six pieces turned green (as in Fig. 1b) and participants were asked "Which piece is better?" Critically, half of our participants were tasked with identifying which was better of the two potentially high-value options, and the other half were tasked with identifying which was better of the two potentially low-value options. Three participants were excluded because their mean performance in the comprehension questions was 2 standard deviations below the mean. These participants were replaced, until a total of 60 participants was reached. All methods and analyses were pre-registered (https://aspredicted.org/wd5it.pdf).

4.1.2. *Results and discussion*

Response accuracy and response times for the single trial were recorded for each participant. Only response times where participants responded correctly were included in the analysis. Participants who chose between the potentially high-value options responded faster in the decision phase ($M = 2.56$ s, $SD = 1.77$ s) than participants who chose between the potentially low-value options ($M = 5.20$ s, $SD = 3.82$ s), as depicted in Fig. 5a ($t(43.37) = 3.65$, $p < .001$, $d = 1.06$ over the logarithm of the response time). Including the incorrect answers did not yield any different results, $t(55.41) = 3.51$, $p < .001$, $d = 0.91$. There was no significant difference between the percentage of people who responded accurately when choosing between potentially high-value options vs. potentially low-value options (86.67% vs. 66.67%, Fisher's exact: $p = .125$). Thus, even after switching over from a self-report method to this implicit response time method, we still get the same basic effect: Participants' decision-making times suggest that during the thinking period, they prioritized and refined their representations of the potentially high-value options.

4.2. *Experiment 2b*

In Experiment 2a, potentially high-value options always had more blocks, while potentially low-value options always had fewer blocks. Might people simply have been attracted to larger pieces, independent of their value? To ensure that our results were not just a matter of the number of blocks, Experiment 2b flipped the size-value relationship: Smaller puzzle pieces were now more valuable than larger puzzle pieces.

4.2.1. *Method*

This experiment was identical to Experiment 2a, except as noted. Seventy new participants from the broader community participated. The sample size was determined before data collection began based on a power analysis run on pilot data (see Supplemental Materials for details). Participants first read a brief set of task instructions that explained the logic of the task. In contrast to Experiment 2a, participants learned that the number of additional blocks in a piece that fits now reflected the number of points that would be deducted. For instance, if a piece fit into the puzzle base and had 7 additional blocks, then 7 points would be deducted. If the piece did not fit, then 10 points would be deducted. In this case, the potentially high-value pieces were the ones with fewer blocks, and the potentially low-value pieces were the ones with more blocks (and thus, the difference in the number of points lost for getting a choice between the smaller piece options wrong versus the number of points earned for getting it right is higher than for choices between larger piece options). Participants began with a score of 50 points and the thinking and decision phase proceeded in the same way as Experiment 2a. Two participants were excluded because their mean performance in the comprehension questions was 2 standard deviations below the mean. These participants were replaced, until a total of 70 participants was reached. All methods and analyses were pre-registered (https://aspredicted.org/25ux2.pdf).

4.2.2. *Results and discussion*

Response accuracy and response times for the single trial were recorded for each participant. Only response times where participants responded correctly were included in the analysis. Participants who chose between the potentially high-value options responded faster in the decision phase ($M = 2.37$ s, $SD = 1.73$ s) than participants who chose between the potentially low-value options ($M = 3.99$ s, $SD = 2.23$ s), as shown in Fig. 5b ($t(40.37) = 3.88$, $p < .001$, $d = 1.08$ over the logarithm of the response time). Including the incorrect answers did not yield any different results ($t(65.24) = 4.14$, $p < .001$, $d = 0.99$). While we found a marginal difference between the percentage of people who responded accurately when choosing between potentially high-value options vs. potentially low-value options (71.43% vs. 91.43%, Fisher's exact: $p = .062$), accuracy was nonetheless high in both conditions. And this marginally higher when choosing between low-value options may reflect participants' determination to not make a wrong choice. Thus, our results suggest that people are not just attracted to thinking about

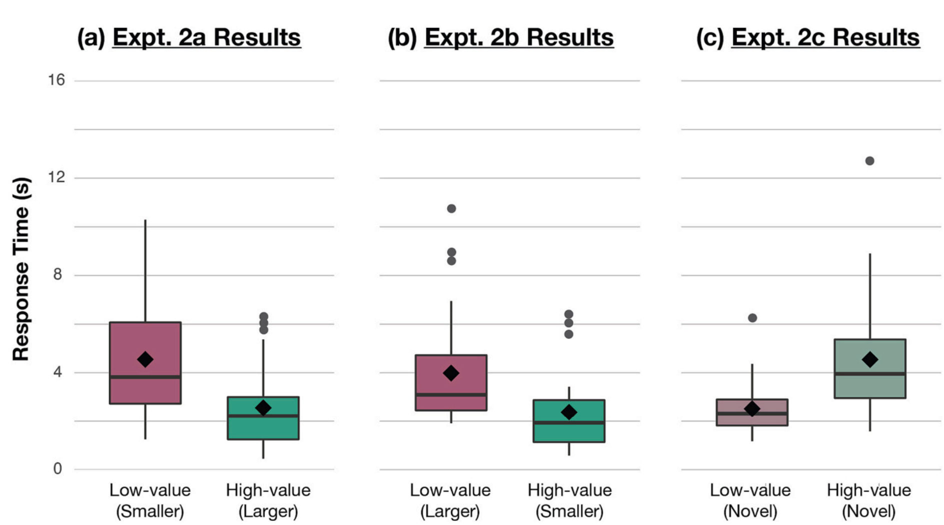

Fig. 5. (a) Results from Experiment 2a. Diamonds reflect means. (b) Results from Experiment 2b. (c) Results from Experiment 2c.

visually larger puzzle pieces, and they instead think about particular puzzle pieces based on the pieces' underlying values.

