Generating options and choosing between them rely on distinct forms of value representation

Adam Morris[1], Jonathan Phillips[2], Karen Huang[3], & Fiery Cushman[1]

[1]Harvard University, [2]Dartmouth College, [3]Georgetown University

1 Abstract

Humans have a remarkable capacity for flexible decision making, deliberating among actions by modeling their likely outcomes. This capacity allows us to adapt to the specific features of diverse circumstances. In real-world decision making, however, people face an important challenge: There are often an enormous number of possibilities to choose among, far too many for exhaustive consideration. There is a crucial, understudied "pre-choice" step in which, among myriad possibilities, a few good candidates come quickly to mind. How do people accomplish this? We show across nine experiments ($N = 3972$, U.S. residents) that people use computationally-frugal "cached" value estimates to propose a few candidate actions based on their success in past contexts (even when irrelevant for the current context). Deliberative planning is then deployed just within this set, computing more accurate values based on context-specific criteria. This hybrid architecture illuminates how typically-valuable thoughts come quickly to mind during decision making.

Keywords: Consideration sets | Value | Planning | Reinforcement Learning | Decision making

2 Statement of Relevance

A core challenge in decision science is understanding what happens *before* choice: how, from an enormous space of initial possibilities, the mind proposes a few candidate actions to deliberate among. How do good candidates come quickly to mind for subsequent appraisal? We bridge the literature on "consideration sets", which has distinguished between these two parts of the decision process, with work on the neurocognitive architecture of value-guided decision making, which has distinguished between two forms of value estimation. We show that people differentially apply these two forms of value estimation – one that "caches" value representations by generalizing from past experience, and one that applies context-specific knowledge at decision time – to the two aspects of the decision process, generating candidate actions via cached value and then selecting among them via context-specific evaluation. This architecture offers a simple and cognitively-plausible approach to how humans make deliberation tractable in decision making.

3 Introduction

We often study decision making by giving people a choice between just two things, such as eating fruit or cake. In the real world, however, the number of possibilities is often much larger; there are thousands of different things, for instance, that we could eat any evening. We cannot deliberate about every possible action because estimating and comparing their values takes time (Krajbich, Armel, & Rangel, 2010; Simon, 1955). Rather, from this sea of possibilities, just a few surface into consciousness. Constructing a "consideration set" (Howard & Sheth, 1969) requires generating a few good candidates with minimal effort. How does the mind do this?

Our proposal begins with a distinction, central to the contemporary cognitive neuroscience of decision making, between two ways of estimating an action's value (Dolan & Dayan, 2013). When planning, people estimate action values from a causal model of their potential outcomes. Planning can thus incorporate knowledge unique to the current context ("tonight's guest is vegetarian, so steak would be bad"; Daw, Gershman, Seymour, Dayan, and Dolan (2011)). However, people also pre-compute (or "cache") estimates of an action's value by averaging over past experiences across variable contexts (resulting in a representation like "steak = good"). This approach is unresponsive to unique features of the current context (Daw, Niv, & Dayan, 2005). Cached value representations – computed in dopaminergic circuits of the basal ganglia – are imprecise but efficient, and capture key features of "habitual" as opposed to planned behavior (Dickinson & Balleine, 1994; Dolan & Dayan, 2013).

In traditional formulations of this distinction, the two value types are competitors, vying for behavioral control (Daw et al., 2005). However, combining them reveals an appealing approach to the consideration set problem. Cached values can identify typically-good actions from among many alternatives with minimal computation, making them an attractive basis on which to construct a consideration set. Online planning can precisely evaluate context-specific values with greater effort, making it ideal for choosing from among the limited set of generated options. We propose that the mind employs this hybrid architecture, using cached values to generate candidate actions (a kind of "habit of thought"), and online planning to select from among those candidates. Like other "bounded rationality" approaches to decision-making (Simon, 1955), restricting consideration to candidates with high cached values will sometimes miss optimal choices in unusual contexts, but offers a large gain in efficiency.

This hybrid value-guided architecture complements several classic psychological models of consideration set construction, which, while demonstrating that people generate high-quality options, focus mostly on the contribution of memory search and retrieval processes (Johnson & Raab, 2003; Kaiser et al., 2013; Klein, 1993) rather than the role of value representations. Other research highlighting rule-based strategies that people use to narrow down their consideration set (like eliminating options that are missing a desired feature; Hauser (2014); Tversky (1972)), or research characterizing which thoughts are more "accessible" in other cognitive contexts (Gigerenzer & Todd, 1999; Kahneman, 2003; Tversky & Kahneman, 1973), also tend to not spotlight the role of different types of value estimation.

Past treatments of value-based consideration set construction do, however, echo the broad contours of our approach. Peters, Fellows, and Sheldon (2017) found that patients with damage to a brain region crucial for storing value representations showed deficits in generating candidate options, and suggested that value is important for option generation. Hauser and Wernerfelt (1990) theorized that consumers' consideration sets could be econometrically predicted by expected utilities analogous to the context-free cached values we propose. Kalis, Kaiser, and Mojzisch (2013) highlighted the potential and understudied role of value, hypothesizing a distinction between "intrinsic" and "extrinsic" value which maps closely onto the distinction we employ here.

We build on this work by drawing on a precise distinction between two types of value (Daw et al., 2005), specifying the different roles they play in option generation versus choice, simulating the architecture to illustrate why it is beneficial, and showing that people spontaneously employ it during decision-making.

4 Simulations

4.1 Methods

We first illustrate the advantages of this architecture by defining it formally and simulating its performance in the contextual bandit setting (Sutton & Barto, 1998). (All code and data throughout this manuscript are available for download at https://github.com/adammmorris/consideration-sets.) As shown in Figure 1A and elaborated in Supporting Information (SI), an agent must choose between many possible actions. Each action is associated with a value that varies by context. The agent can use online computation to derive these precise "context-specific" values (e.g. by model-based reinforcement learning; Sutton and Barto (1998)). The agent also stores cached value estimates for each action that are insensitive to the current context, and instead generalize across past contexts (e.g. like those computed in model-free reinforcement learning; Sutton and Barto (1998)). Naturally, in any given context, these cached estimates are correlated to some degree with the context-specific action values. In our hybrid architecture, cached value estimates guide consideration set construction while context-specific value estimates guide choice. (Note that, while our model focuses on how value guides consideration set construction and choice, we do not mean to imply that value is the *only* factor involved; we highlight other such factors in the General Discussion.)

Specifically, an agent can choose one from a set of possible actions $A = \{A_1, A_2 \ldots A_N\}$ in a context $C_j \in \{C_1, C_2 \ldots C_J\}$. She constructs her consideration set, $CS \subseteq A$, by calling to mind K actions with probability proportional to their context-free cached value(s) $\hat{V}(A_i)$. Then, from this consideration set, she chooses a single action with probability proportional to its context-specific value $V(A_i, C_j)$. The correlation between cached and context-specific values is parameterized by ρ, and is set to either 0.25, 0.5, 0.75, or 1.

To illustrate the benefits of this architecture, we contrast it with alternative architectures which use only one type of value, either cached or context-specific. The "cached-only" architecture chooses a single option with probability proportional to the option's cached value. (This architecture is formally equivalent to our hybrid model with a consideration set size of 1, since, with only one generated candidate, context-specific values are not used; hence, in Figure 1B, we visualize its performance at the point $K = 1$.) This approach, akin to choosing based purely on habit, will of course be less accurate than the hybrid model, but could make up for this loss of accuracy with a gain in efficiency. The "context-specific-only" architecture chooses a single option with probability proportional to the option's context-specific value. (This architecture is formally equivalent to our hybrid model with a consideration set size of N, i.e. one which evaluates all the options; hence, in Figure 1B, we visualize its performance at the point $K = N$.) This approach, akin to performing exhaustive deliberative planning, will be maximally accurate but inefficient. Finally, we also simulate an architecture that generates consideration sets randomly and then chooses among the generated candidates with context-specific value.

When making any given choice, each architecture will vary in two key dimensions: accuracy and efficiency. To identify which architecture is favored overall in different settings, it is helpful to put these two dimensions on a common "scale". To accomplish

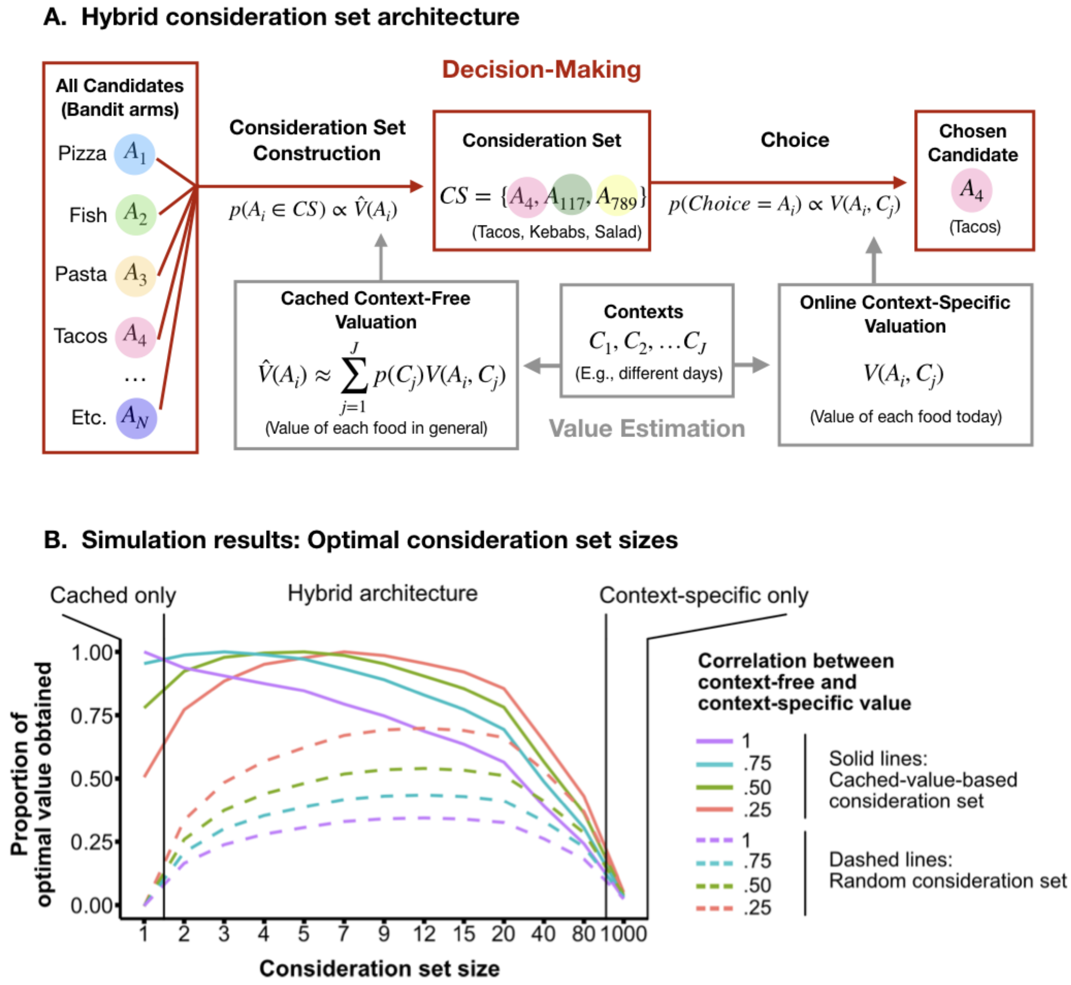

Figure 1. A. Proposed architecture. B. Simulation results: The effect of consideration set size K (X axis) on accumulated reward (Y axis), normalized where 1 is the maximum possible value for that parameter set. Color indicates the correlation between cached and context-sensitive value. Solid lines show consideration sets determined by cached value; dotted lines show randomly constructed sets. When cached and context-sensitive value are imperfectly correlated, our hybrid architecture is preferred to pure a "cached-only" ($K = 1$) or "context-sensitive-only" ($K = N$) model.

this, we follow the method of Vul, Goodman, Griffiths, and Tenenbaum (2014): We simulate agents making multiple choices within a given "time budget", and assume a temporal cost of planning and action. Devoting time to computing context-specific values improves the average value of each choice, but permits fewer total choices (and thus less overall accumulation of reward). Thus, in this framework, an architecture is favored if it balances accuracy and efficiency – i.e. if it produces a high average value for each individual choice without spending too much computation time doing so.

Figure 1B illustrates the case where there are 1000 possible actions ($N = 1000$), queries of cached values are instantaneous, computing context-specific values costs 1 time unit per candidate action, performing an action costs 25 units ($AC = 25$), and the time budget is $B = N + AC = 1025$. The “context-specific-only” architecture ($K = N$) can choose the best option exactly once: the budget allows it to evaluate all 1000 actions at a cost of 1 unit each, and then to perform the reward-maximizing action at a cost of 25 units. The “cached-only” architecture ($K = 1$) can choose sub-optimal actions 41 times by never computing any context-specific values at all. Crucially, by varying the consideration set size, agents can interpolate between these extremes and integrate both cached and context-sensitive values into our proposed hybrid architecture.

