Value-guided choice sets support efficient planning

Jonathan Phillips (phillips01@g.harvard.edu), Adam Morris (adammorris@g.harvard.edu)
& Fiery Cushman (cushman@fas.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St. Cambridge, MA, 02138

Abstract

Real-word decision making often involves selecting a single choice from an arbitrarily large set of possible options. Given that it is typically not feasible to evaluate every possible option in real world decision making, how are human decision makers able to efficiently make good decisions? We propose and evaluate a two-step architecture according to which people first sample a small subset of options weighted by their previously learned value, and then evaluate those options within the current decision-making context. We demonstrate that a version of this model captures human decision making in problems where time and resource constraints prevent the evaluation of every option, and connect this research to the growing literature on the representation of non-actual possibilities.

Keywords: value-guided decision making; choice sets; modal cognition; possibility

Introduction

It is a striking feature of ordinary life that we have far too much to think about. Imagine a psychology student, Sally, deciding where to eat lunch. She has certain preferences (e.g. she likes Mexican food, dislikes walking long distances) and constraints (e.g. she's deathly allergic to walnuts) that she should factor into her decision. Ideally, she would carefully evaluate all her options, and choose the one with highest expected value. For example, she might appraise each option based on how close it is to her office, how Mexican its cuisine is, the likelihood that it uses walnuts, etc., and then choose the option with the highest aggregated value. This process—computing the expected values of options at decision time by planning over a causal model of the environment—has been intensively studied, and we have some idea how it could be accomplished for a small set of options (Dolan & Dayan, 2013; Doll, Simon, & Daw, 2012).

But in any real-world decision, there are an overwhelming number of potential options (Cushman & Morris, 2015; Phillips & Knobe, 2018). Our own workplace is within a moderate walk of hundreds of restaurants, and thousands are within a short taxi ride. And the problem is even worse than this, because Sally has more options than just restaurants: She could also grow the crops herself, or catch a wild animal to eat, or steal food from the communal refrigerator, etc. She couldnt possibly plan over all her options—she would die of starvation before she finished.

Yet people like Sally are able to make these decisions with speed and ease. How? Intuitively, people don't consider all possible options. Rather, they construct a small subset of options to evaluate, and ignore all the rest (Newell, Shaw, & Simon, 1958). For instance, Sally might only consider Chipotle and Taco Bell, and choose one of those. The process by which people narrow down the enormous set of potential options to a small set of relevant choices is known as *choice set construction* (Ben-Akiva & Boccara, 1995). (Colloquially, options in the choice set just seem to "come to mind".)

The aim of this paper is to characterize how choice sets are constructed. Not all options are equally likely to make it into someones choice set; people clearly favor some options (e.g. Chipotle) over others (e.g. catching a wild animal). What determines which options come to mind? We focus on one potentially important factor: how good an option has been in the past (i.e. the option's past value). Options that have been good in the past tend to be good in the future. Moreover, prior research has demonstrated that people spontaneously compute and maintain a representation of how good an option has been, on average, in the past (Dolan & Dayan, 2013); the past values of options are pre-computed, or "cached", before decision time. (In the reinforcement learning framework, this process is often called "model-free learning" (Sutton & Barto, 1998).) Hence, the mechanism that constructs choice sets might be designed to propose options with high past values—narrowing down options to consider without incurring a high computational cost.

We explore this idea in three ways. First, we construct a computational model of value-guided choice set construction, and simulate its performance (and the performance of two alternative models) in various environments. When options that have been good in the past tend to be good in the future, the choice set model achieves good accuracy at low computational cost.

Second, we present an experimental paradigm designed to elucidate the role of value-guided choice set construction in decision-making. We fit the model to people's choices in the experiment, and find that the best-fitting model constructs choice sets guided by the prior value of options. These results suggest that people spontaneously construct choice sets when faced with difficult decisions, and are often more likely to include options with high prior values in those choice sets.