4.3. Experiment 2c

So far, the results from Experiments 2a-2b suggest that people prioritize thinking about potentially high-value options, even in novel situations. This strategy produced a response time benefit when people had to decide between two potentially high-value options, relative to when they had to decide between two potentially low-value options. To further demonstrate that this response time benefit is specific to puzzle pieces that participants mentally rotated during the thinking phase (recalling their value during the decision phase), we tested people on pieces that were not available during the thinking phase. This experiment would provide a baseline of how long people would take to decide between two pieces if they did not have the time to think about these pieces in an earlier thinking phase.

4.3.1. Method

This experiment was identical to Experiment 2a, except as noted. Sixty new participants from the broader community participated. The sample size was determined before data collection began based on a power analysis run on pilot data (see Supplemental Materials for details). During the decision phase, a new pair of high-value or low-value options were generated and presented to the participants. One participant was excluded because their mean performance in the comprehension questions was 2 standard deviations below the mean. This participant was replaced, until a total of 60 participants was reached. All methods and analyses were pre-registered (https://aspredicted.org/8mk78.pdf).

4.3.2. Results and discussion

Response accuracy and response times for the single trial were recorded for each participant. Only response times where participants responded correctly were included in the analysis. This time, participants who chose between potentially high-value options were in fact slower ($M = 4.54$ *s*, $SD = 2.51$ *s*) than participants who chose between potentially low-value options ($M = 2.52$ *s*, $SD = 1.12$ *s*), as depicted in Fig. 5c ($t(41.76) = 4.10$, $p < .001$, $d = 1.21$ over the logarithmic transformations of the distributions)—perhaps because the stakes were higher, such that participants would earn more by making the right choice between the potentially high-value options. Including the incorrect answers again did not yield any different results ($t(55.15) = 2.77$, $p = .007$, $d = 0.72$). There was no significant difference between the percentage of people who responded accurately when choosing between potentially high-value options vs. potentially low-value options (76.67% vs. 76.67%, Fisher's exact: $p = 1$). To compare these results with those of Experiment 1a, we ran a 2 choice types (old vs. novel) x 2 choice values (potentially high-value options vs. potentially low-value options) ANOVA. There was no main effect of choice type, $F(1,88) = 0.12$, $p = .728$, $\eta^2 = 0.002$, or of choice value, $F(1,88) = 0.33$, $p = .570$, $\eta^2 = 0.004$. Crucially, there was a significant interaction, $F(1,88) = 20.99$, $p < .001$, $\eta^2 = 0.193$. These results suggest that participants in Experiments 2a-2b had a response time benefit for potentially high-value pieces because they learned their specific value during the thinking phase, and not because they developed a general strategy for thinking about those pieces more quickly during the decision phase.

5. Experiment 3: a “Snake-in-the-stack” effect in explicit thought reports

Experiments 1 and 2 showed that, in novel situations, people can preferentially think about options that are likely to be of high value, and that the process of developing this thinking plan is fast and does not depend on longer-term cached values. This behavior was consistent with our computational model, which predicted that a rational thinking strategy should prioritize thinking about options with a higher expected value. At the same time, this experiment only considered situations where thinking about the good was the best strategy. We therefore do not know if people have a good-action bias in all situations, or if they can revise this strategy when necessary.

To test this possibility, we used our computational framework to search for problems where the best strategy would be to think about low-value options first, and tested if participants can adjust their thinking strategy accordingly. As we show below, this search led to a task structure that we call “snake in the stack.” This task is structurally similar to the tasks that we used in Experiments 1 and 2, with the difference that we introduced a “snake”, or a piece that comes with a relatively large cost for participants. In this type of situation, our model predicts that the best strategy is to first find the snake (i.e., think about potentially low-value pieces) and then switch to thinking about potentially good options.

This “snake in the stack” decision problem may be reminiscent of real-world situations in which a broad space of possible options can become significantly constrained at decision time, such that you will have to select from a smaller set of options, which may not always have the best possible options. For instance, consider delicate social situations, in which one might want to think first about what one shouldn't say to not make the situation worse. In this context, one can plan ahead about what the best or worst things to say might be, but ultimately the space of possible things to say can also change and be significantly reduced based on how the other person might respond, such that you might not be able to say any of the best possible options, and need to instead know which options to *avoid* or not say.

The strategy derived from our model implies that the amount of time used to think about good pieces depends on how quickly one finds the bad piece. For instance, if a participant was lucky and happened to identify the bad piece on their first try, this would give them enough time to also evaluate high-value pieces and show the same traces of the good-action bias from Experiments 1 and 2. However, if a participant is unlucky and slow to find the bad piece, this would come at the cost of not having the opportunity to evaluate the high-value pieces, and the traces of the good-action bias from Experiments 1 and 2 should disappear.

To test this effect experimentally, participants in the next experiments were told that large pieces had higher expected values, but that one of the small pieces would lead to a very high cost. Half of the participants were then shown a puzzle where *all* the small puzzle pieces were snakes (Easy-Snake condition). Participants searching for the snake in this condition would always find one on the first try (perhaps believing they got lucky), and thus should switch to thinking about the high-value options. The other half of participants were shown a puzzle where none of the small pieces were snakes (Impossible-Snake condition). Participants searching for the snake in this condition would never be able to find it, and thus might fixate on the low-value options (perhaps believing they made a rotation error).

5.1. Model simulation and results

To search for problem structures where the good-action bias no longer applies, we modified the decision problem along two dimensions: (1) the number of options an agent could choose between in the decision phase (set to either 2, 3, 4, 5 and 6 out of 6), and (2) the number of low-value options in the puzzle among which the snake could be hidden (set to 2, 3, 4, or 5). For instance, the problem where the number of choices = 3 and the number of apparent snakes = 4 corresponds to a puzzle with two large (i.e., potentially high-value) pieces and four small (i.e., potentially low-value) pieces, one of which is believed to be the snake. At the decision phase, the model would be allowed to select one of three randomly selected pieces to win (or lose) points.