4.2 Results

The signature of our proposed hybrid architecture being favored is when reward is maximized by considering more than one action (which would be equivalent to the cached-only model), but fewer than all actions (which would be equivalent to the context-specific-only model). Naturally, when cached and context-specific values are perfectly correlated—i.e. the context never changes—reward is maximized by the cached-only model, i.e. by acting purely habitually. The agent can instantaneously retrieve the action with the best cached value, and it is always optimal. Crucially, however, for the correlations less than 1 illustrated in Fig. 1B, the optimal consideration set sizes range from 2 to 12, indicating that a hybrid architecture is favored. In no case is the context-specific-only model, i.e. exhaustive planning, favored. And, in all cases, value-based consideration set construction is highly preferred to random consideration set construction.

In the SI, we present simulation results varying the other parameters: the number of total possible actions (N), and the time cost of performing an action relative to performing a context-specific evaluation (AC). The primary effect of action time cost is that, when actions take very little time (and the correlation between cached and context-specific values is high), the cached-only model becomes preferred, because it becomes advantageous to rapidly perform many sub-optimal actions rather than deliberate to find a better action. The primary effect of N is that, when there are very few possible actions (e.g. $N = 20$), the context-specific-only model, i.e. exhaustively deliberating about every possible action, becomes more tenable (although still not preferred to the hybrid model). This is especially true when the time cost of performing an action is high; in other words, for weighty decisions with relatively few options, exhaustive deliberation can become an effective strategy. (The exact results are shown in Figure S1.) Nonetheless, over a broad range of plausible parameters, our hybrid architecture is preferred.

5 Experiments

5.1 Studies 1-3: Deconfounded cached and context-specific values in food choice

We next test whether people employ this architecture in decision-making.

Study 1: Methods Our experimental methods involve two key elements. First, it is necessary to dissociate cached from context-specific value estimates. We therefore adapted the classic "devaluation" procedure (Dickinson & Balleine, 1994) in which the experimenter induces a sudden change in the current value of actions. Context-specific evaluation can immediately incorporate this change, but cached values cannot. We can then test whether the candidates that come to mind are principally guided by cached values, while choice from within this consideration set is principally guided by context-specific value. (Note that, by using devaluation tasks which dissociate cached and context-specific values, we create situations where using cached values for candidate generation is unhelpful and seemingly irrational. These situations are ideal for testing whether people are employing cached values; they are not, however, meant to typify situations where employing cached values would be useful. In most real-world decisions, cached and context-specific values would be correlated, and hence—as demonstrated in the simulations above—our architecture would be on balance beneficial.)

Second, it is necessary to measure people's consideration sets. To accomplish this, after presenting people with a decision problem (e.g., choose a food to eat for dinner), we asked them to retrospectively report every candidate they had considered before settling on their final answer, even if the candidate was quickly rejected. This retrospective method has some disadvantages compared to real-time methods like online thought listing protocols, namely that people's memory of their decision process may be imperfect. However, it has the advantage of not interfering at all with people's decision process while they are making it, and we wanted to err on the side of non-interference. In the SI we detail several "sanity checks" of this retrospective self-report method. We found that people typically reported considering between 2 and 4 items, which is consistent with prior research (Hauser & Wernerfelt, 1990). We also found that, on average, people reported considering more items when given more time for consideration, and fewer items when given less time.

In Study 1, we exploit a familiar, everyday context: choosing what to eat for dinner. Yet, we create an unfamiliar version of this everyday decision in order to dissociate cached and context-specific values. We told people:

> Imagine that you just got dental surgery, and your doctor gives you food restrictions for the night. You're supposed to avoid food with seeds, foods that require too much chewing, and foods that are moist. What would you cook yourself for dinner tonight?

Participants had 30 seconds to choose a food. Here, participants' cached, general food preferences likely diverge from the best context-specific choice that they would compute online (i.e., taking into account the peculiar dental demands). Immediately after participants reported their choice, we asked them to retrospectively record all the options that came to mind, allowing them to list up to 8. Importantly, we emphasized that they should list any options that came to mind, even if those options were quickly rejected (Klein, Wolf, Militello, & Zsambok, 1995).

We then asked people to rate each food item in that set on two dimensions (presented in a random order). First, on their context-specific values: "In this situation, how good it would be to make this food (given the current doctor restrictions)?", reported on a scale ranging from 1 (labeled as "this food is among the **worst** dishes I could make in this situation") to 7 ("this food is among the **best** dishes I could make in this situation”), with 4 labeled as “this food would be **average** among the dishes I could make in this situation”. Second, on their general cached values: “In general, how much do you like eating this food (ignoring the current doctor restrictions)?”, reported on a similar scale (where 1 was labeled as “this food is among my **least favorite** dishes”, 4 as “this food is **average** for me”, and 7 as “this food is among my **favorite** dishes”).

202 participants were recruited on Amazon Mechanical Turk. We excluded participants who failed to answer a rating question, or who gave the same rating for all foods in their consideration set. This left 185 participants for analysis. All experiments were approved by Harvard’s Committee on the Use of Human Subjects; all participants gave informed consent. For all of our experiments, all sample sizes were chosen in advance of data collection based on preliminary effect sizes from pilot data, and we did not compute the effects of interest before terminating data collection. All exclusion criteria were also chosen in advance based on pilot studies.

In these experiments, we cannot calculate the exact probability of a food making it into a participant’s consideration set, because we do not know each participant’s set of unconsidered (but possible) foods. Instead, employing assumptions described in SI, we can compute a number that is proportional to this probability (the “inferred probability” of a food coming to mind). To test the influence of each value estimate on candidate generation, we estimated a linear mixed-effects model, regressing the inferred probability that a food came to mind on its general and context-specific value ratings. To test the influence of each value estimate on choice from among the considered candidates, we estimated a mixed multinomial logistic model, regressing choice from among the consideration set on cached and context-specific value ratings. Mixed-effect models included maximal random intercept and slopes (unless they prevented convergence); see SI for further details.

Throughout all our analyses, numbers in brackets represent 95% confidence intervals, b represents unstandardized regression coefficients, and β represents standardized regression coefficients. (For the logistic regressions, coefficients are standardized only with respect to the predictor variable.)

Study 1: Results People's ratings of general and context-specific values for the generated food items were correlated at $r = .072$ $([.0087, .13], t(957) = 2.2, p = .026)$. Though this correlation is significant, it is low, indicating that our method successfully dissociated general from context-specific values and that the ratings could be entered as simultaneous predictors on our outcomes of interest.

As predicted, people tended to generate options that were high in general value: 66% $([63\%, 69\%])$ of generated options were above the scale midpoint in general value, with only 13% $([11\%, 15\%])$ below. In contrast, they showed a much smaller tendency to generate options high in context-specific value: 49% $([46\%, 52\%])$ of generated options were above the midpoint in context-specific value, with 33% $([30\%, 35\%])$ below. (The full distributions of each rating are presented in Fig. S3.)

This pattern was borne out in our inferred probability analysis. Figure 2A shows the inferred probability of considering a candidate as a function of its specific and general value. When entered as simultaneous predictors, general value strongly predicted the probability of consideration $(b = .26\ [.23, .29], \beta = .47\ [.41, .54], t(737) = 15, p < .001)$, while context-specific value did to a smaller extent $(b = .085\ [.052, .12], \beta = .16\ [.093, .22], t(737) = 4.9, p < .001)$. These slopes differ significantly; see SI. (Though our analysis focuses on demonstrating the role of cached value representations in guiding consideration set construction, people are likely also relying on other mechanisms such as semantic or contextual memory (Johnson & Raab, 2003; Kaiser et al., 2013; Klein, 1993; Smaldino & Richerson, 2012). For example, when selecting a dinner that was not too chewy or moist, participants' option generation process presumably involved not only cached value, but also retrieving foods from memory with features such as "crispy" and "dry". This may explain the modest effect of context-specific value on candidate generation; we return to this point in the General Discussion.)

In contrast, the final choice of a candidate from within the consideration set was strongly guided by context-specific value. Of the options that were ultimately chosen, 85% ([80%, 90%]) of those options were above the scale midpoint in context-specific

value, compared to 3.2% ([.70%, 5.8%]) below. (Restricting to chosen options did not affect the distribution of general value ratings, suggesting that general value does not influence final choice.) This pattern was borne out in our analysis, with context-specific value strongly predicting choice (Figure 2A;
$b = 3.3\,[1.5, 5.1], \beta = 5.1\,[2.4, 7.8], z = 3.7, p < .001$) but not general value (which instead trended in the opposite direction,
$b = -.20\,[-.42, .016], \beta = -.32\,[-.65, .013], z = -1.9, p = .059$).

Study 2: Methods In Study 2, we replicated these results in a similar task with a modified cover story. In Study 1, the context-specific values (what you'd like to eat, given the dental restrictions) were still likely related to the cached general values of food options (what you like to eat in general). In Study 2, by using a social context (cooking for a friend), we could further distance the context-specific values from the cached ones. Specifically, we told people:

> Imagine that, as a gift, you're going to cook dinner for a friend who is currently on crutches and can't cook for themselves. You know that your friend is allergic to anything with seeds, prefers foods that don't require too much chewing, and dislikes foods that are moist. What would you cook them? (The food is only for your friend; you won't eat any of it.)

100 participants were recruited on Amazon Mechanical Turk. We used the same exclusion criteria as in Study 1, leaving 90 for analysis.

Study 2: Results People's ratings of general and context-specific values for the generated food items had a correlation of .24 ($[.16, .32], t(485) = 5.5, p < .001$). Again, this correlation is low enough to reasonably enter the ratings as simultaneous predictors on our outcomes of interest. (See Fig. S3 for full distributions of each rating.)

As in Study 1, people's option generation was strongly influenced by generalized, cached value (Fig. 2B). General value strongly predicted the probability of consideration ($b = .31\,[.27, .35], \beta = .57\,[.48, .65], t(357) = 13, p < .001$), while context-specific value did to a smaller extent
($b = .091\,[.046, .14], \beta = .17\,[.084, .25], t(357) = 3.9, p < .001$). These slopes differ

significantly; see SI. In contrast, choice from among the considered candidates was again guided by context-specific value ($b = 3.5\,[.76, 6.2], \beta = 5.3\,[.99, 9.6], z = 2.5, p = .014$), not general value ($b = .065\,[-.39, .52], \beta = .10\,[-.61, .81], z = .29, p = .77$; although a Bayes factor analysis provided ambiguous evidence against the null hypothesis, $BF_{alt} = 4.7$). (All Bayes factors were computed with the BIC approximation; Wagenmakers (2007).) These results again suggest that people are calling foods to mind based on their general cached value, and then choosing among the generated foods according to their value in the current context.

Study 3: Methods In Study 3, we explored the relative influence of two factors intimately associated with habit-like behaviors: historical cached value (Dolan & Dayan, 2013) and historical choice frequency (Miller, Shenhav, & Ludvig, 2019; Tversky & Kahneman, 1973). We propose that people use value as a guide to consideration set construction, but a plausible alternative explanation of our results is that people are using historical choice frequency instead. Naturally, these are often correlated because we tend to choose the things we value most. Because this correlation is imperfect, however, we can adjudicate between these alternatives by asking whether the options that come to mind in our "food choice" paradigm are most closely associated with participants' reports of how much they generally value those foods, or instead with their reports of how often they eat those foods.

To test this question, we replicated Study 1 with an additional follow-up question. In addition to the general and context-specific value questions, for each food people had generated, we asked them: "How often do you eat each of these foods, compared to all the types of food you eat?", reported again on a scale from 1-7, where 1 was "this food is among the least common dishes I eat", 4 was "I eat this food an average amount", and 7 was "this food is among the most common dishes I eat".

To test the relative influence of value and choice frequency on candidate generation, we estimated a linear mixed-effects model, regressing the inferred probability that a food came to mind on its general cached value, its context-specific value, and its historical choice frequency.

114 participants were recruited on Amazon Mechanical Turk. We used the same exclusion criteria as in Studies 1-2, leaving 105 for analysis.

Study 3: Results People's ratings of historical choice frequency of the generated food items was significantly correlated with the general value ratings ($r = .58\,[.52, .63], t(564) = 17, p < .001$) and not with context-specific value ($r = .061\,[-.022, .14], t(564) = 1.4, p = .15$). These results makes sense; how often someone eats a food is related to how much they like it in general but less related to how appropriate it is given strange dental restrictions.

When all three ratings were entered as simultaneous predictors, as in Studies 1 and 2, the general value of a food strongly predicted the probability of its consideration ($b = .13\,[.11, .15], \beta = .34\,[.28, .41], t(836) = 11, p < .001$) while the context-specific value did to a smaller extent, ($b = .039\,[.015, .063], \beta = .10\,[.039, .16], t(836) = 3.2, p = .0015$). Historical choice frequency also predicted a food's probability of consideration to a smaller extent, with a similar magnitude as context-specific value ($b = .047\,[.023, .071], \beta = .12\,[.060, .19], t(836) = 3.8, p < .001$). The slope for general value was significantly larger than the slopes for either context-specific value or historical frequency, and the latter two slopes did not differ from each other. This result again suggests that people disproportionately rely on cached values to generate candidate options.