Third, we connect our model to recent work on modal cognition—that is, the representation of hypothetical and counterfactual possibilities. Previous work has argued that for many real world problems (e.g., how to get to the airport when your car breaks down), people's default or implicit representations of which actions are "possible" is constrained by the value of possibilities (Phillips & Cushman, 2017; Phillips & Knobe, 2018). Following this finding, we ask whether the learned value of the options in an experimental context influences whether or not participants represent those options as possible choices. Mirroring the previous work, we find that implicit, but not explicit representations of possibility tend to exclude options with a low value.

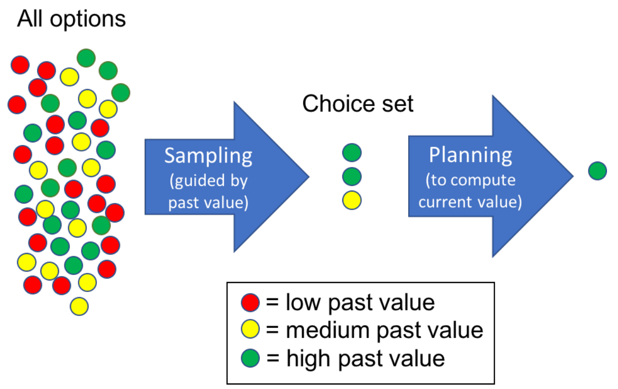

Figure 1: A schematic depiction of the choice set construction model.

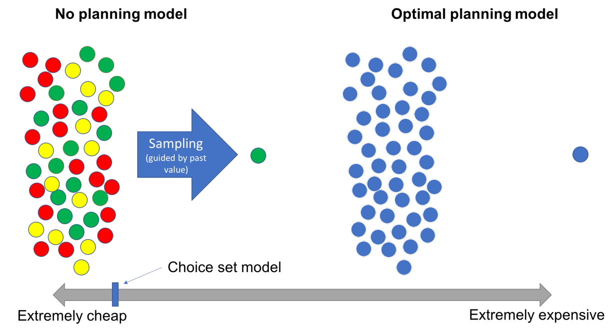

Figure 2: A schematic depiction of two alternative models. The no planning model is extremely computationally cheap; the optimal planning model is extremely expensive; and our choice set model falls somewhere in between.

Computational model

A schematic of the choice set model is depicted in Figure 1. There is a large pool of N potential options, each marked with a pre-computed past value. We assume that agents have learned these values from past experience, and do not explicitly model the learning process. The agent samples a small number of K options, without replacement, from this pool.

The sampling process is non-uniform, and is more likely to sample options with high past values. Let Q_i^p be the cached the past value of option i. Then the probability of sampling an option i is:

$$Prob(option\ i\ in\ choice\ set) = \frac{e^{\beta_1 Q_i^p}}{\sum_{j=1:N} e^{\beta_1 Q_j^p}}$$

where β_1 is an inverse temperature parameter controlling the degree to which sampling is biased towards options with high past values. (This formula employs a *softmax function* over the past values; we will use this terminology throughout the paper.)[1]

Once the choice set is sampled, the agent uses a laborious planning process to compute the current value Q_i^c of each option in the choice set, and chooses probabilistically among them with a softmax function over current value:

$$Prob(choosing\ option\ i\ from\ choice\ set) = \frac{e^{\beta_2 Q_i^c}}{\sum_{j=1:N} e^{\beta_2 Q_j^c}}$$

We do not explicitly model the planning process. Instead, we treat it as a black box that the agent can use to compute the current values of options at high computational cost.

[1]Non-uniform sampling without replacement is tricky, because maintaining stable sampling probabilities as the pool shrinks seems to require a costly renormalization after every sample. Fortunately, there is a simple, highly parallelizable algorithm that can sample without replacement in one pass over the options, without having to constantly renormalize. The agent simulates an exponentially distributed random number (with rate parameter 1) for each option, divides it by the options probability, and chooses the K options with the lowest resulting numbers. This algorithm achieves the desired weighted sample. See (Efraimidis & Spirakis, 2006) for proof and elaboration.