We solved POMDPs for all 20 problems in the 5×4 parameter space (see Supplemental Methods for full implementation details). We then began by testing whether any region in this space led to strategies that prioritized thinking about bad options. Fig. 6 shows the results from this analysis, coding the first thinking action chosen by the model. When the agent can choose among all, or nearly all, the options during the decision phase (rightmost columns), the model begins by searching for high-value options. This makes sense: When an agent can choose from all options, knowing the option with the highest value should suffice. However, when the agent will have to select the best piece from a limited subset (e.g., two options; leftmost column), the model no longer prioritizes the potentially high-value options. Instead, it allocates time depending on the number of apparent snakes. The higher the number of apparent low-value options, the more the agent thinks about them (indicated by how the green squares turn red as we increase the number of low-value options along the y-axis in Fig. 6). This again makes sense: The more low-value options, the more likely that one of them will be part of the decision phase, in which case knowing whether the piece is the snake or not is critical.

Of the space of 20 models, we chose a setting with four low-value options where a decision phase with two pieces (third row from the bottom, leftmost column). We chose this setting because it was one of the closest to Experiment 1 (critically involving only two choices at the decision phase), but was where our analysis suggested that people might prioritize thinking about bad options. Within this setting, the time an agent devotes to evaluating potentially high-value options depends on whether an agent finds the snake immediately or not. As we discussed above, if the agent finds the snake on the first try, then it can switch to thinking about the two high-value pieces. But if it never finds the snake, it will keep thinking about the apparent snakes until time runs out (or at least realizes or determines that it's futile to search for the snake). To probe this intuition, we tested our model in two modified situations. In one condition, we gave the POMDP policy a puzzle where all low-value options were snakes (Easy-Snake condition). In the other condition, we gave the POMDP policy a puzzle where none of the low-value options were snakes (Impossible-Snake condition).

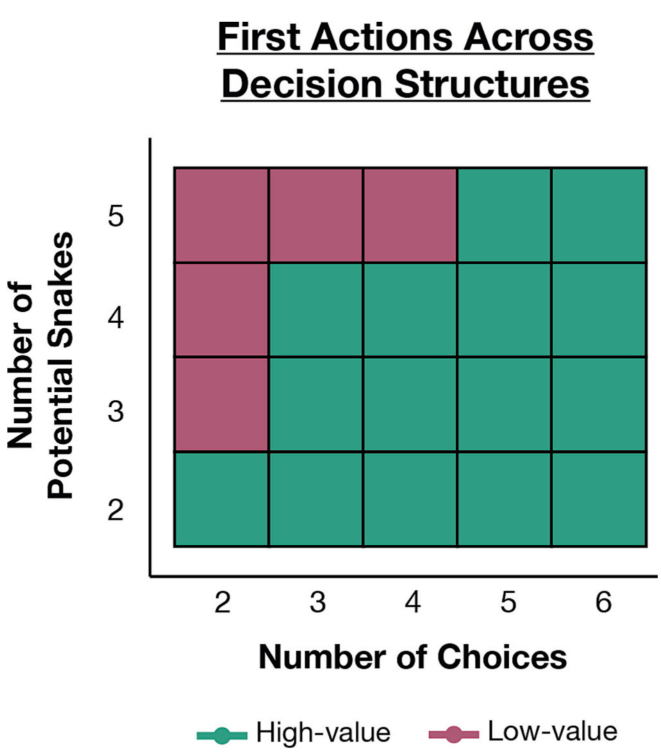

Fig. 6. First actions of each model systematically varying the number of apparent low-value options (among which one was is a "snake" that provides negative cost), as well as the number of choices the agent is during the decision phase.

Critically, the POMDP was always solved under the assumption that one (and only one) of the small puzzle pieces would be a snake. Testing the POMDP on these new problems therefore created situations where the POMDP would be led to believe it had found the unique snake on its first try in the Easy-Snake condition, and that it had failed to find the snake in the Impossible-Snake condition. Fig. 7a-b show our model's thinking strategies across both conditions. In the Easy-Snake the model first searches for the snake (considering low-value options) and immediately switches to evaluate high-value options on the second action, having believed that it found the snake. In contrast, in the Impossible-Snake condition, the model devotes more time to searching for the snake, at the cost of being unable to spend as much time evaluating the potentially high-value pieces. Given these thinking strategies, we then asked what people will do in these situations.

5.2. Explicit self-report method

As in Experiment 1, we began by using a self-report method to see which options people would report thinking about.

5.2.1. Method

This experiment is identical to Experiment 1 except where noted. 300 participants were recruited on the online Prolific platform (150 for each of the two conditions below) for monetary compensation. The sample size was determined based on a power analysis run on pilot data. Participants first read a brief set of task instructions that explained the logic of the task. Here, participants were told that the six puzzle pieces consisted of two potentially high-value options (where one fit and the other did not) and four potentially low-value options (where three pieces fit and only one did not [the snake]). Participants were told that choosing the small piece that did not fit would lead to a decrease in their score of 10 points. As in Experiment 1, participants were told to simply study the pieces, and that they would be asked questions about them afterwards (i.e., participants were not told which options they would have to choose between). In the Easy-Snake condition, all the low-value options did not fit—in which case any option people think of should be the snake. In the Impossible-Snake condition, none of the low-value options did not fit—in which case people would never actually be able to find a snake. Participants began the task with a score of 0. After five seconds, participants were asked to click on the pieces that they thought about. When they had clicked on all the pieces they had thought about, they pressed a key to complete the study. All methods and analyses were pre-registered (https://aspredicted.org/zp4tv.pdf).

5.2.2. Results and discussion

People reported thinking of an average of 2.86 pieces (as in Fig. 8a). Fig. 8b shows that the proportion of potential snake pieces that people thought about was greater in the Impossible-Snake condition than in the Easy-Snake condition (permutation test over 10,000 permutations, $p = .001$). These initial self-report results suggest that people do not just default to thinking about potentially high-value options—they do switch thinking plans based on the structure of the problem.