When choosing an option out of the consideration set, people again seemed to rely primarily on context-specific value, although in this study the effect was marginal ($b = 5.1\,[-.25, 10], \beta = 7.7\,[-.34, 16], z = 1.9, p = .061$). When choosing among the options they considered, people did not show a significant effect of general value ($b = .071\,[-.52, .66], \beta = .11\,[-.79, 1.0], z = .24, p = .81$) or frequency ($b = -.10\,[-.73, .53], \beta = -.16\,[-1.1, .80], z = -.32, p = .75$), although the Bayes factors favored keeping them in the regression ($BF_{alt} = 207$ and 486, respectively).

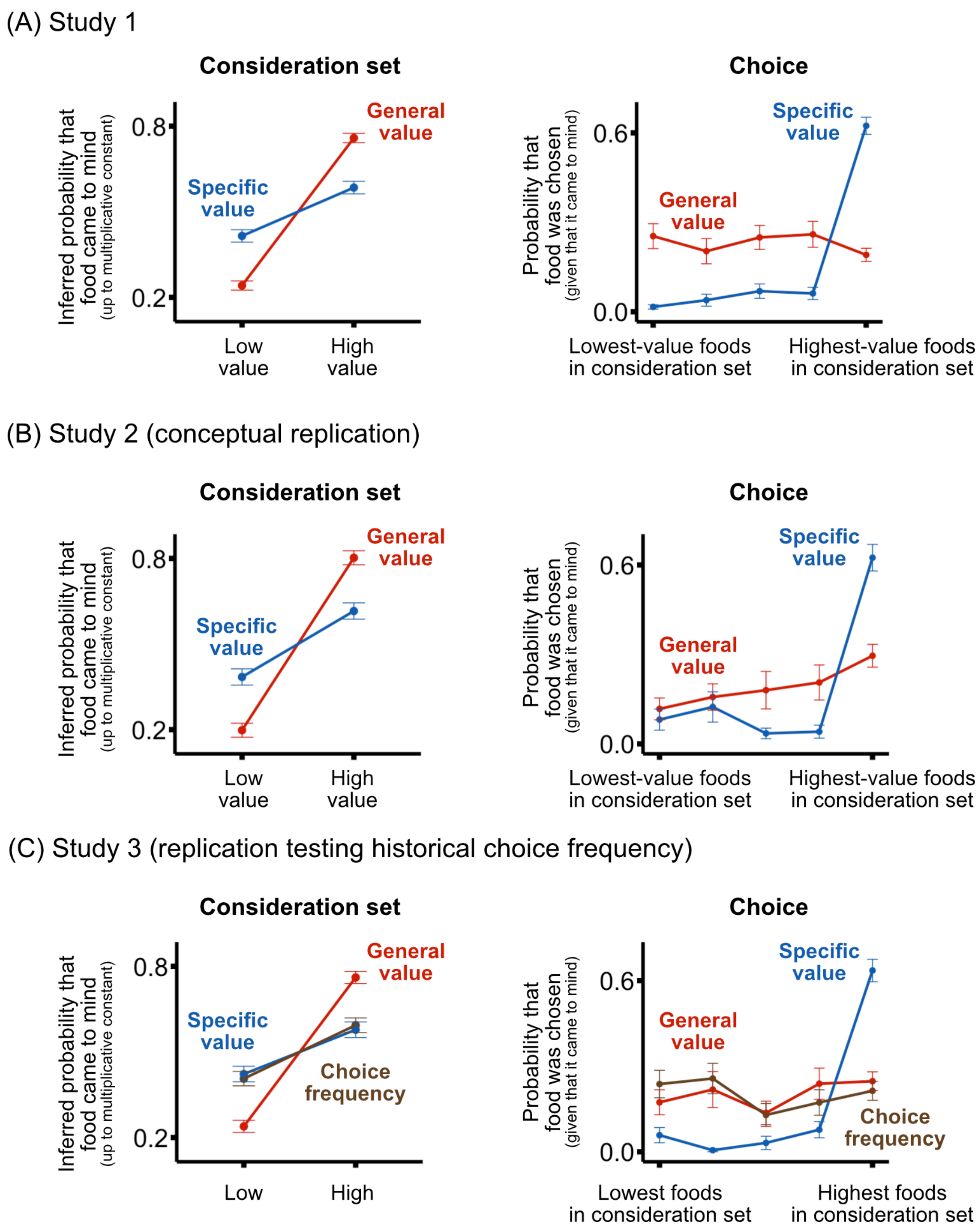

Figure 2. Results of Studies 1-3. General value strongly determines consideration and weakly determines selection (given consideration), while context-specific value strongly determines selection and weakly determines consideration. Choice frequency has a weak effect on both. Error bars are SEM.

5.2 Studies 4-6: Uncorrelated cached and context-specific values in a novel task

We next demonstrate the same pattern in a setting with more experimental control, where cached and context-specific values are orthogonal.

Methods This new experimental paradigm had two stages. In Stage I, we trained participants to associate twelve words—the months of the year—with monetary rewards. These words became the candidate "actions" used throughout the experiment. Each word was paired with an amount of points between 1 and 12 (e.g. FEBRUARY = 7), randomized across participants. On each Stage I trial, participants were given a choice between two months (e.g. FEBRUARY versus AUGUST), and received points according to their choice. (Participants received a final bonus according to how many points they earned.) Each month was paired with each other month twice, for a total of 132 trials; trial order was randomized across participants. All months were presented an equal number of times, and both month values were revealed after each trial (no matter what the participant chose).

Then, in Stage II, people were asked to make a new decision involving those twelve words: to choose a month whose third letter came late in the alphabet, with better answers earning more money. This method recruits participants' pre-existing knowledge of language as a "model" used for context-specific evaluation. There was no systematic relationship between Stage I and Stage II value, and we emphasized this to participants. Participants had 25 seconds to make their choice.

Finally, we asked participants to report which months had come to mind as candidates for the decision in Stage II. If people use cached values to generate candidates, then they would be more likely to have considered words that were good in Stage I, despite the known irrelevance of Stage I value in the present context.

Study 4 was the initial test of this prediction; Study 5 was a pre-registered replication of Study 4 (http://aspredicted.org/blind.php?x=3aa3vq).

Study 6 was a follow-up experiment with two purposes. It allowed yet another test of the main prediction (that Stage I value would influence which months came to mind in Stage II), and also tested for an additional effect: whether how frequently a word was chosen in Stage I would influence whether it came to mind in Stage II. Study 6 was identical to Studies 4 and 5, except with a modified Stage I training regime that produced some words which varied only in choice frequency (but not in value), and some words which, like those in Studies 4-5, varied in both choice frequency and value. (For details about this training regime, see SI.) The latter words were used as a replication test of our main prediction, and the former words were used to test for an additional effect of choice frequency on consideration.

For Studies 4-6, 502, 605, and 500 participants were recruited on Amazon Mechanical Turk respectively. We excluded participants who didn't complete the study, who chose the better alternative in Stage 1 training on fewer than 70 percent of trials, who failed to enter a month within the time limit in Stage 2, who failed a Stage 2 comprehension check, or who wrote things down physically during the experiment (as measured by self-report). All exclusion criteria were chosen before analyzing any of the data reported here. After exclusion, in Studies 4-6, we had 324, 373, and 326 participants for analysis respectively.

In Studies 4-6, we do know the total set of possible actions, so the analysis is simpler. To test the influence of value on candidate generation, we estimated a logistic mixed-effects model, regressing whether each month came to mind on its cached (Stage I) and context-specific (Stage II) values. To test the influence of value on selection (given consideration), we again estimated a mixed-effects multinomial logistic regression, regressing Stage I and Stage II values on whether a word was chosen out of the consideration set. We again used maximal random effects, unless they prevented convergence; see SI for details of the statistical models.

Results Indeed, people were more likely to think of words in Stage II that were good in Stage I (Study 4, Figure 3A; $b = .032\,[.0085, .056], \beta = .23\,[.052, .41], z = 2.5, p = .011)$. Study 5, the pre-registered replication of Study 4, showed the same effect (Figure 3B; $b = .021\,[-.0056, .043], \beta = .15\,[.0027, .30], z = 2.0, p_{\text{one-tailed}} = .023)$. In Study 6, there was no discernible effect of choice frequency on candidate generation ($b = .12\,[-.096, .34], \beta = .12\,[-.10, .35], z = 1.1, p = .29, BF_{null} = 25.8$), but there was again an effect of Stage I value (Figure 3C; $b = .031\,[.00096, .062], \beta = .28\,[.0083, .54], z = 2.0, p = .043)$.

Figure 3D aggregates the data to visualize the effect of Stage 1 value at a more granular level. This visualization reveals that, in addition to our effect, people were also somewhat more likely to think of the very worst words. Statistically, inclusion in people's consideration sets was predicted by an additional quadratic term: Stage I value squared ($b = .038\ [.019, .057], \beta = .20\ [.10, .29], z = 4.0, p < .001$). This pattern suggests an effect of "relative extremity" on what comes to mind in our experiment; the options that were either the best or worst among those presented saw a boost in consideration. (This can be interpreted as a kind of salience effect; Kahneman (2003); Lieder, Griffiths, and Hsu (2018).) We return to consider this effect further in the discussion. Note that this relative extremity effect is in addition to (rather than being an alternate explanation for) our primary result, that options with higher Stage I values are more likely to come to mind; when including the quadratic term, the main effect of value was still significant ($b = .051\ [.026, .076], \beta = .18\ [.092, .28], z = 3.9, p < .001$).

Stage I values did not, in contrast, significantly influence the final item chosen from within the consideration set (Fig. 3E; Study 4, $b = -.00081\ [-.18, .18], \beta = -.0012\ [-.28, .27], z = -.0088, p = .99, BF_{null} = 312$; Study 5, $b = .10\ [.090, .29], \beta = .15\ [-.14, .44], z = 1.0, p = .30, BF_{null} = 57$; Study 6, $b = .13\ [-.086, .35], \beta = .18\ [-.13, .49], z = 1.1, p = .25, BF_{null} = 85$; combined, $b = .054\ [-.050, .16], \beta = .08\ [-.075, 2.3], z = 1.0, p = .31, BF_{null} = 313$). Rather, once a candidate came to mind, its probability of being chosen was uniquely influenced by its Stage II value (Fig. 3F; Study 4, $b = 2.2\ [1.4, 3.0], \beta = 3.4\ [2.2, 4.6], z = 5.6, p < .001$; Study 5, $b = 2.6\ [1.8, 3.4], \beta = 3.9\ [2.7, 5.1], z = 6.3, p < .001$; Study 6, $b = 2.3\ [1.6, 3.0], \beta = 3.5\ [2.4, 4.6], z = 6.3, p < .001$; combined, $b = 2.3\ [1.9, 2.7], \beta = 3.5\ [2.9, 4.1], z = 10.5, p < .001$). In sum, these results support our proposed architecture: Cached values influence candidate generation, and then selection from among the generated candidates relies on context-sensitive evaluation.

5.3 Study 7: Salience or value?

There is a potential confound in Studies 4-6: The options with high cached values were also the furthest from zero—i.e., had the most extreme absolute magnitudes.