Alternative models We compare the choice set model to two alternatives, which anchor the two ends of a spectrum of computational complexity (Figure 2). The "no planning" model does not perform any forward planning or evaluation of options in the current context, and simply samples an option with probability proportional to its past value. Because past values are cached before decision time, this process is computationally cheap, but can be highly inaccurate if circumstances change.

In contrast, the "optimal planning" model plans over all options in the current context, and chooses the best. By planning over a causal model of the environment, this approach achieves high accuracy, but at a high computational cost.

The choice set model falls somewhere between these two extremes; it evaluates some options via planning, but much less than the optimal planning model. We show that, for a plausible range of environments, the choice set model provides major gains in accuracy over the no planning model, at a fraction of the cost of optimal planning.

Simulation setup To show this, for each of the three models, we simulated 10,000 agents using that model to make decisions in five different environments. Each agent made a single decision in each environment, which consisted of choosing among $N = 1000$ options based on their past and/or present value. The past and current value of each option were simulated anew for each agent-environment pair. The no planning model sampled according to the options past values; the optimal planning model deterministically chose the option with the highest current value; and the choice set model used the past values to construct a choice set of size $K = 10$, from which it chose an option with probability proportional to the current values. Again, we assumed that actual agents would acquire these values through prior learning episodes (for the past values) or online planning (for the current values), but we did not explicitly model the learning/planning processes.

The five environments differed solely in the simulated correlation between past and current values. The values were

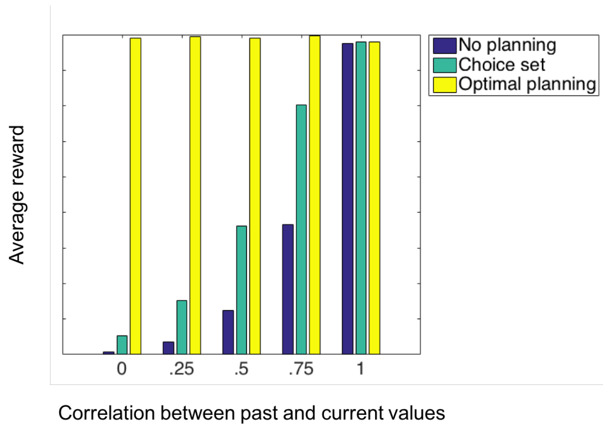

Figure 3: Simulated earnings of the three models across five environments.

drawn from lognormal distributions with correlation either 0, .25, .5, .75, or 1. (The lognormal distribution embodied the assumption that most of the options available to us are bad, while only a few are good.)

Simulation results The simulation results are shown in Figure 3. As expected, the optimal planning model achieves the highest accuracy across all environments (but at an extreme computational cost). Moreover, when the correlation between past and current values is low—i.e. when the past value of options is not indicative of their current value—then both the no planning and choice set models perform poorly.

Crucially, however, for a broad range of intermediate correlations between past and current value, choice set models are adaptive. When past values are highly but not perfectly predictive (e.g. $r = .75$), the choice set model performs almost twice as well as the no planning model, with accuracy approaching the optimal planning model, at a low computational cost. This result suggests that, as long as decision environments do not change too quickly, constructing value-guided choice sets is an efficient way to make effective decisions.

Behavioral Experiment

Next, we tested whether people construct value-guided choice sets when faced with difficult decisions. To test this, we employed an experiment with two stages. The idea was to expose people to a large set of different-value options in Stage 1, and then ask them to make decisions using those options in Stage 2. The resulting decision patterns could be tested for signatures of value-based choice set construction.

All parts of this experiment were pre-registered; the pre-registration can be found at `https://aspredicted.org/blind.php?x=33tr23`. (Note that, in the pre-registration document, the experiment stages are labeled differently.)