6. Experiment 4: a "Snake-in-the-stack" effect in implicit decision-making times

We employed the same implicit measure in Experiment 2 to confirm our results from Experiment 3. At the end of the thinking phase, all participants were prompted to select which is better of two high-value pieces. If participants simply prioritize thinking about good pieces (as in Experiments 1 and 2), our manipulation should have no effect on participants, as they would not take more or less time searching for the snake. However, if participants switch their thinking strategy by first searching for the snake before thinking about good options, then participants in the Easy-Snake condition should be faster in the decision phase relative to participants in the Impossible-Snake condition

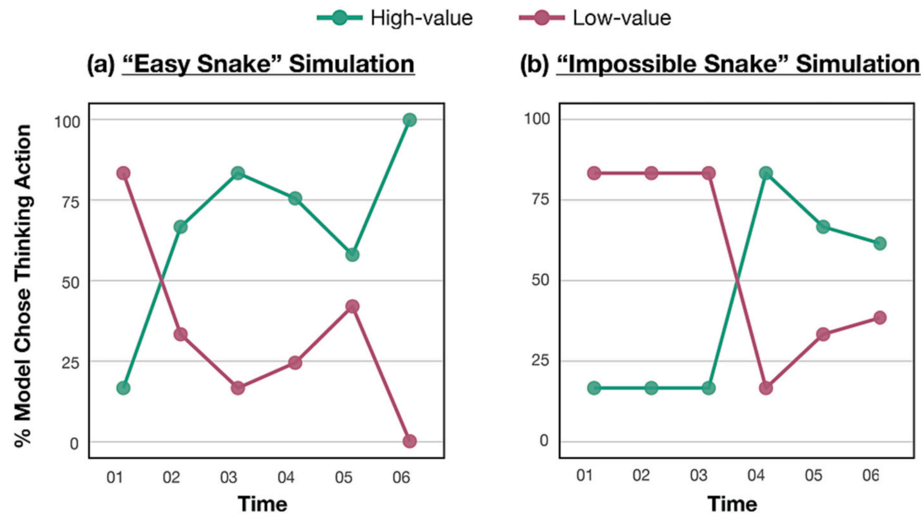

Fig. 7. Model predictions about what pieces to think about as a function of time step in Experiment 3. The model behaves differently depending on whether the potential snakes are easy or difficult to find. In either case, our rational model suggests that the best strategy is to think about the low-value pieces.

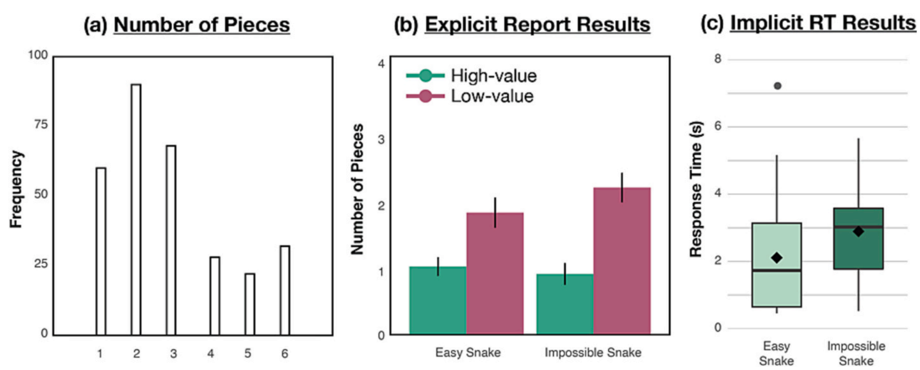

Fig. 8. Results from Experiment 3 and 4. (a) Histogram of how many pieces people selected. The x-axis depicts the total number of pieces per participant, with the maximum number of pieces being 6. (b) The average number of pieces people selected of a particular value, either: high, medium, or low. There were two possible high-value pieces and four possible low-value pieces. Error bars reflect 95% confidence intervals. (c) Results from Experiment 4. Diamonds reflect means.

(because only participants in the Easy-Snake condition had the time to evaluate the high-value pieces).

6.1. Method

This experiment was identical to Experiment 2a, except as noted. Seventy new participants participated. The sample size was determined before data collection began based on a power analysis run on pilot data (see Supplemental Materials for details). As in Experiment 1, participants were told to simply study the pieces, and that they would be asked questions about them afterwards (i.e., participants were not told which options they would have to choose between). In the Easy-Snake condition, all the low-value options did not fit—in which case any option people think of should be the snake. In the Impossible-Snake condition, none of the low-value options did not fit—in which case people would never actually be able to find a snake. Participants began the task with a score of 0. In the decision phase, participants were always asked to decide between the two potentially high-value options. Three participants were excluded because their mean performance in the comprehension questions was 2 standard deviations below the grand population mean. These participants were replaced, until a total of 70 participants was reached. All methods and analyses were pre-registered (https://aspredicted.org/67gf7.pdf).

6.2. Results and discussion

Response accuracy and response times for the single trial were recorded for each observer. Only response times where participants responded correctly were included in our analysis. Participants in the Easy-Snake condition responded faster to the potentially high-value options (M = 2.11 s, SD = 1.67 s) 690 than participants in the Impossible-Snake condition (M = 2.89 s, SD = 1.33 s; t(46.36) = 2.60, p = .013, d = 0.70 over the logarithm of the distributions; see Fig. 8c). Including the incorrect answers did not yield any different results, t (57.25) = 2.85, p = .006, d = 0.68). There was no significant difference in accuracy across the Easy-Snake and Impossible-Snake conditions (62.86% vs. 80.00%, Fisher's exact: p = .185). This pattern of results suggests that people in the Impossible-Snake condition might have fixated more on the apparent snakes, giving them less time to think about the potentially high-value options. Moreover, while we did not observe differences in people's explicit reports of whether they thought about high-value options, these response time results support the idea that high-value options were nonetheless treated differently across conditions, at least to the extent that people were faster at deciding between them in the Easy-Snake condition than in the Impossible-Snake condition. Implicit in this is that people must have started thinking about the potentially low-value options first—ultimately suggesting that they flexibly switched their thinking strategy to think about the worst options instead of the best options.