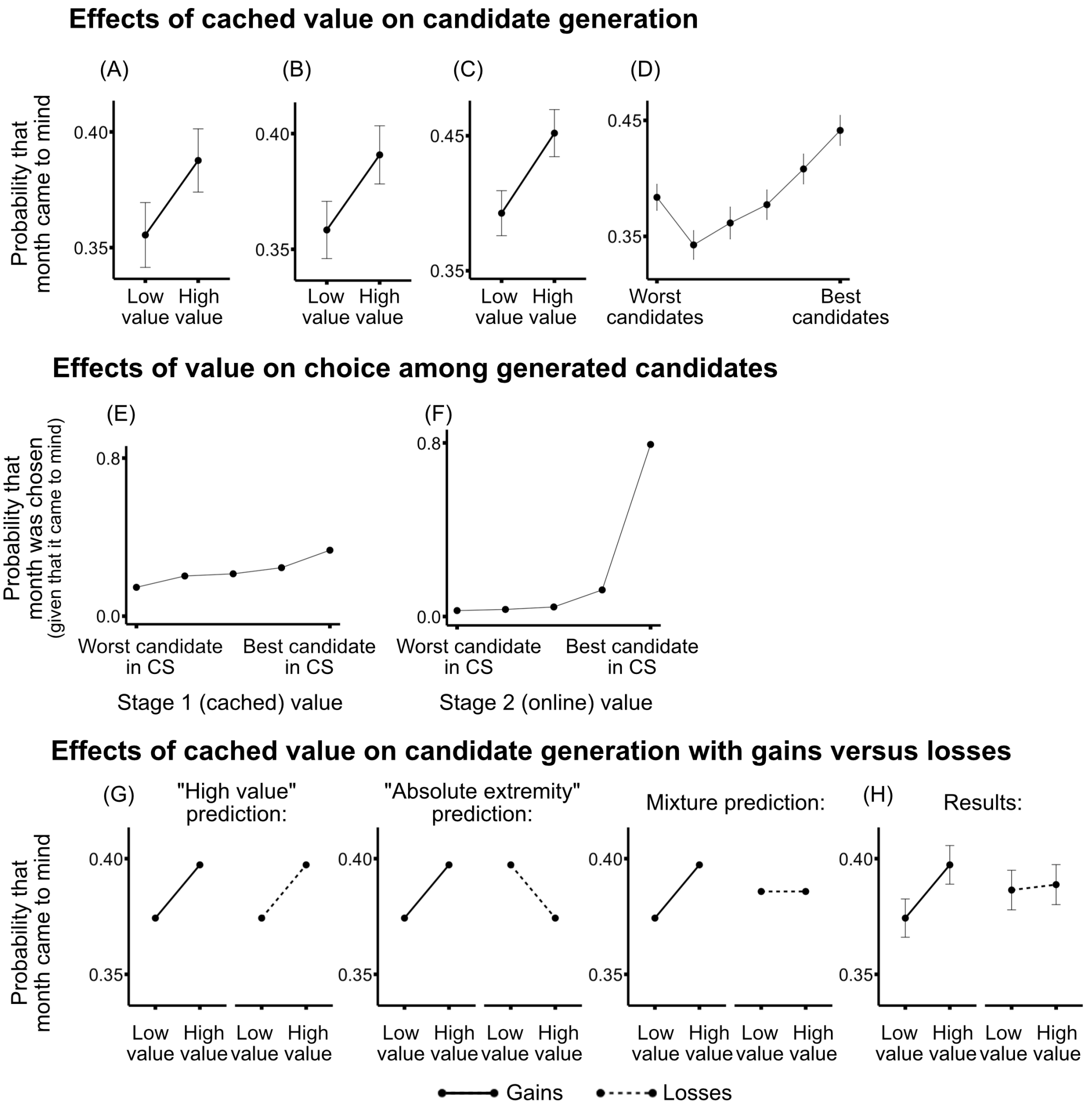

Figure 3. Effects of value on consideration and choice in Studies 4-7. (A-C) Months with (irrelevant) high Stage I values more often came to mind as candidate solutions in the Stage II decision. (D) Aggregating the data reveals a “checkmark” shape: Increases in Stage I value increase the probability of a candidate month coming to mind, except for an increased probability for the very worst candidates. (E) Once a word was generated, its Stage I value had little influence on whether it was chosen. (F) Instead, final choices were determined mostly by Stage II value. (G) Predictions of “high value” versus “absolute extremity” hypotheses when Stage I values are gains versus losses in Study 7. (H) People exhibit a main effect of Stage I value, consistent with the “high value” hypothesis.

Perhaps people generate options with more extreme cached values (whether good or bad), rather than more positive cached values. This is a plausible alternative, given the connection between extremity and salience (Kahneman, 2003; Lieder et al., 2018).[1] If

[1] Note that this “absolute extremity” alternative differs from the “relative extremity” effect documented above.

absolute extremity explains the entire effect of value on consideration set construction, then we should observe an equally sized *negative* relationship between value and consideration when months of the year are associated with variable monetary *losses*.

Study 7: Methods To test this, we reran the design in Studies 4-5 with an additional condition: For half the participants, the options in Stage I all gained money, and for the other half of participants, the options all lost money (with participants choosing options that would lose them the least money). The gains condition was an exact replication of Studies 4 and 5. In the losses condition, the methods were identical except that, in Stage I, the twelve months were randomly assigned point values of -1 to -12. (Participants were instructed of this in advance, given an initial point endowment of 2000 points, and told to choose the option on each trial that lost them the least amount of points.)

In this design, each hypothesis – "high value" and "absolute extremity" – predicts a unique pattern of results. If consideration sets tend towards options with high cached values, people will show an overall main effect of value in which words come to mind. If they tend towards options whose cached values deviate furthest from zero, people will show a crossover interaction between cached value and gain-versus-loss condition: They will tend to generate words with more positive cached values in the gains condition but words with more negative cached values in the loss condition. And if both hypotheses are correct, then people will exhibit both a main effect of cached value and an interaction between value and condition. We visualize these predictions in Fig. 3G (these predictions are schematic; the directional effects are what is predicted, not the magnitude of each effect).

To test these predictions, we estimated a logistic mixed effect regression model, regressing whether each month came to mind on its Stage I value, participant condition (gains versus losses), and their interaction. All aspects of this experiment were preregistered (https://aspredicted.org/blind.php?x=xk43pf). Following our pre-registered analysis plan, we calculate one-tailed p values for the predicted effect of each hypothesis (the main effect of Stage I value and the interaction between Stage I value and condition).

2745 participants were recruited on Amazon Mechanical Turk. We employed the same exclusion criteria as in Studies 4-6, leaving us with 1787 for analysis.

Study 7: Results As predicted by our proposed "high value" mechanism, people were on average more likely to think of words with high cached values; there was a main effect of Stage I value ($b = .013\,[.0036, .022], \beta = .090\,[.022, .16], z = 2.6, p_{\text{one-tailed}} = .0045$; Fig. 3H). There was a significant simple effect of Stage I value in the gains condition ($b = .021\,[.0065, .036], \beta = .15\,[.048, .25], z = 2.9, p = .0040$), and no significant simple effect in the losses condition ($b = .0038\,[-.012, .019], \beta = .027\,[-.083, .14], z = .48, p = .63$). The results also provided weak evidence for an additional "absolute extremity" effect; there was a marginal interaction between Stage I value and condition, where the effect of value was attenuated in the loss condition ($b = .017\,[-.0085, .042], \beta = .060\,[-.028, .15], z = 1.3, p_{\text{one-tailed}} = .09$).

Importantly, participants were not simply more confused in the losses condition: They chose the better month with roughly equal frequency in Stage I (87% [86.5%, 87%] of trials in the gains condition, 88% [87.5%, 88.5%] of trials in the losses condition), and they chose the best month from among those considered with roughly equal frequency in Stage II (82% [79%, 85%] of trials in the gains condition, 81% [78%, 84%] of trials in the losses condition).

Study 7: Discussion These results suggest that our primary effect—people tend to generate options with high cached values, even when uncorrelated with their value in the present context—is not primarily due to the options' more extreme absolute magnitude.

Why do we observe no simple effect of cached value in the loss condition? Perhaps option generation is biased towards *both* high-value options and options with higher absolute value magnitude. In the loss condition, these two tendencies would cancel out. Or, perhaps only cached values derived from gains influence consideration set construction. The cognitive-neural mechanisms underlying cached value learning do indeed involve different pathways for appetitive and aversive learning (Liu et al., 2007; Yacubian et al., 2006). Future research should explore these possibilities.

5.4 Studies 8 & 9: Anti-correlated cached and context-specific values

We conclude with an especially strong test of our model. We construct a task in which cached and context-specific values are anti-correlated by asking people to choose the option that, historically, had the lowest value. In other words, in this task, context-specific value is the mirror image of cached value. Our model predicts that cached value will "intrude" on consideration set construction—i.e. people will habitually call to mind historically good candidates even though they are the worst candidates for the present context.

Study 8: Methods We begin by implementing this design in the "food choice" paradigm used in Studies 1-3. In Study 8, we divided participants into two conditions. In the "think of best" condition, we asked participants:

> Imagine that someone has offered to cook you dinner tonight. What meal would you most want for the dinner? (Please limit your answer to normal meals that someone would reasonably cook.)

And in the "think of worst" condition, we asked:

> Imagine that someone has offered to cook you dinner tonight. What meal would you least want for the dinner? (Please limit your answer to normal meals that someone would reasonably cook.)

We then administered our consideration set measure from Studies 1-3, and asked people to rate how much they generally liked each food that had come to mind (using the same scale as before). (Note that, here, since cached and context-specific values are being dissociated by the "think of worst" condition, we don't need to use the unusual circumstances of, e.g., dental restrictions that we employed in Studies 1-3.)

To test whether general cached values persisted in influencing candidate generation in this setting, we examined how often people generated foods that were the "opposite" valence from the current decision. Specifically, we examined how often people thought of above-average foods (greater than 4 on the 1-7 scale) when they were supposed to be thinking of bad ones, compared to how often people thought of below-average foods when they were supposed to be thinking of good ones. We estimated a mixed-effects logistic regression model, regressing whether each generated food was the opposite valence from the current decision on the participant's condition.

Just as in Studies 1-3, this analysis depends on several assumptions. For instance, it assumes that there are not simply a greater number of above-average foods than below-average foods; if there were, people would be more likely to experience intrusions of above-average foods even without an effect of cached value on option generation. See SI for further details and discussion. This potential confound is addressed in Study 9.

500 participants were recruited on Amazon Mechanical Turk, and we excluded 113 (using the same exclusion criteria as in Studies 1-3), leaving 387 for analysis.

Study 8: Results As predicted, historically good candidates intruded on people's consideration sets in the "think of worst" condition. Specifically, people were over six times more likely to report a liked food coming to mind when trying to choose disliked foods than a disliked food coming to mind when trying to choose liked foods ($b = 3.1$ $[2.5, 3.7], \beta = 3.8$ $[3.1, 4.5], z = 11, p < .001$; Figure 4A).

Study 9: Methods Study 9 adapted the "months" paradigm in a similar manner. Using the paradigm from Study 4, in Stage I we trained people to assign value to each of the twelve month names. In this version, however, each month generated rewards stochastically. (Specifically, rewards were drawn from a normal distribution with a constant, randomly assigned mean between 1-12 for each month and a standard deviation of 1.75. Everything else in Stage I was identical to Study 4.) Then, in Stage II, we asked half of participants, "What was the best month to choose in Part 1?", and we asked the other half, "What was the worst month to choose in Part 1?". We then administered the same consideration set measure as in Studies 4-7.

We incidentally collected data in this paradigm in two separate instances. Both showed the same directional effect; one yielded a significant result and the other did not. We report the total, aggregated results here and the disaggregated results in the SI, and note that the results of Study 9 should be viewed as suggestive but preliminary.

Study 9 had $N = 457$ total, and we excluded 62, leaving 395 for analysis. We used the same exclusion criteria as in Studies 4-6, with two changes. First, since the Stage I training was now stochastic, we dropped the criterion that people had to make the correct Stage I choice in over 70% of trials. Second, we added a question asking whether people had done an experiment very similar to this one before, and excluded anyone who answered yes. These criteria were chosen before analyzing the data.

The analysis was identical to Study 8. We characterized a month as "opposite-valence" if it was above-average in the "think of the worst month" condition (i.e. had a Stage I mean value greater than 6) or below-average in the "think of the best month" condition (i.e. had a mean value less than 7).

Study 9: Results The results are similar to Study 8 (Fig. 4B): Months that were good in Stage I came to mind while deciding which month was worst, more than bad months came to mind while deciding which month was best ($b = 0.50$ [.23, .77], $\beta = .57$ [.25, .89], $z = 3.5, p < .001$).

Importantly, for both Studies 8-9, people were equally good at choosing candidates from within their consideration set (see SI). This suggests that participants understood the decisions equally well in both conditions. Moreover, it supports our general finding that cached values intrude primarily at the point of consideration set construction, not choice.

Studies 8 & 9: Discussion When asked to name the worst of a category, people still habitually call to mind the best. This speaks against a reductive account that cached value is simply one option feature used as a cue for memory retrieval (Gigerenzer & Todd, 1999; Johnson & Raab, 2003) or screening (Tversky, 1972)—like a car buyer narrowing their search to cars over $20,000. If this were true, then people should be equally able to screen for high-value and low-value options (just as a buyer can narrow their search to cars over or under $20,000). In contrast, we find that calling to mind high-cached-value options cannot be easily reversed.

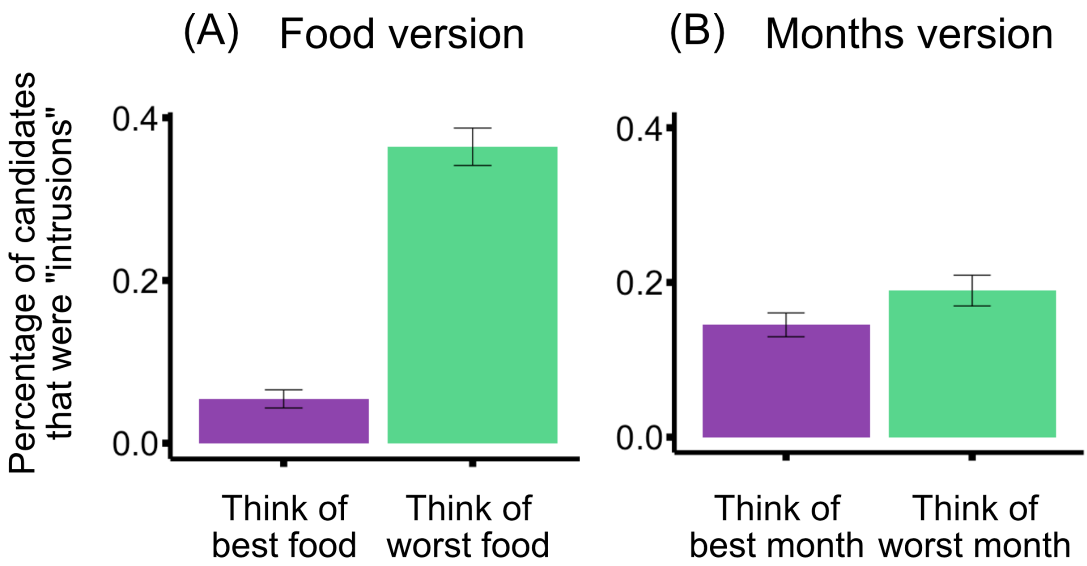

Figure 4. Percentage of intrusions in Studies 8-9. Historically good candidates more often intruded into people's minds, indicating a persistent influence of cached value on candidate generation.