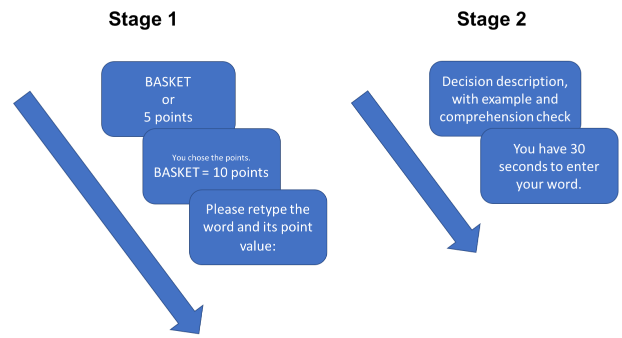

Figure 4: Design of the experiment.

Design In Stage 1 of the experiment, participants were exposed to a set of fourteen common English nouns (e.g. "basket", "community", "machine"). Each word was associated with some amount of bonus points. For instance, "basket" might have been worth 10 points, and "community" might have been worth 0. Half of the words were randomly chosen to have a low point value (either 0, 1, or 2 points), and half to have a high point value (either 8, 9, or 10 points). (Points were translated into bonus money at the end of the experiment.)

In order to learn these word-value associations, in Stage 1 participants played a game where they repeatedly chose between a word and a fixed number of points (Fig. 4). For instance, on one trial, a person might have had to choose between basket and 5 points. If they chose the word, they earned however many points its worth. If they chose the fixed number of points, they received that many points. Thus, participants were incentivized to learn the word-value associations and use that knowledge to win more bonus points throughout the game.

Participants completed 8 trials per word, for a total of 112 trials. Importantly, no matter what they chose, we showed the words point value on each trial. This procedure guaranteed that people were exposed to each word an identical number of times. To further ensure that people learned the word-value associations, we asked participants to retype the word and its value after each trial.

Then, in Stage 2, participants faced a series of decisions such as: "Give us a word from Stage 1 with the most number of vertical lines in its letters. You'll win 10 points for each vertical line in the letter of your word." In these questions, the potential options were the words from Stage 1, and each options current value (e.g. the number of vertical lines in the word) was difficult to evaluate. There were 8 decisions in total. For each decision, participants were given an example and a comprehension check. All decisions had a time limit, which was calibrated *a priori* to the difficulty of each decision (e.g. the vertical lines decision had a time limit of thirty seconds.)

Each options current value in the Stage 2 decisions (e.g. the number vertical lines in the word) were uncorrelated with its past value in Stage 1. Participants were explicitly instructed of this fact. Nonetheless, we hypothesized that, to make these decisions, people would construct a small set of words to evaluate, and that words with high Stage 1 point values would be more likely to enter this choice set.

So far we have omitted two important details about the design. First, at the end of Stage 1, participants made a series of possibility judgments about the words from Stage 1. Second, at the end of Stage 2, participants took a free recall test, recalling as many of the Stage 1 words as they could. We will return to both of these key details below.

Choice results 300 participants took the experiment. We excluded participants for whom any of the following was true: They didn't complete the study, they successfully rewrote less than 75% of the words or values during Stage 1 training, they showed a Pearson correlation between Stage 1 value and Stage 1 choices of less than .75, they failed to give a Stage 1 word for more than 2 of the 8 Stage 2 trials, they repeated an answer in Stage 2 more than twice (people were not allowed to repeat words on consecutive trials), they passed less than 50% of the Stage 2 comprehension checks, they recalled less than 5 words in the free recall question, or they wrote things down physically during the experiment (as measured by a probe at the end). We also excluded any Stage 2 trials in which the participant did not give a response that matched a Stage 1 word. After exclusion, 205 participants remained. All participants give informed consent, and the study was approved by Harvards Committee on the Use of Human Subjects.

We tested for the presence of value-guided choice sets in two ways. First, if people are constructing value-guided choice sets, then their word choices in Stage 2 should show an influence of both past value (i.e. the point values in Stage 1) and current value (i.e. the values of the words in the current decision).

To test for an influence of current value, for each Stage 2 decision, we ranked all the words according to their current values (from 1, the worst word, to 14, the best word). Then, we computed the average rank of each participants Stage 2 choices (Fig. 5). Peoples choices were ranked significantly above chance, demonstrating that they were influenced by the current values of the words (one-sample t-test, $t(204) = 51, p < .001$).