7. Experiment 5: thought flexibility within-subjects

So far, all experiments have been run as between-subjects experiments, but we can also probe flexibility *within* a single individual. In a case where a participant first encounters a situation in which they should think about high-value options and then subsequently encounters a situation in which they should think about low-value options, will they be able to flexibly switch thinking plans? Here, we ran an experiment in which each participant first encountered the decision problem from Experiments 1 and 2 (where the best strategy is to think about high-value options) and then the decision problem from Experiments 3 and 4 (where the best strategy is to think about low-value options). The key question was whether participants would be able to flexibly change strategy when they switched to a different decision problem.

7.1. Method

This experiment combines Experiments 1 and 3. Three hundred participants were recruited on the online Prolific platform (150 for each of the two conditions described below). All participants first encountered the exact decision problem from Experiment 1 (the "No-Snake" phase)—after which they were told that they would be presented with a different puzzle, where now there could be potential "snakes" in the set. They were then presented with the exact decision problem from Experiment 3 (the "Potential-Snake" phase), where half of participants were assigned to the Easy-Snake condition and the other half, to the Impossible-Snake condition. All methods and analyses were pre-registered (https://aspredicted.org/gd5ng.pdf).

7.2. Results and discussion

Fig. 9 shows the proportion of high-value options selected of the pieces that participants selected, across the No-Snake and Potential-Snake phases. First, participants were more inclined to think of high-value options in the No-Snake phase than in the Impossible-Snake condition of the Potential-Snake phase (permutation test over 10,000 permutations, $p = .005$). Second, within the Potential-Snake phase, they reported thinking about more high-value options in the Easy-Snake condition than in the Impossible-Snake condition (permutation test over 10,000 permutations, $p = .041$). These results suggest that participants were able to switch from thinking about high-value options to thinking about low-value options when the decision problem had changed.

We then asked whether participants who thought of more high-value pieces in the No-Snake phase would tend also to think of more high-value pieces in the Potential-Snake phase. Fig. 9 shows the association between the proportion of high-value options selected in the No-Snake phase versus the Potential-Snake phase (Fig. 9a: Easy-Snake, and Fig. 9b: Impossible-Snake). There was a positive correlation between responses in the two phases both in the Easy-Snake condition ($r = 0.31$, $p < .001$) and in the Impossible-Snake condition ($r = 0.27, p < .001$). In short, although participants were able to change their strategies between the two phases, there was still a significant correlation between responses in the first phase and responses in the second. Future work can explore whether such correlations reflect a robust individual difference trait.

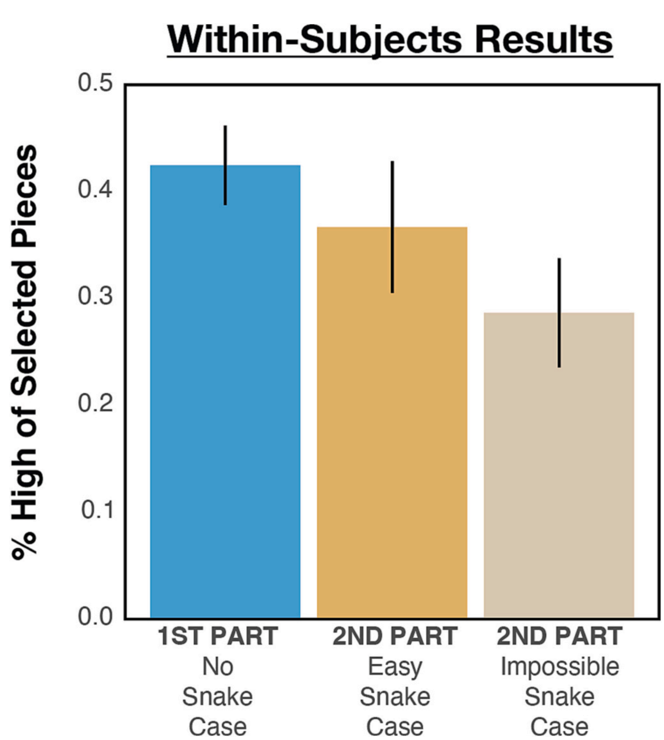

Fig. 9. Results from Experiment 5. The average proportion of high-value options from participants' selected pieces. Error bars reflect 95% confidence intervals.

8. General discussion

People often think about possible actions they can perform even before they are faced with an actual decision. A question arises about which specific actions people tend to consider when they are engaged in this type of cognition. To address this question, we used computational methods to determine which actions would be best to consider in various settings and conducted a series of experiments to determine which actions people do tend to consider.

Computationally, we formalized this class of problems as a partially observable Markov decision process (POMDP), with thinking itself treated as a type of action. We then found the best thinking plans for different specific decision problems. The results indicated that there are certain decision problems for which the best strategy is to think about the potentially high-value options and others for which the best strategy is to think about the potentially low-value options.

The modeling results revealed an interaction between (a) the number of options the agent can choose between at decision time and (b) the proportion of options that are potential "snakes". When the number of options the agent can choose between at decision time is high, it is always best to consider the potentially high-value options. By contrast, when the number of options the agent can choose between at decision time is low, the best strategy depends on the proportion of options that are potential snakes (i.e., if some of these options are extremely bad, as in the snake in the stack). When that proportion is small, the best strategy is to consider the potentially high-value options, whereas when that proportion is large, the best strategy is to consider the potentially low-value options.