6 General Discussion

In decisions involving many potential actions, a few good candidates rapidly come to mind. Our experiments show that this process is guided in part by cached values generalized from past contexts. In contrast, choice among those candidates is strongly influenced by values estimated from a model of the current context.

"Model-free" and "model-based" value estimation processes are typically conceptualized as competitors (Daw et al., 2005; Kool, Gershman, & Cushman, 2017). Our results contribute instead to a recent interest in their cooperation (Cushman & Morris, 2015; Keramati, Smittenaar, Dolan, & Dayan, 2016; Kool, Cushman, & Gershman, 2018). Model-free values may support tractable planning by directing limited deliberation towards promising actions. An intriguing implication is that many presumed instances of "habitual" action may in fact reflect its influence on candidate generation: habits not of action, but of thought (Cushman & Morris, 2015; Graybiel, 2008; O'Reilly & Frank, 2006).

Although we model candidate generation and choice as two separable steps for convenience, presumably people often generate some candidates, evaluate them, decide whether to generate more, and continue until reaching a threshold (Smaldino &

Richerson, 2012) – akin to a "satisficing" procedure (Simon, 1955). Our basic insight may be applied to this dynamic process: Cached values guide ongoing candidate generation while context-specific evaluation guides choice.

Consideration set construction certainly relies on more than cached values. Some mechanisms allow people to generate options via criteria specific to the current decision context; for instance, when generating options in Studies 1-2, people seemed partially able to search their memory directly for foods that had relevant context-specific features (e.g. "not chewy", like soups). Indeed, consideration sets rely substantially on memory retrieval processes (Johnson & Raab, 2003; Kaiser et al., 2013; Klein, 1993). Yet, people cannot always directly query their memory for answers that match current contexts. For instance, it is hard to query memory for months with third letters close to Z. (When people can or can't query memory in a "content-addressible" way (McElree, 2000) is a complex question ripe for future research (Barsalou, 1983).) Our experiments show that cached values can help fill such gaps.

Salience effects, such as recency, frequency of consideration, or "extremity", likely also contribute to consideration (Kahneman, 2003; Tversky & Kahneman, 1973). Our results supported at least one salience effect: In Studies 4-6, in addition to our primary effect of high cached value, options with more extreme cached values relative to the mean also tended to come to mind (the "checkmark" shape in Figure 3D). Salience effects like this may have a functional basis, such as conserving scarce cognitive resources (Lieder et al., 2018). An ideal general theory would specify how these diverse factors – including many others, like personality traits, social roles, and cultural norms (Smaldino & Richerson, 2012) – form a coherent, adaptive design for option generation.

A growing body of work suggests that value influences what comes to mind not only during decision-making, but in many other contexts such as causal reasoning, moral judgment, and memory recall (Bear & Knobe, 2017; Braun, Wimmer, & Shohamy, 2018; Hitchcock & Knobe, 2009; Mattar & Daw, 2018; Phillips, Morris, & Cushman, 2019). A key inquiry going forward will be the role of cached versus context-specific value estimation in these cases.

In sum, the fact that certain actions just “come to mind” is a fundamental and mysterious part of human cognition: Fundamental because it is necessary for making efficient decisions in the real world, but mysterious because we don’t know how the mind does it so quickly and effectively. Cached values may be a key piece of that puzzle.

7 Acknowledgements

We thank Adam Bear, Joshua Knobe, Samuel Gershman, Joshua Greene, Wouter Kool, Max Kleiman-Weiner, and Tomer Ullman for their advice. Funded by ONR Grant N00014-19-1-2025.

8 References

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, *167*, 25–37.

Braun, E. K., Wimmer, G. E., & Shohamy, D. (2018). Retroactive and graded prioritization of memory by reward. *Nature Communications*, *9*(1), 4886.

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

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

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

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* Oxford University Press, USA.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Reviews of Neuroscience*, *31*, 359–387.

Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, *67*(8), 1688–1699.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, *16*(4), 393–408.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *11*, 587–612.

Howard, J. A., & Sheth, J. N. (1969). *The theory of buyer behavior.* New York: Wiley.

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes*, *91*(2), 215–229.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kaiser, S., Simon, J. J., Kalis, A., Schweizer, S., Tobler, P. N., & Mojzisch, A. (2013). The cognitive and neural basis of option generation and subsequent choice. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 814–829.

Kalis, A., Kaiser, S., & Mojzisch, A. (2013). Why we should talk about option generation in decision-making research. *Frontiers in psychology*, *4*, 555.

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, *113*(45), 12868–12873.

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes*, *62*(1), 63–69.

Klein, G. A. (1993). A recognition-primed decision (rpd) model of rapid decision making. In J. Klein, R. Orasanu, C. Calderwood, & C. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). New York: Ablex Publishing Corporation.

Kool, W., Cushman, F. A., & Gershman, S. J. (2018). Competition and cooperation between multiple reinforcement learning systems. In *Goal-directed decision making: Computations and neural circuits* (pp. 153–178). New York: Elsevier.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, *28*(9), 1321–1333.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292.

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), 1.

Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007). Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *Journal of Neuroscience*, *27*(17), 4587–4597.

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of psycholinguistic research*, *29*(2), 111–123.

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological review*.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.

Peters, S. L., Fellows, L. K., & Sheldon, S. (2017). The ventromedial frontal lobe contributes to forming effective solutions to real-world problems. *Journal of Cognitive Neuroscience*, *29*(6), 991–1001.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in cognitive sciences*, *23*(12), 1026–1040.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99–118.

Smaldino, P. E., & Richerson, P. J. (2012). The origins of options. *Frontiers in Neuroscience*, *6*, 50.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (1st ed.). Cambridge: MIT press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, *79*(4), 281.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic bulletin & review*, *14*(5), 779–804.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006). Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience*, *26*(37), 9530–9537.

Supplemental Information:

Generating options and choosing between them rely on distinct forms of value representation

Adam Morris[1], Jonathan Phillips[2], Karen Huang[3], & Fiery Cushman[1]

[1]Harvard University, [2]Dartmouth College, [3]Georgetown University

This document reports the materials and methods for the simulations and experiments reported in the main text. All data and code can be found at https://github.com/adammmorris/consideration-sets.

1 Simulations

In this section, we describe our simulations in more detail. These simulations illustrate how our proposed hybrid architecture, when facing a common type of decision problem with many options, can make good choices at low computational cost.

Consider an agent facing a series of decisions with N possible actions; for instance, a person choosing what to eat for dinner. Each action has an associated reward that differs based on context (e.g. sometimes steak is good, but other times you have vegetarian guests). This is an example of a "contextual multi-armed bandit" problem (Sutton & Barto, 1998). The agent's goal is to maximize its accumulated reward. Following Vul et al., we assume that the agent has a fixed "budget" of B units that it can spend either performing actions to earn reward, or thinking about which actions to choose (Vul, Goodman, Griffiths, & Tenenbaum, 2014). (The budget can be thought of as a constraint on time or on resources.)

We consider two methods available to the agent for estimating the value of each arm. First, we assume that the agent has cached a "context-free" value for each action based on its reward across past contexts (e.g. through model-free reinforcement learning; Sutton and Barto (1998)). Accessing these cached values is costless (i.e. instantaneous, if the budget is interpreted as temporal). Second, we assume the agent can spend some of its budget to reveal an action's reward in the present context (e.g. by doing model-based planning over a model of the present environment; Sutton and Barto (1998)). Revealing an action's context-specific reward costs 1 unit.

After spending K units on "thinking" (i.e., revealing the context-specific rewards of some number K of actions out of the total set of N possible actions), the agent can choose an action, pay a cost of AC units to perform the action, and receive the reward. We assume that typically AC is much greater than 1; for the types of real-world decisions we're considering (e.g. what to eat for dinner), it takes much longer to perform an action than it does to evaluate the context-specific reward of that action by deliberating. On the other hand, we also assume that AC is not so much greater than 1 that performing a context-specific evaluation is trivial; for instance, we are not simulating decisions that are only performed once in a person's life (e.g. choosing where to go to college).

To reduce the number of free parameters, we normalize the agent's budget to the number of possible actions N; we assume that, if the agent evaluates every possible action, then it will have the remaining budget to perform exactly one action. In other words, $B = N + A$.

We consider four types of agents. **Cached-only** agents simply choose actions with the highest cached value. **Context-specific-only** agents evaluate the context-specific reward for all possible actions before making a choice. **Random consideration-set** agents choose a random subset of K actions to evaluate among the N total, and then choose the action among those with the highest context-specific value. Finally, **hybrid** agents use the architecture we propose: They create a consideration set via cached values, evaluate those, and choose from among the set via context-specific value. In general, we do not make strong claims about exactly how the consideration set is constructed, or how the final choice happens, so long as the former is influenced by cached value and the latter by context-specific value. For the purposes of the simulations, we assumed that the hybrid agents employ a deterministic "hardmax" procedure (rather than a probabilistic "softmax" one): They take the K actions with the highest cached values, evaluate those, and then choose the action in the consideration set with the highest context-specific value.

For simplicity, we do not explicitly simulate the process of learning and caching context-free values, the process of learning a model, or process of planning. Rather, we assume asymptotic convergence of cached value estimates to perfectly match the average reward across contexts, and asymptotic convergence of the model and planning process such that context-sensitive value estimates perfectly match context-sensitive rewards. We directly manipulate the key parameter that emerges from this assumed process: The correlation between cached and context-sensitive values, which captures how representative cached values are of rewards in the present context. To accomplish this, we generated two normal random variables (with mean 0 and variance 1) of pre-specified correlation ρ. This gave us, for each agent, the cached and context-sensitive values for each action.

Thus, the simulations had four free parameters: N (the total number of possible actions), AC (the cost of performing an action, relative to the cost of evaluating an action), K (the size of the consideration set), and ρ (the correlation between cached and context-sensitive values). We manipulated ρ to take the values $0.25, 0.5, 0.75, 1$; K to take the values 1-5, 7, 9, 12, 15, 20, 40, 80; N to take the values 20, 100, 1,000 and 10,000; and A to take the values $2, 5, 10, 25, 50, 100$. Importantly, the case where $K = 1$ is equivalent to a "cached-only" agent, and the case where $K = N$ is equivalent to a "context-sensitive-only" agent. So, to simulate cached-only agents, we simply simulate hybrid agents with a K of 1 but omit any cost of "thought", because there is no advantage to retrieving a context-specific value for the sole option under consideration. To simulate context-sensitive-only agents, we simulate hybrid agents with $K = N$. Thus, the problem of determining when hybrid agents outperform cached-only or context-sensitive-only agents reduces to determining when the optimal K for hybrid agents is greater than 1 but less than N.

The simulations were conducted as follows. For each joint parameter setting, we simulated $10,000$ agents using each architecture, each with a randomly generated set of cached and present action values. Each agent made a single decision and earned a reward; we then averaged the earnings of the 10,000 agents to get the average earning for a single decision made by a particular architecture. We then calculated the number of possible actions that could be taken by each architecture (based on how much of their budget they spent on computing context-specific values), and multiplied that number by the average earnings for that architecture to get its total earnings. Finally, we divided each architecture's total earnings by the maximum earnings achieved by an architecture for that joint parameter setting. The result is shown in Figure S1.

Of the four architectures, context-sensitive-only and random consideration sets are never ideal. When the action cost is low (i.e. performing an action take little time or effort, relative to evaluating an action), and when the correlation between cached and online values is high, acting on pure cached values is optimal; constructing a consideration set and evaluating the context-specific rewards is not worth the time or effort. For example, see $AC = 5$ and $\rho = .5$ in Fig. S1. And when the action cost is high and there are few possible actions to deliberate about (e.g. $AC = 100$ and $N = 20$), the context-specific-only model becomes nearly as good as the hybrid model; exhaustively deliberating over all the possible actions becomes more tenable (due to a lower number of possible actions) and more important (due to a higher time cost of performing an action). Nonetheless, for a large range of parameter settings, agents do best by generating a consideration set guided by cached value.

2 Pilot experiment: Validating our consideration set measure

Next, we describe the methods of our experiments in more detail. All experiments were approved by Harvard's Committee on the Use of Human Subjects; all subjects were recruited on Amazon Mechanical Turk and gave informed consent.

In every experiment, we asked people to retrospectively report which candidates came to mind during a previous decision. Here, we report a pilot experiment (Pilot 1, $N = 599$) that performs two basic sanity checks on this measure: Do people report having generated a reasonable number of candidates? And, do they report having generated more candidates when they had more time to think during the decision?