To test for an influence of past value, we employed a similar procedure. We computed the percentage of each participants word choices which had high point values in Stage 1 (Figure 6). People chose words with high Stage 1 values significantly more than chance, suggesting that they were also influenced by the past values of the word (one-sample t-test, $t(204) = 2.5, p = .01$).

Of course, the fact that people are influenced by both the past and current values of the words does not prove that they are constructing value-guided choice sets. There are other ways that the past and current word values could combine to

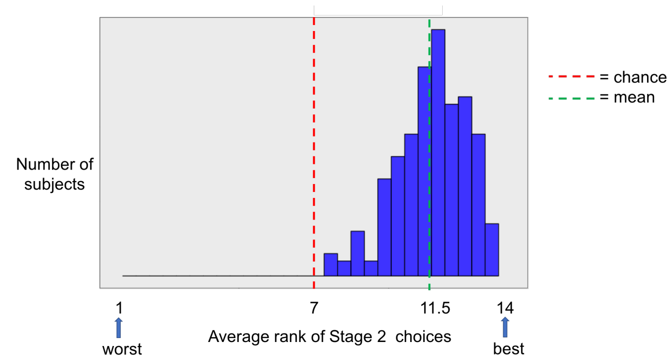

Figure 5: Average rank of each participant's word choices in Stage 2, according to the current values of the words. People chose words with high current values significantly above chance, suggesting that their choices were influenced by current value.

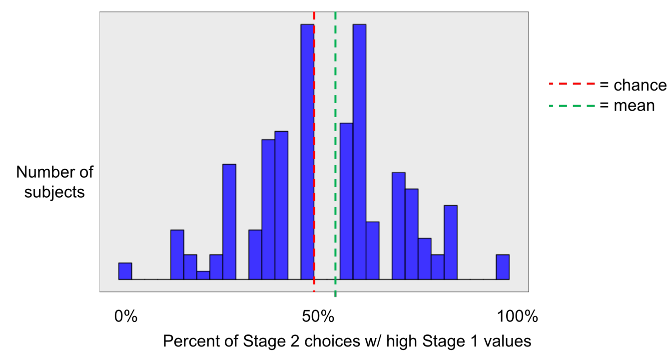

Figure 6: Percentage of each participant's Stage 2 word choices which were high-value in Stage 1. People chose words with high past values significantly above chance, suggesting that their choices were influenced by past value as well.

influence choice. People could be alternating between the two approaches, computing the current values in some trials and relying on past values in others; or, people could be basing their choices on a linear mixture of past and current values.

To rule out these alternatives, we fit our choice set model and non-choice-set alternatives to peoples choices, and performed formal model comparison. For each type of model (choice set, no choice set), we fit several variants, shown in Table 1. We computed the maximum a posteriori estimates for all parameters, using a Gamma prior for the inverse temperatures and a uniform prior for the mixture weights and choice set size. The possible choice set sizes were restricted to $2, 3, 4$. We then performed Bayesian model selection, approximating the model evidences with the Laplace method and treating model as a random effect across subjects (Stephan, Penny, Daunizeau, Moran, & Friston, 2009).[2]

[2] To demonstrate that this model comparison was a valid technique, we first simulated choices in this task from all the models. As predicted, the models involving value-guided choice sets were

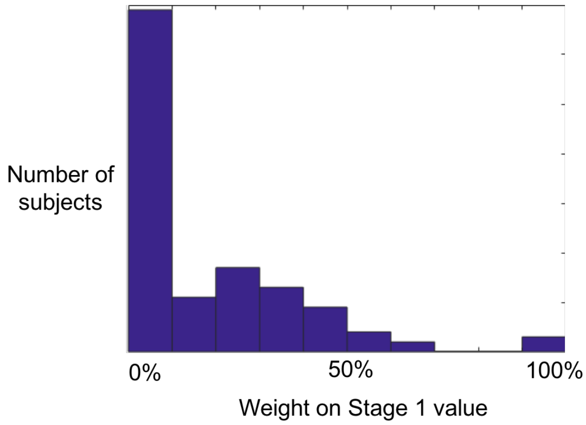

Figure 7: Distribution of weights of the influence of Stage 1 value on choice set sampling, across subjects.