Empirically, we used two approaches—a self-report method and an implicit response time method that helped us tap into what people were thinking about *before* they had to make a decision. In the thinking phase, participants were given different options and had the opportunity to think about whichever options they wanted to. Participants could be explicitly asked what they thought about, or in a decision phase, participants were confronted with just two of these options and asked to choose between them. Response times in the decision phase thereby provided evidence about which options participants were considering in the thinking phase. Specifically, the shorter a participant's response time in response to a pair of options in the decision phase, the more reason we have to conclude that the participant already considered those options in the thinking phase.

Using these methods, Experiments 1–2 looked at decision problems for which the formal model indicated that the best strategy would be to think about the high-value options. In those experiments, the results indicated that participants did tend to think about the high-value options, even in a novel context. These results clarify whether the good-action bias operates over model-based vs. model-free values, in that insofar as people can quickly learn and use arbitrary value functions, this suggests that the underlying mechanism may be more model-based than model-free. Experiments 3–4 then turned to decision problems for which the model indicated that the best strategy would be to think about the low-value options. In those experiments, the results indicated that people did tend to think about the low-value options. These results suggest that people do not just default to the high-value options, but rather utilize a more flexible mechanism of thought-planning. Finally, Experiment 5 showed that, in a within-subjects design, people were able to flexibly adapt thinking plans as the problem structure changed. Taken together, the studies therefore suggest that people can respond flexibly, thinking about either high-value options or low-value options depending on which strategy is best for the specific decision problem they face.

8.1. Selecting thoughts vs. selecting actions

In certain respects, the problem of choosing which options to consider can be seen as closely analogous to the problem of choosing which actions to perform. Indeed, the formal tools we have used to model the former problem are borrowed directly from research on the latter. Yet, although the problems are similar in their formal structure, many of the cognitive processes people use when choosing among possible actions to perform simply would not make sense when they are choosing among possible options to consider.

In particular, when people are choosing among different possible actions, they often proceed by carefully considering the pros and cons of certain possibilities. However, this same approach would almost never make sense when choosing between different possible options to consider. After all, in the time a person spent considering the pros and cons of considering an option, they could always instead have been actually considering the pros and cons of the option itself. For example, in the task we used in our experiment, participants would be unlikely to spend much time carefully considering the pros and cons of mentally rotating a specific shape, because they could always instead have used that time to actually mentally rotate the shape itself.

Given this, existing research has suggested that people might choose among possible options to consider using a relatively simple heuristic. One of the most promising such suggestions is that people employ what we have called the "good-action bias," i.e., that they show a very general tendency to think more about actions they regard as having high value than about actions they regard as having low value (Bear et al., 2020; Icard et al., 2017; Mattar & Daw, 2018; Morris et al., 2021). Of course, this simple heuristic will not always enable people to allocate their thinking in a way that maximizes expected utility, but given the computational constraints people face, it might sometimes turn out that this algorithm is the best of all the algorithms it would be feasible for them to use.

Work in adjacent fields may also provide possible dimensions that could be imported when considering *what we think about*—such as how consumers know which products to consider first (Weitzman, 1978), or which brands to deliberate between more carefully (Hauser & Wernerfelt, 1990). Here, consumers similarly use surface-level features to select products (e.g., a box of cereal) that they then think about further (e.g., look at the labels) to realize their true values. But of course, there remain crucial differences between selecting products versus selecting thoughts, and it could be interesting to consider whether there may be more general principles that operate across different types of cognitive "selection" tasks.

In the present studies, we find evidence that the options people consider are not simply determined by the good-action bias and that the criteria people use are more complex. Nonetheless, the broader point clearly stands. Whatever cognitive process people are using to select options to consider, it is almost certainly not a process that involves carefully considering the pros and cons of considering each option.

8.2. Exploring the underlying process

Future work should continue to explore people's ability to show this sort of flexibility in determining which options to think about. Such work will require a mixture of computational and empirical research. At an empirical level, our experiments used an explicit self-report method after the thinking phase, and an indirect measure of response times during decision-making to tap into what people thought about offline, *before* actually having to make any decision. Future work can employ more direct measures (such as eye-tracking; Callaway, van Opheusden, et al., 2021, Callaway, Rangel, & Griffiths, 2021) during the actual thinking phase to obtain more detailed traces of people's thinking plans.

At a computational level, we face further questions about the conditions under which it is best to think about potentially good options vs. potentially bad options. In the present paper, we looked at two specific dimensions along which decision problems could vary, but future research could continue this investigation by looking at other dimensions. Such research would presumably uncover other dimensions that affect what thinking plan people should deploy, which could then be tested experimentally. In particular, our work did not explore two important dimensions. First, we did not manipulate how varying degrees of time pressure might affect thinking plans and flexibility. It is possible that, under extreme time pressure, people might always default to a simpler good-action bias. Conversely, under no time pressure, people might have less of a need to develop an efficient thinking strategy, as they might have the luxury to consider all options. Second, we did not consider problems of a sequential nature. That is, in many situations, we do not have immediate access to a set of things to think about. Instead, we must generate them sequentially and decide when to stop generating new possibilities. Our work leaves open the question of what thinking strategies are best in these situations, and whether people should show flexibility accordingly.

Work in this area might eventually lead to the development of more general theories that specify the conditions under which the best strategy is to think about potentially high-value vs. potentially low-value options. Such theories would not be limited just to one setting but would provide more general insights about when each strategy is best. For example, it might turn out that all possible decision problems that have certain features will be problems for which the best strategy is to think about potentially low-value options—which has been alluded to by previous work as well (Hamrick & Griffiths, 2014; Lieder et al., 2018).

To the extent that we can develop such an account, we open up the possibility of a new explanation of the effects observed in the present studies. It might be that people are not determining which options to consider by using a process that is even remotely like solving a POMDP. Instead, it might be that people are simply checking for certain features that serve as reliable cues to whether it is better to think of potentially high-value or potentially low-value options. If we do find that people are responding to certain characteristics that are reliable indicators of which options are best to consider, we would face a further question as to how people come to be able to identify these in the first place. One possible answer would be that people never need to learn them. Instead, the use of these features could simply be built into people's decision-making mechanisms. A second possibility would be that people are actually learning the use of these features over the course of numerous episodes of decision-making.