We told people: "Imagine that someone offered to cook you any meal you wanted for dinner tonight. What would you ask them to cook?" We gave people either 5, 10, 15, 20, or 25 seconds to answer. (We excluded anyone who wrote nothing in the answer box.) Then, we gave them our consideration set measure:

> Now, we want to know: Which foods did you consider while making your decision? In other words, which potential meals came to mind while you

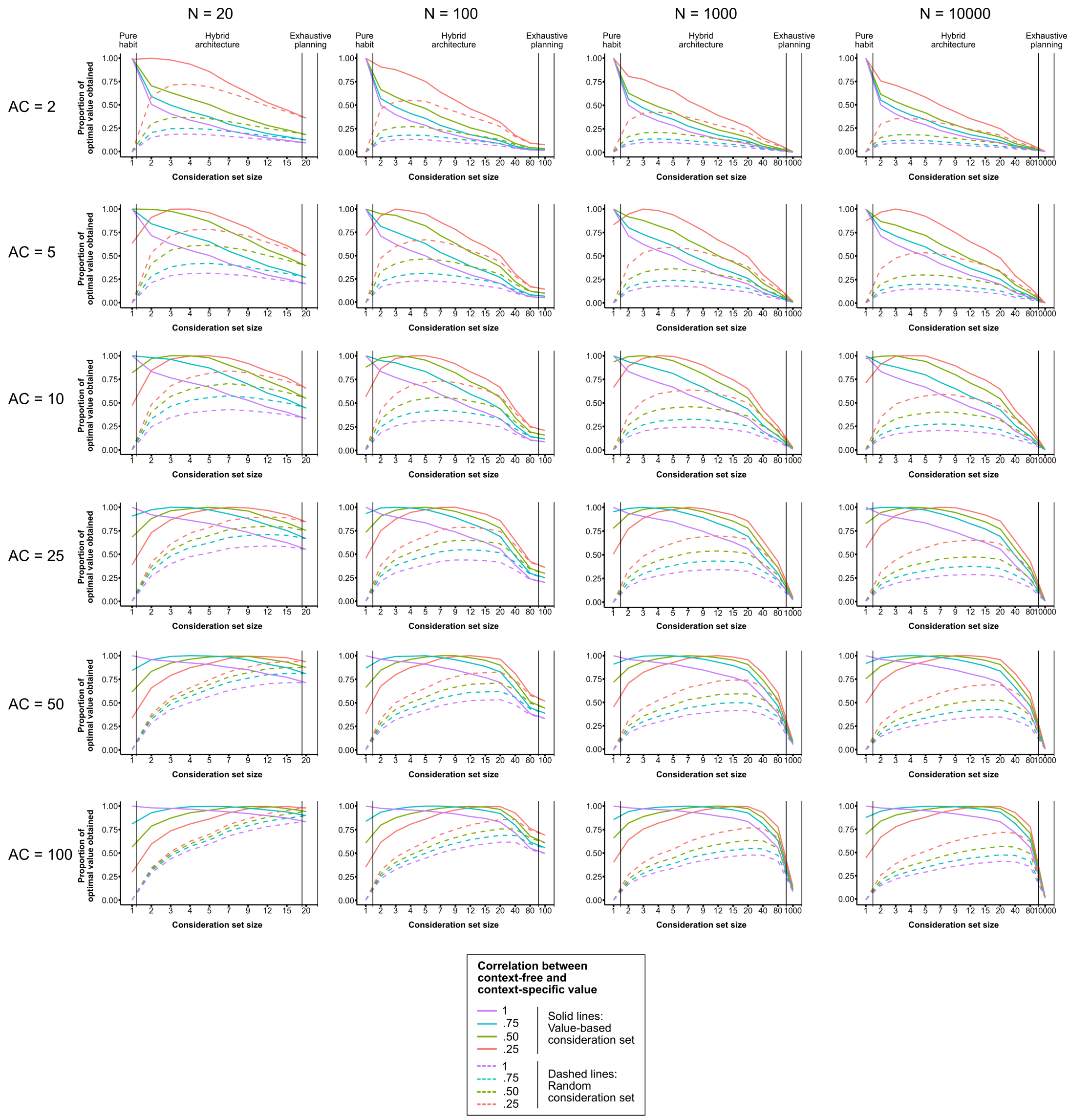

Figure S 1. Simulation results. *AC* represents the cost of performing an action (relative to evaluating it); *N* represents the total number of possible actions. As long as *AC* is not too low and context-free cached values are not too highly correlated with context-specific values, agents do best with our hybrid architecture that uses cached values to generate a consideration set (of size > 1 and less than *N*) and context-specific value to select among the generated actions.

were trying to make your choice?

Please list all the foods you considered while making your decision. It's completely okay to list foods that you thought about, but then immediately realized they weren't good options.

We're giving you 8 spaces to list foods you thought of, but you don't have to use all 8. Don't make up foods that you could have thought of, but didn't;

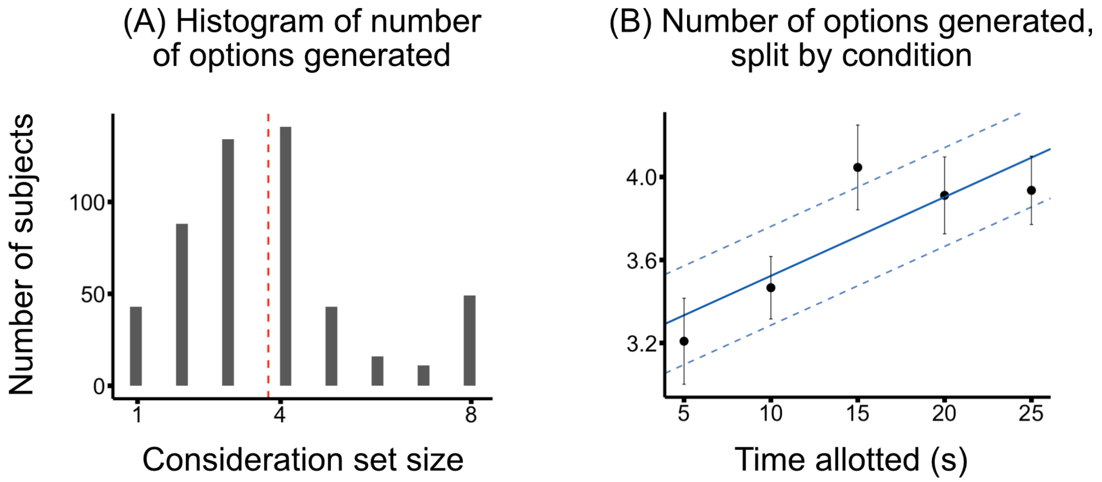

Figure S 2. The results of the first pilot experiment. (A) People reported considering an average of 4 candidates, which aligns with previous work. (B) People reported considering more candidates when they had more time during the decision. (Dashed lines indicate ±1 standard error.)

just list all the ones that really came to mind during the decision.

Averaging across all time pressure conditions, people reported generating a mean of 3.7 [3.5, 3.9] foods (Figure S2A). This aligns with consideration set sizes reported in previous work (Hauser & Wernerfelt, 1990). Moreover, people reported generating more foods when they had been given more time to make the decision (Fig. 2B; linear regression, $b = .036\,[.012, .060], \beta = .13\,[.047, .22], t(523) = 3.04, p = .0025$). These results suggest that people can retrospectively report which options came to mind.

3 Studies 1-2

Here, we give more detail about the first two experiments reported in the main text (the "dinner" experiments; Figure 2 in the main text). In both of these experiments, we ask people to think of a food to make for dinner, where the cached, general values around different foods are deconfounded from their context-specific values in the decision. The two experiments were identical except for the text describing the decision.

3.1 Methods

In Study 1, we asked:

> Imagine that you just got dental surgery, and your doctor gives you food restrictions for the night. You're supposed to avoid food with seeds, foods that require too much chewing, and foods that are moist. What would you cook yourself for dinner tonight?

In Study 2, we administered the same task, but now with a "social" version of the cover story:

> Imagine that, as a gift, you're going to cook dinner for a friend who is currently on crutches and can't cook for themselves. You know that your friend is allergic to anything with seeds, prefers foods that don't require too much chewing, and dislikes foods that are moist. What would you cook them? (The food is only for your friend; you won't eat any of it.)

People had thirty seconds to answer. Then, after eliciting people's consideration sets, we asked people to report the general, cached value of each food they considered ("In general, how much do you like this food?"), and the context-specific value of each food they considered ("In this situation, how good it would be to cook your friend this food?"). They responded to these questions on a scale from 1 to 7. When reporting the general values, 1 was labeled as "this food is among my **least favorite** dishes", 4 as "this food is **average** for me", and 7 as "this food is among my **favorite** dishes". When reporting the context-specific values, 1 was labeled as "this food is among the **worst** dishes I could make in this situation", 4 as "this food would be **average** among the dishes I could make in this situation", and 7 as "this food is among the **best** dishes I could make in this situation". The two value questions were presented in a random order. We excluded anyone who failed to answer all the value rating questions, or who gave the same rating for all foods. There were 202 subjects in the first experiment, and we excluded 17, leaving 185 for analysis; there were 100 subjects in the second experiment, and we excluded 10, leaving 90 for analysis. (Note that our decisions effectively deconfounded cached and situation-specific values; the correlation between cached and situation-specific value ratings was .07 in the first experiment and .24 in the second experiment. Though both of these correlations were significant, they were low enough that we could reasonably perform a statistical control for either when testing for an effect of the other on an outcome of interest.)

3.2 Analysis

We predicted that that people would tend to think of foods that were above average in general value (after controlling for context-specific value), but then would choose among the foods they generated primarily according to context-specific value. The two predictions, one about option generation and one about choice, require different analytic approaches. For the first prediction (candidate generation), we don't know the universe of possible candidates from which they were sampling. Thus, we cannot directly test whether the candidates they generated have higher cached values than the candidates they didn't generate, because we do not know the members of the total set of "thinkable" items from which considered candidates were drawn. Instead, we can ask: Did people generate options with above *average* cached values, as measured by their ratings on the 1-7 scale? If, after controlling for context-specific value, the general values of the generated candidates were significantly above the scale midpoint (the rating for an "average" option), this would provide evidence that candidate generation was influenced by cached value.

The raw data are shown in Figure S3. People tended to generate candidate foods that were above-average in general value, even after controlling for context-specific value. To test this statistically, we regressed the general value ratings on the context-specific value ratings, and found that the intercept was significantly greater than 4 (the midpoint of the scale and the designated rating for an "average" food). In Study 1, the intercept was 5.0 $[4.9, 5.1]$ (comparison to 4, $t(159) = 14, p < .001$); in Study 2, the intercept was 5.2 $[5.0, 5.4]$ (comparison to 4, $t(83) = 14, p < .001$). This tells us that, controlling for context-specific value, the foods that come to mind are more likely to be ones which are good in general.

This result is consistent with people's candidate generation process being influenced by general, cached value. However, there are two concerns with this inference. First, there could be confounding factors that are correlated with general value, but which are actually the ones influencing candidate generation. This concern is impossible to rule out with this design; we address it in the "months" experiments below by performing an exogenous manipulation on cached values. The second worry is that there are simply more foods with above-average cached values (i.e. the distribution of food values is left-skewed), such that random sampling would produce more foods in the upper range of the rating scale. There is some reason to think that this is not a concern—in general, real-life reward distributions tend to be right-skewed (Gigerenzer & Todd, 1999; Stewart, Chater, & Brown, 2006), which means that random sampling would actually produce foods with *lower*-than-average cached values. The distribution of food rewards, of course, could be different. We rule out this concern more rigorously in the "months" paradigm, where the members of the total set of possible actions (i.e., the twelve months of the year) are known.

In sum, in order to draw inferences about the role of value in candidate generation from these data, we rely on a minimal set of assumptions. We now describe these assumptions formally, and describe how we produced the graphs in Figure 2 in the main text.

Let G represent general (i.e., context-free) value, S represent context-specific value, and CS represent whether a candidate was included in the consideration set. For the purposes of this analysis, we consider G and S to be dichotomous: either "high" (above average) or "low" (below average). Dichotomizing the variables allows us to analyze the results using weaker assumptions, as described below. (In order to dichotomize the variables ranging from 1-7, we divided responses at the midpoint rating of the scale (4) evenly into either high or low bins, at random. All results were similar using undichotomized variables.)

From the current design, we can obtain the joint distribution of people's value ratings for the candidates that did come to mind; in other words, we have $Prob(G = g, S = s \mid CS = 1)$ (where $g, s \in \{high, low\}$; Fig S3). However, what we ultimately care about is the probability that a candidate would come to mind, given its value ratings; in other words, we want $Prob(CS = 1 \mid G = g, S = s)$. (Once we have these inferred probabilities, we can test whether a food was more likely to come to mind given $G = high$ versus $G = low$, and the same for S.)

By making some assumptions, we can derive the latter from the former. By Bayes' rule:

$$Prob(CS = 1 \mid G = g, S = s) = \frac{Prob(G = g, S = s \mid CS = 1) \cdot Prob(CS = 1)}{Prob(G = g, S = s)}$$

Our first simplifying assumption is that there are no confounding factors correlated with the joint distribution of G, S that influence whether an item comes to mind; hence, $Prob(CS = 1)$ is constant with respect to G, S. (There can, of course, be other factors that influence CS, so long as they are uncorrelated with $Prob(G, S)$.) Our second simplifying assumption is that the joint distribution of G, S is uniformly distributed; in other words, the universe of foods has an equal number of foods that are above-average and below-average value. (This is why dichotomizing the variables permitted us to use weaker assumptions; otherwise, we would have to assume that G, S are uniformly distributed across the entire 1-7 scale on which participants made their ratings.) Although this assumption is unlikely to be met perfectly, it is probably conservative with respect to our hypothesized effect; food values are in fact most likely left-skewed (Gigerenzer & Todd, 1999; Stewart et al., 2006), so by assuming a uniform distribution we're making it more difficult to find that value increases the likelihood of a food being considered. Under this assumption, $Prob(G, S)$ is constant.