The overwhelmingly preferred model, with a protected exceedance probability of over .999, was the choice set model which used both Stage 1 (past) and Stage 2 (current) value to construct choice sets. This result suggests two things. First, as hypothesized, people were using past value—the point values from Stage 1—to guide their choice set construction.

Second, before going through the laborious computation of current values (e.g. counting the number of vertical lines), people likely had access to some cue that correlated with the current value of the word (e.g. the words length). By using this cue as an additional influence on choice set construction, peoples choice sets also appeared to be influenced by Stage 2 value.[3]

According to the preferred model, when sampling a choice set, people employed a weighted mixture of Stage 1 and Stage 2 values:

$$Prob(option\ i\ in\ choice\ set) = \frac{e^{\beta_1 (w * Q_i^{Stage-1} + (1-w) * Q_i^{Stage-2})}}{\sum_{j=1:N} e^{\beta_1 (w * Q_i^{Stage-1} + (1-w) * Q_i^{Stage-2})}}$$

where w captures the influence of Stage 1 value. As a final test, we extracted each participants best-fit w and examined the distribution (Figure 7). About half of participants showed little influence of Stage 1 value ($w < 0.1$), but the other half showed a range of influence. The mean w was 0.17, indicating that, on average, Stage 1 value contributed 17% of the weighting on which words entered people's choice sets.

Ruling out a memory encoding effect One worry about our design is that people might be failing to choose low-value words not because they're excluding the words from their choice sets, but because they simply couldn't remember them; they never encoded the low-value words in the first place.

To rule this out, at the end of Stage 2, we asked participants to recall as many of the Stage 1 words as they could. (Participants were able, on average, to remember most of the words; the average number recalled was 10.7). Then, when fitting the computational models to each person's choices, we restricted the models to only consider words which that person could recall. Thus, when the preferred model estimated the influence of Stage 1 value on choice set construction, it was only calculating that influence among words which the participant was able, in principle, to recall. This rules out that our effect is due solely to memory encoding.

Including the free recall test allowed us to run an additional, exploratory test of our hypothesis. The free recall test itself can be thought of as a decision, where some options (i.e. words) will come more easily to mind than others. The words that come more easily to mind will, on average, be recalled earlier in the free recall test.

Consistent with our hypothesis, words that were high-value in Stage 1 were consistently recalled earlier. We estimated a linear mixed effects model, regressing, for each participant, the order in which each word was recalled on the Stage 1 value of the word (with maximal random intercepts and slopes for subject and word). Words that were high-value in Stage 1 were recalled earlier ($\beta = 0.7, t(28.6) = 3.6, p = .001$).

Possibility judgments As described above, participants were asked to make judgments of whether it each word was a "possible" option in Stage 1. Prior work on high-level modal

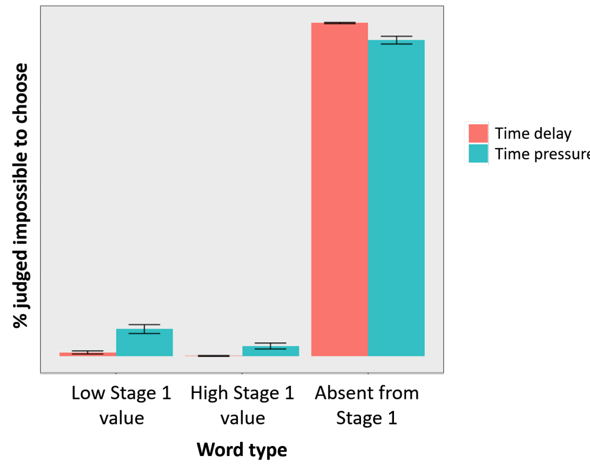

Figure 8: Mean percentage of words judged to be impossible to select in Stage 1 as a function of whether the words had been randomly assigned a low value (left), a high value (middle) or were absent entirely (right). Red bars indicate responses made after a time delay; blue bars indicate responses made under time pressure. Errors bars depict $+/-1SEM$.

only preferred when the simulated choices were produced by value-guided choice sets.