If people are indeed learning to use the relevant features, a question would arise as to how this learning takes place. One possible answer would be that people are making use of familiar mechanisms of model-free learning (e.g., Gläscher, Daw, Dayan, & O'Doherty, 2010). On this view, people would have to be capable of model-free learning at an extremely abstract level. For example, over the course of numerous episodes of playing chess, people would have to be learning not only about the game of chess in particular but also about very abstract patterns regarding which options tend to be most worth thinking about. The development of algorithms for such learning is an important problem for future research—which some researchers have in fact already begun to address (He, Jain, & Lieder, 2021; He & Lieder, 2022; Jain, Callaway, & Lieder, 2019).

8.3. Flexibility and inflexibility

Overall, we find evidence that seems to point to striking flexibility in people's thinking plans—but there is also some important evidence that seems to point to inflexibility. A key task for future research will be to further explore the evidence on each of these sides.

First, in Experiment 5, we obtain different results that might initially seem to point in different directions. In that study, participants went through an experiment that had two distinct phases. In the first phase, they faced a decision problem in which the best strategy was to begin by thinking about the potentially high-value pieces, while in the second phase, they faced a decision problem in which the best strategy was to begin by thinking about the potentially low-value pieces. On one hand, we found that people were successfully able to shift their thinking strategies between the two phases. Thus, when participants were in a condition in the second phase that required shifting to thinking about the bad pieces, there was a significant effect such that participants did indeed shift to thinking more about the bad pieces. This is clear evidence of flexibility. But on the other hand, there was also a significant correlation between what participants reported thinking about in the two phases. That is, participants who thought more about good pieces in the first phase tended also to think more about good pieces in the second phase. A question now arises as to how to explain this latter result.

One plausible explanation would be that people are showing both a certain amount of flexibility and a certain amount of inflexibility. On this explanation, when the decision problem changes, people do indeed show a capacity to change their thinking plans, but this flexibility is not complete. There is at least a certain degree to which the thinking plans people form in the first phase are "sticky" and do not change when the decision problem changes.

However, the results could also be interpreted in other ways. To begin with, as Fig. 10 shows, there are participants in all four quadrants, with some participants thinking about the good pieces in the first phase and then shifting over completely to thinking about the bad pieces in the second phase. It is therefore possible that there are individual differences, such that some participants are more flexible than others. Alternatively, it might be that the correlation we observe here does not show any degree of stickiness in people's thinking plans. For example, it could be that the correlation arises because of an individual difference such that some participants have a general tendency to think more about the good options, which then influences their thinking in both phases. Future research could delve further into these different possibilities.

Second, our results are striking in light of previous research that suggests that the mechanisms people use to figure out which options to consider are highly inflexible. Specifically, a number of studies suggest that people show an inflexible tendency to think of options that are generally good and statistically frequent (Bear et al., 2020; Morris et al., 2021; Phillips, Morris, & Cushman, 2019). For many decision problems, it will be helpful to think about options that are statistically frequent and generally good, but these studies seem to suggest that people tend to think about those options even when the decision problem is structured in such a way that it is obviously not helpful to think about those options.

For example, in one study, participants were asked to name the food they would *least* want for dinner (Morris et al., 2021). Clearly, in answering this question, it is helpful to think about foods that are generally bad, but the results indicated that participants actually showed a tendency to think first of foods that are generally good. Results like this one seem to suggest that there is an inflexible cognitive mechanism that generates options for people to consider when planning.

One way in which the present studies depart from previous work, however, is in the type of decision problems participants were presented with. In previous studies, people were typically asked about decision problems for which they already may have model-free values (e.g., how many hours of TV to watch a day). In such cases, it may be that the pre-existing value assignment gives rise to the inflexible mode of thought that previous studies have observed. In contrast, the present studies look at the options people think about in novel situations, where they have not already assigned model-free values to the various options. The absence of model-free values may have allowed for the flexibility observed in our experiments.

Another important distinction between the present studies and previous work is that previous work looked at the conditions under which an option "comes to mind," while the present studies look at the conditions under which people actually simulate forward what would happen if they chose an option. Within existing research on which options come to mind, there is some evidence that the options that come to mind are determined by an inflexible mechanism. Yet, despite this, it might be that people make use of a more sophisticated and flexible mechanism to determine which simulations to run. Thus, in the present studies, it might turn out that the options with potentially high values are the first that come to mind in all conditions, but then it might be that people make use of a different psychological mechanism to determine

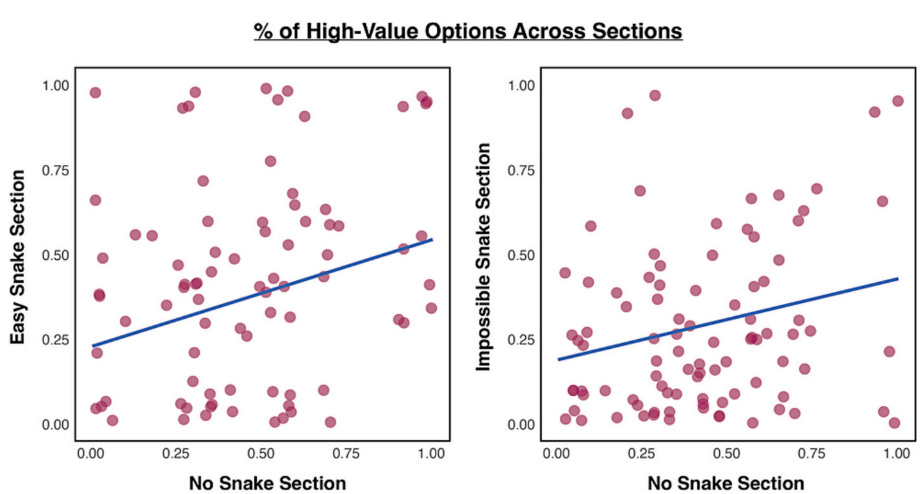

Fig. 10. The average proportion of high-value options from participants' selected pieces across the two sections of the Experiment. Plots show positive correlations between the proportion of high-value options selected in the first section vs. the proportion of high-value options selected in the second section.

which simulations to run, and that this mechanism sometimes selects the options with potentially low values.