Under these two assumptions, $Prob(CS = 1 \mid G = g, S = s) \propto Prob(G = g, S = s \mid CS = 1)$; the two are equal up to some multiplicative constant (which depends on the total number of possible foods). Hence, we can analyze the variable we care about, $Prob(CS \mid G, S)$, by analyzing the variable we have, $Prob(G, S \mid CS)$. Using this logic, we computed $Prob(CS = 1 \mid G = g, S = s)$ for each participant, and regressed the resulting probabilities on two dummy variables encoding whether $G = high$ or low and $S = high$ or low. The inferred probability of a food coming to mind was significantly higher when general value was above-average, controlling for specific value (Study 1, $b = .26\,[.23, .29], \beta = .47\,[.41, .54], t(737) = 15, p < .001$; Study 2, $b = .31\,[.27, .35], \beta = .57\,[.48, .65], t(357) = 13, p < .001$). The inferred probability of a food coming to mind was also higher when specific value was above average, controlling for general value (Study 1, $b = .085\,[.052, .12], \beta = .16\,[.093, .22], t(737) = 4.9, p < .001$; Study 2, $b = .091\,[.046, .14], \beta = .17\,[.084, .25], t(357) = 3.9, p < .001$). This effect was, however, significantly smaller; both coefficients were below the 95% confidence intervals of the general-value coefficients.

(This statistical approach estimates the effect of each value type while controlling for the other, e.g. the effect of $G = high$ controlling for S. For simplicity of visualization, in the graphs in Figure 2 of the main text, we instead show the marginal effect of each value type. In other words, in Figure 2 of the main text, we show the inferred $Prob(CS = 1 \mid G = high$ or $low)$ and $Prob(CS = 1 \mid S = high$ or $low)$.)

In contrast to candidate generation, analyzing people's choices among the generated candidates is straightforward because we do know the members of each participant's consideration set. For each participant, we ranked the items in their choice set according to their general values, and then separately according to their context-specific values. We ranked items in ascending order (with higher-valued items getting higher rank numbers). Moreover, since relatively few people considered more than five items, we collapsed all items that were fifth-best or worse into the lowest rank of 1. Hence, the highest-valued item was always ranked as 5, and the lowest-value items were either ranked as 1 (if the person considered at least 5 items), or 2 (if the person only considered 4 items), etc. (Ties were assigned equal rank, with the rank of the next-highest value penalized to match the number of ties in the rank preceding it.) This procedure produced ranks that were easy to visualize and interpret. For instance, suppose that the items in a person's consideration set had general value ratings $(5, 6, 4, 2, 7)$ and context-specific value ratings $(2, 2, 3, 6, 7)$. Then their general-value

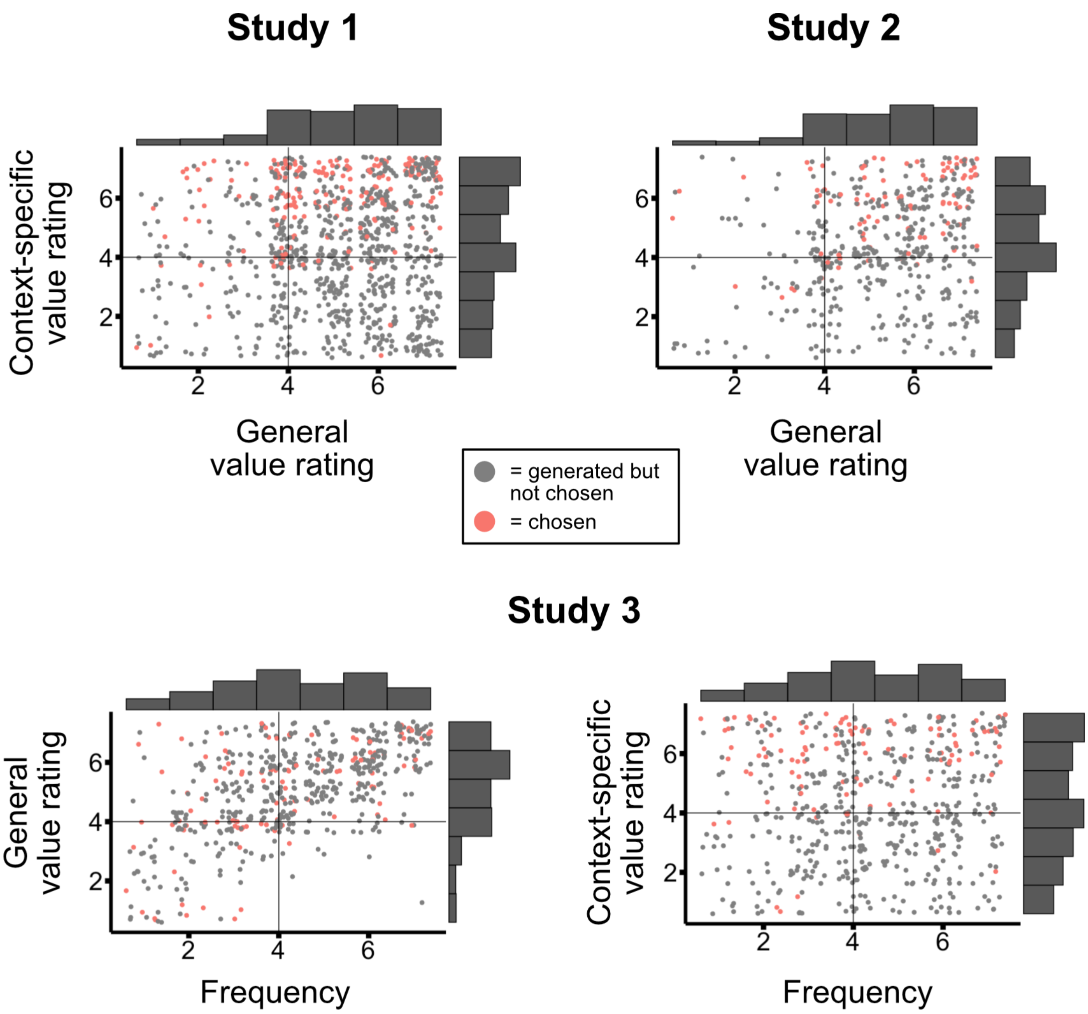

Figure S 3. Raw data from Studies 1-3. Each dot represents a food that came to mind; dot color indicates whether the food was chosen or not. The dots are jittered. People tended to generate items with high general, context-free values and then choose among those according to context-specific value. Historical choice frequency (Study 3) had a relatively weak effect on both generation and choice. Histograms represent the frequency of considered items (i.e., gray plus red dots) within each value septile.

ranks would be $(3, 4, 2, 1, 5)$ and their specific-value ranks would be $(1, 1, 3, 4, 5)$. (Throughout all our analyses of choice from among generated candidates, we use these ranks, rather than the raw values, as predictors. Using ranks avoids some conceptual issues, such as whether the raw values should be normalized by the other values in the consideration set, and generally fits the data better. All our results are similar, however, when using the raw values.)

We then estimated a mixed-effect multinomial logistic regression model, with both general and context-specific values as predictors. We used the "mlogit" R package (Croissant et al., 2012) with normally distributed random parameters per subject and the default Halton sequence for parameter estimation; in these experiments, there were no consistent "items", so we omitted item-specific intercepts. People were much more likely to choose words with high context-specific values (Study 1, $b = 3.3$ $[1.5, 5.1]$, $\beta = 5.1$ $[2.4, 7.8]$, $z = 3.7$, $p < .001$; Study 2, $b = 3.5$ $[.76, 6.2]$, $\beta = 5.3$ $[.99, 9.6]$, $z = 2.5$, $p = .014$); general value either had no, or a slightly negative, effect (Study 1, $b = -.20$ $[-.42, .016]$, $\beta = -.32$ $[-.65, .013]$, $p = .059$; Study 2, $b = .065$ $[-.39, .52]$, $\beta = .10$ $[-.61, .81]$, $z = .29$, $p = .77$).

4 Study 3

In Studies 1-2, we pitted cached values against context-specific values. In Study 3, we also included choice frequency—i.e. how often a food was eaten in the past. Testing for an effect of frequency is interesting because frequency is considered a primary influence on what comes to mind in other types of judgment (Tversky & Kahneman, 1973).

To test this, we replicated Study 1 with an additional question. For each food they had generated, we asked them: "How often do you eat each of these foods, compared to all the types of food you eat?"We gave them a scale from 1-7, where 1 was "this food is among the least common dishes I eat", 4 was "I eat this food an average amount", and 7 was "this food is among the most common dishes I eat".

The raw data are shown in Figure S3. The primary result is that people tended to generate foods with high general value, with both context-specific value and historical choice frequency having a relatively smaller effect. From the inferred probability analysis (Figure 1C in the main text), when all three predictors are entered as simultaneous regressors, the slopes for general value, context-specific value, and choice frequency are (respectively): $b = .13$ $[.11, .15]$, $b = .039$ $[.015, .063]$, $b = .047$ $[.023, .071]$. The slope for general value is above the 95% confidence interval for the latter two slopes.

5 Months experiments (Studies 4-7)

Here, we give details for Studies 4-7 (the "months" experiments; Figure 3 in the main text). In these experiments, we specify a set of available candidates (the twelve months of the year) and manipulate their cached values directly. We describe Study 4, then Study 5 (which was a pre-registered replication of Study 3), then Study 6 (which was a follow-up experiment with minor modifications), then Study 7 (which was a follow-up experiment designed to deconfound value and salience). The results from Studies 4-6 are visualized in Figure S4, and the results from Study 7 are visualized in Figure S5.

5.1 Study 4

Study 4 was our initial months experiment.

Methods All three month experiments were divided into two stages. In Stage I, participants were trained to associate each month with an arbitrary monetary reward. The reward amounts were 1 points through 12 points, assigned randomly to the twelve months for each participant. The training phase consisted of 132 trials; on each trial, participants were given a choice between two of the months (e.g. FEBRUARY vs AUGUST), and won the number of points associated with the month they chose. After the participant made their choice, they were shown the value of both months (e.g. FEBRUARY = 7, AUGUST = 12). Each month was paired with each other month twice, and all months were presented an equal number of times. For the first third of trials, participants were allowed to press a “get hint” button that would reveal the two month values before making a choice, allowing them to learn the values more easily.

Then, in Stage II, participants were told that they were going to make some decisions involving the month names, and that the decision was completely unrelated to the point amounts in Stage I. After reading these instructions, they were told:

> We’ll ask you to think of a word from Stage 1 whose **third letter is late in the alphabet.** You’ll earn bonus money based on the position in the alphabet of the third letter in your word (i.e. A = 1 cent, Z = 26 cents).
>
> For example, the third letter of the word IMPACT is P, which is the 16th letter in the alphabet. Thus, if IMPACT had been one of the words in Part 1, it would earn 16 cents.

Participants were then given a comprehension check question:

> If the word FIZZLE had been one of the words in Part 1, how many cents would FIZZLE earn you?

After receiving these instructions and answering the comprehension check, participants were given 25 seconds to make a decision. Finally, after making a decision, participants were asked to report which words came to mind during the decision:

> Because the last question had a time limit, most people are unable to consider all the words from Part 1 before making their choice.
>
> In Question 2, we'll ask you: Which words **did** you consider while answering the last question, before choosing your final answer? In other words, which words came to mind while you were trying to answer Question 1?
>
> We'll show you one word at a time. Select 'Yes' if that word came to mind at all while answering the question, and 'No' if it never came to mind. (This won't affect your bonus pay, so please answer honestly.)

We then showed participants each month (in random order), and asked them: "Did this word come to mind while you were answering the last question?". Participants could respond with "Yes" or "No".

To parse participants' choices, we compared their response to the list of month names using the restricted Damerau-Levenshtein distance method in the "amatch" function of R package "stringdist" (with a maximum distance of 2; Van der Loo (2014)). We excluded any participants who didn't complete the study, who chose the better alternative in Stage 1 training on less than 70 percent of trials, who failed to choose a month within the time limit in Stage 2, who failed the Stage 2 comprehension check, or who wrote things down physically during the experiment (as measured by self report at the end of the experiment). These exclusion criteria were chosen before analyzing the results.

Study 4 had $N = 502$ subjects, and we excluded 178, leaving 324 subjects for analysis.

Analysis Since people only had 25 seconds to make a decision, they did not have time to consider all the months. We predicted that, when generating months to consider, people would be more likely to think of months that were higher value in Stage I (the month's "cached" value), but would choose between generated months according to their value for the specific situation at hand (i.e. how late the third letter of the word was).

To test the first prediction, we estimated a mixed-effects logistic regression, regressing whether each month came to mind on the month's Stage I value (with random intercepts and slopes for each subject and month). Stage I value had a significant positive influence on the likelihood of a month coming to mind ($b = .032\,[.0085, .056], \beta = .23\,[.052, .41], z = 2.5, p = .011$). Stage II value also ostensibly influenced the likelihood of a month coming to mind ($b = .050\,[.0088, .091], \beta = .72\,[.13, 1.3], z = 2.4, p = .016$). However, this effect is uninterpretable because, in this paradigm, the Stage II value of a month is confounded with the identity of the month itself; the likelihood of October coming to mind, for instance, could be because of October's third letter (i.e. its Stage II value), but it could also be because of extraneous associations specific to October. As in the rest of the paper, we do not make claims about when context-specific values can influence candidate generation. (This issue does not arise for the Stage I values since they were randomly assigned to different months for each participant.)