[3]The existence of these cues does not stop a person from planning once the choice set is constructed. The cues would be very rough estimates of current value, and further evaluation of the options in the choice set would still be beneficial.

cognition demonstrates that under time pressure (but not time delay), participants show a tendency to regard 'low-value' options (e.g., immoral or irrational actions) as strictly 'impossible' (Phillips & Cushman, 2017; Phillips & Knobe, 2018). In order to extend this finding to the present context, some participants made possibility judgments under time pressure and others under time delay; we hypothesized that the former group would show a stronger effect of option value on possibility judgment.

We therefore analyzed participants' possibility judgments using a mixed 2×2 *ANOVA* with word type (Low Stage 1 value vs. High Stage 1 value) as a within-subjects factor and time condition (Time delay vs. Time pressure) as a between-subjects factor. This analysis revealed a main effect of time condition, $F(200) = 30$, $p < .001$, a main effect of word type $F(200) = 21$, $p < .001$, and critically, an interaction effect $F(200) = 10$, $p = .002$ (see Fig. 8). More specifically, when participants had to answer quickly, they exhibited a tendency to judge that low-value words were actually 'impossible' to select in Stage 1. This tendency was absent when participants underwent a time-delay before responding, and occurred much more strongly for low-value words than for high value words.

Conclusion

Our work builds on the idea that people narrow down the set of options to consider (Newell et al., 1958; Browne et al., 2012; Huys et al., 2012), and that certain options are more "available" and thus more likely to be evaluated (Tversky & Kahneman, 1973). We proposed a specific feature—high past value—that makes options more available, and demonstrated that decisions with large option sets can be made effectively with this decision strategy. We then used a novel behavioral paradigm to show that a large percentage of people appear to spontaneously employ this architecture. Finally we directly connected this implicit representation of a set of available options in an experimental context to prior work on modal cognition for real-world problems and demonstrated that the prior value of an option plays a similar role in both. These finding represent a key step toward understanding how people make quick, effective decisions in environments of real-world complexity.

Acknowledgments

This research was supported by Grant N00014-14-1-0800 from the Office of Naval Research.

References

Ben-Akiva, M., & Boccara, B. (1995, May). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, *12*(1), 9–24. Retrieved from http://www.sciencedirect.com/science/article/pii/016781169500002J doi: 10.1016/0167-8116(95)00002-J

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., ... Colton, S. (2012, March). A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, *4*(1), 1–43. doi: 10.1109/TCIAIG.2012.2186810

Cushman, F., & Morris, A. (2015, November). Habitual control of goal selection in humans. *PNAS*, *112*(45), 13817–13822. doi: 10.1073/pnas.1506367112

Dolan, R., & Dayan, P. (2013, October). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. doi: 10.1016/j.neuron.2013.09.007

Doll, B. B., Simon, D. A., & Daw, N. D. (2012, December). The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol*, *22*(6), 1075–1081.

Efraimidis, P. S., & Spirakis, P. G. (2006, March). Weighted random sampling with a reservoir. *Information Processing Letters*, *97*(5), 181–185. Retrieved 2017-01-03, from http://www.sciencedirect.com/science/article/pii/S002001900500298X doi: 10.1016/j.ipl.2005.11.003

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012, March). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Comput Biol*, *8*(3), e1002410. doi: 10.1371/journal.pcbi.1002410

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151–166. doi: 10.1037/h0048495

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649-4654. Retrieved from http://www.pnas.org/content/114/18/4649.abstract doi: 10.1073/pnas.1619717114

Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*. doi: 10.1111/mila.12165

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009, July). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811909002638 doi: 10.1016/j.neuroimage.2009.03.025

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA: MIT Press.

Tversky, A., & Kahneman, D. (1973, September). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. doi: 10.1016/0010-0285(73)90033-9