To the extent that this latter answer turns out to be correct, a further question arises as to how far the flexibility extends. The present studies explore people's thinking in a context in which they are aiming to achieve a specific goal, but much of our thinking is not goal-directed—as in mind-wandering, in which people are not trying to address any particular decision problem (e.g., Irving & Thompson, 2018; Mooneyham & Schooler, 2013). When people's minds are wandering, do they also show the sort of flexibility observed in the present studies? For example, does the degree to which they think about potentially high-value vs. potentially low-value options depend in part on the number of options they expect to be choosing between at decision time? Regardless of what the answer to this question turns out to be, such research promises to give us a real insight into the scope or boundary conditions of the phenomena we have been exploring here.

8.4. Conclusion

When people face a choice from a set of options, they often face a difficult problem. They find themselves faced with so many different options that it would not be possible to consider them all, and they therefore need to have some way of picking out certain specific options that are especially worthy of consideration.

In a series of studies, we looked at the options people tend to consider and found that people appear to be selecting options using remarkably sophisticated criteria. People don't simply show a general tendency to consider the potentially high-value options. Instead, they consider the high-value options for certain decision problems and the low-value options for others. A key task now will be to explain how people are able to show this flexibility and how to reconcile the flexibility they show on the sort of problems in this current study with inflexibility they show on other, seemingly related problems.

CRediT authorship contribution statement

Joan Danielle K. Ongchoco: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Joshua Knobe:** Conceptualization, Writing – original draft, Writing – review & editing, Supervision. **Julian Jara-Ettinger:** Conceptualization, Writing – original draft, Writing – review & editing, Supervision.

Data availability

Data and code for all experiments reported here are available on: https://osf.io/n8em7.

Acknowledgment

This work was supported by NSF award IIS-2106690. For helpful comments, we would like to thank Fred Callaway, and members of the Yale Computational Social Cognition Lab.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105669.

References

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition, 194*, Article 104057.
Braun, E. K., Wimmer, G. E., & Shohamy, D. (2018). Retroactive and graded prioritization of memory by reward. *Nature Communications, 9*, 1–12.
Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Computational Biology, 17*, Article e1008863.
Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P., Lieder, F., & Griffiths, T. (2021). Rational use of cognitive resources in human planning. *Nature Human Behavior, 6*, 1112–1125.
Cassandra, A. R. (1998). A survey of POMDP applications. In *Working Notes of AAAI 1998 Fall Symposium on Planning with Partially Observable Markov Decision Processes*.
Chen, H., Chang, H. J., & Howes, A. (2021). Apparently irrational choice as optimal sequential decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7*, 20–43.
Gelly, S., & Silver, D. (2011). Monte-Carlo tree search and rapid action value estimation in computer go. *Artificial Intelligence, 175*, 1856–1875.
Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*, 585–595.
Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences, 29*, 24–30.
Hamrick, J., & Griffiths, T. (2014). What to simulate? Inferring the right direction for mental rotation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research, 16*, 393–408.
He, R., Jain, Y. R., & Lieder, F. (2021). Measuring and modelling how people learn how to plan and how people adapt their planning strategies to structure of the environment. In *International Conference on Cognitive Modeling*.
He, R., & Lieder, F. (2022). *Learning-induced changes in people's planning strategies*.
Hsu, D., Lee, W. S., & Rong, N. (2008). A point-based pomdp planner for target tracking. In *2008 IEEE international conference on robotics and automation* (pp. 2644–2650).
Icard, T., Cushman, F., & Knobe, J. (2018). On the instrumental value of hypothetical and counterfactual thought. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
Icard, T. F. (2018). Bayes, bounds, and rational analysis. *Philosophy of Science, 85*, 79–101.
Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition, 161*, 80–93.
Irving, Z. C., & Thompson, E. (2018). The philosophy of mind-wandering. In *The Oxford handbook of spontaneous thought* (pp. 87–96).
Ito, T., & Cacioppo, J. (2005). Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition and Emotion, 19*, 1–26.
Jain, Y. R., Callaway, F., & Lieder, F. (2019). Measuring how people learn how to plan. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
Kurniawati, H., Hsu, D., & Lee, W. S. (2008). SARSOP: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems*.
de Leeuw, J. R., Gilbert, R. A., Petrov, N., & Luchterhandt, B. (2023). Simulating behavior to help researchers build experiments. *Behavior Research Methods, 55*, 1863–1873.
Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review, 124*, 762–794.
Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43*.
Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125*, 1–32.
Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience, 21*, 1609–1617.
Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology, 67*, 11–18.
Morris, J., Phillips, A. E., Huang, K., & Cushman, F. (2021). Generating options and choosing between them rely on distinct forms of value representation. *Psychological Science, 32*, 1731–1746.
Ng, A. Y., & Jordan, M. I. (2000). Pegasus: A policy search method for large MDPs and POMDPs. In *Uncertainty in Artificial Intelligence Proceedings*.
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.
Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences, 23*, 1026–1040.
Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*, 296–320.
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.
Sutton, R. S., Barto, A. G., et al. (1998). *Introduction to reinforcement learning. 135*. MIT Press Cambridge.
Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*, 1039–1061.
Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38*, 599–637.
Weitzman, M. (1978). *Optimal search for the best alternative. 78*. Department of Energy.