(Note that in all of our months experiments, when testing for the influence of value on consideration sets, we fit separate regression models for Stage I and Stage II value rather than estimating one regression with both predictors. This approach is valid since Stage I values are randomized for each participant, and it facilitated convergence during statistical parameter estimation. However, all results are similar if both predictors are entered into the same regression model.)

To test the second prediction, we estimated a mixed-effects multinomial logit, regressing people's choice from among the months that came to mind on the Stage I and Stage II value rankings. (We used the same procedure as in the food experiments,

with one exception: We included month-specific intercepts, unless they prevented model convergence.) Choice among the considered months was determined by their Stage II value rank ($b = 2.2\,[1.4, 3.0], \beta = 3.4\,[2.2, 4.6], z = 5.6, p =< .001$); we did not find an effect of Stage I value rank
($b = -.00081\,[-.18, .18], \beta = -.0012\,[-.28, .27], z = -.0088, p = .99, BF_{null} = 312$). Moreover, the slope for Stage I value ranking was significantly below the 95% confidence interval of the Stage II value ranking slope. (When analyzing consideration set construction above we entered Stage I and Stage II predictors into separate regression models because these predictors were orthogonal by design. But, when analyzing choice out of the consideration set, we enter them into the same model. This is necessary because *rankings* of items by Stage I and Stage II value *within* people's consideration sets were no longer orthogonal. Entering these in a single statistical model allows us to estimate their unique effects.)

Note that, when graphing the choice results in the main text, for simplicity we plotted the effects of Stage I and Stage II value rankings separately (Fig. 3E-F in the main text). However, since the rankings are correlated, a more complex but accurate representation of the data can be obtained by plotting the effect of Stage II value rank at each level of Stage I value rank. Those are the plots we present at the bottom of Figure S4.

"Checkmark" shape In Study 4, the effect of Stage I value on candidate generation appeared nonlinear; the worst month also showed an increased likelihood of coming to mind. We explored this nonlinearity in two ways. First, we tested whether our main effect—months with higher Stage I values being more likely to come to mind in Stage II—was robust to excluding the extreme (best and worst) months. It was; after excluding the best and worst months,
$b = .042\,[.0087, .75], \beta = .25\,[.047, .45], z = 2.4, p = .016$. Second, we tested whether there was a quadratic component to the effect. We regressed whether each month came to mind on both the Stage I value and the square of the Stage I value (after centering Stage I value to prevent collinearity). In this regression, both the linear and quadratic

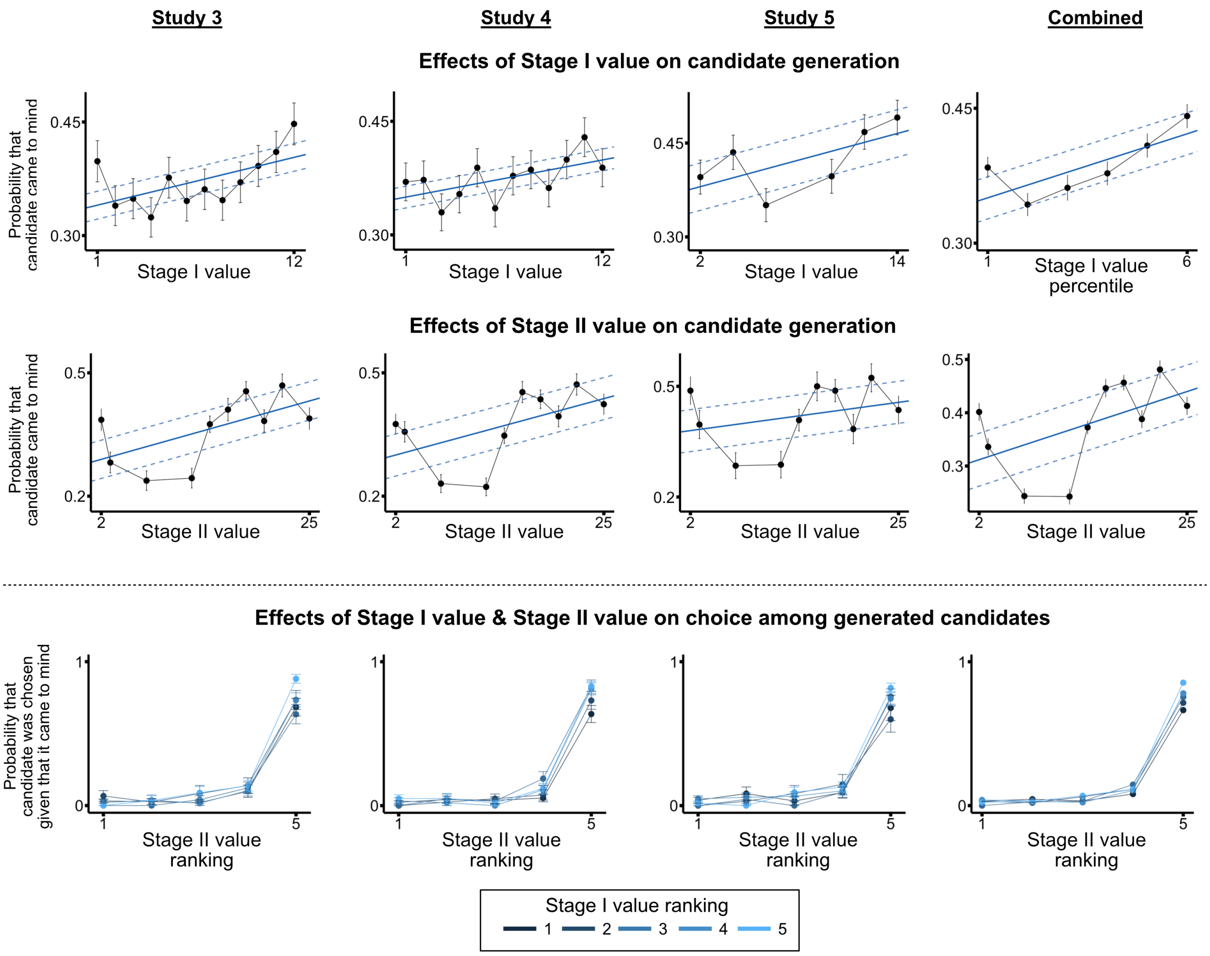

Figure S 4. Results from Studies 4-6, with best-fit lines (dashed lines represent ±1 standard error).

regressors were significant ($b_{\text{linear}} = .029\,[.0055, .052], \beta = .21\,[.036, .38], z = 2.4, p = .018$; $b_{\text{quadratic}} = .012\,[.0038, .020], \beta = .27\,[.091, .45], z = 2.9, p = .003$). This result suggests that there was a nonlinear effect of Stage I value on candidate generation.

5.2 Study 5

Study 5 was a pre-registered replication of Study 4; the pre-registration document can be found at http://aspredicted.org/blind.php?x=3aa3vq. Study 5 had $N = 605$ participants; we excluded 232 (using the pre-registered exclusion criteria, which were identical to the criteria from the first experiment), leaving 373 for analysis.

The results were similar to Study 4. (In this section, as indicated in the pre-registration document, we report one-tailed p values.) The key finding was, again, that people were more likely to think of months for the Stage II decision that were good in Stage I ($b = .021$ [$-.0056, .043$], $\beta = .15$ [$.0027, .30$], $z = 2.0, p_{\text{one-tailed}} = .023$). (People were again more likely to think of months that had high Stage II value ($b = .039$ [$.0037, .074$], $\beta = .55$ [$.052, 1.1$], $z = 2.2, p = .03$) but this result has the same caveat as before.) When choosing among the generated candidates, they again relied on Stage II value rankings ($b = 2.6$ [$1.8, 3.4$], $\beta = 3.9$ [$2.7, 5.1$], $z = 6.3, p < .001$), with no detectable effect of Stage I value rankings ($b = .10$ [$.090, .29$], $\beta = .15$ [$-.14, .44$], $z = 1.0, p = .30, BF_{null} = 57$). Moreover, the slope for Stage I was again significantly below the 95% confidence interval of the Stage II slope.

Recall that in our previous study, Study 4, Stage I value had a "checkmark" shape effect on candidate generation: There was a general positive linear relationship between Stage I value and the probability of a month coming to mind, with a bump for the lowest-value word. To explore this potential nonlinearity, in Study 5 we preregistered the same two tests as in Study 3. First, to ensure that our overall effect—words with higher Stage 1 values being more likely to come to mind—existed without the extreme Stage 1 values, we reran the analysis while excluding the words with the highest and lowest Stage 1 value. The effect remained unchanged ($b = .029$ [$.0016, .056$], $\beta = .17$ [$.0049, .34$], $z = 2.0, p_{\text{one-tailed}} = .022$). Second, to explicitly test for the nonlinearity, we reran the analysis including a quadratic term (i.e. Stage 1 value squared). In Study 5, there was not a clearly detectable checkmark shape; the quadratic term was marginally significant ($b = .0058$ [$-.0013, .013$], $z = 1.6, p_{\text{one-tailed}} = .053$). The checkmark shape, did, however, appear clearly when aggregating Studies 4-6; see the "Combined analysis" section below.

We note two deviations from our pre-registered analysis. First, the pre-registration document indicates a sample size of 600. We posted the study on Amazon Mechanical Turk indicating a desired sample size of 600; the $N = 605$ number includes people who started the study but did not complete it. Second, we pre-registered a slightly different choice analysis. Instead of using value ranks and entering the Stage I and Stage II predictors into the same model, we pre-registered using the raw values and entering the two predictors into separate models. This analysis produces a similar result: Stage II value influenced choice from the consideration set ($b = .63\,[.49, .77], \beta = 4.4\,[3.4, 5.4], z = 8.9, p < .001$), while Stage I value did not ($b = .014\,[-.021, .049], \beta = .050\,[-.073, .17], z = .78, p = .43$). However, for consistency with the other studies (and because of the advantages of using rankings for the choice analysis), we reported the rankings version above.

5.3 Study 6

Studies 4 and 5 demonstrate an effect of cached value on candidate generation. In Study 6, we ran a follow-up designed to test for an additional effect of choice frequency – i.e. how often a month is chosen, controlling for its value. Specifically, we tested whether people would be more likely to think of months that they chose more often in Stage I, even if those months did not have a higher Stage I value. We did not find strong evidence for a role of choice frequency, as described below; but this experiment did provide an additional replication of the effect of Stage I value. (Study 6 had $N = 500$; we excluded 174 participants using the same criteria as the other two month experiments, leaving 326 for analysis.)

In this experiment, we modified the Stage I training such that half the months would have equal Stage I value but would be chosen at different frequencies. Six of the months, chosen at random, were given a Stage I value of 8 points (the “equal-value months”); the other six months had values 2, 4, 6, 10, 12, and 14 points, assigned at random (the “unequal-value months”). Then, in the Stage I training, the equal-value months were presented alongside the unequal-value months such that distinct equal-value months would be chosen at varying frequencies. Specifically, half the equal-value month (“commonly chosen”) were paired with months of lesser value on 14 trials and months of greater value on only 6 trials. The other half of equal-value months (“rarely chosen”) were instead paired with 14 greater-value months and only 6 lesser value months. Thus, the “commonly chosen” equal-value months should be chosen approximately twice as often as the “rarely chosen” equal-value months. By comparing the two types of equal-value months, we can test whether commonly chosen months are more likely to come to mind.

The manipulation worked; people chose the "commonly chosen" equal-value months an average of 13.1 times [13.0, 13.2] and the "rarely chosen" equal-value months an average of 6.3 times [6.2, 6.4]. However, people were not significantly more likely to think of the "commonly chosen" words in Stage II ($b = .12\,[-.096, .34], \beta = .12\,[-.10, .35], z = 1.1, p = .29, BF_{null} = 25.8$).

This experiment also provided a replication of the effect of Stage I values. Among the unequal-value months, months with higher Stage I values were more likely to come to mind ($b = .031\,[.00096, .062], \beta = .28\,[.0083, .54], z = 2.0, p = .043$). (Here, Stage II value did not detectably influence which months came to mind ($b = .02\,[-.019, .059], \beta = .30\,[-.21, .82], z = 1.2, p = .24, BF_{null} = 23$); but, again, this result is not interpretable.) And when choosing among the months which came to mind, people relied on Stage II value rankings ($b = 2.3\,[1.6, 3.0], \beta = 3.5\,[2.4, 4.6], z = 6.3, p < .001$); there was no detectable effect of Stage I value rankings ($b = .13\,[-.086, .35], \beta = .18\,[-.13, .49], z = 1.1, p = .25, BF_{null} = 85$). (Again, the slope for Stage I was significantly below the 95% confidence interval of the Stage II slope.) Thus, this third months experiment provides an additional replication of the patterns observed in the prior two experiments.

5.4 Aggregated analysis

To gain a higher-resolution picture of the small effects in the months paradigm, we aggregated data from the three months experiments. This gave us a total of 1023 subjects for analysis.

We first looked at our key effect: the influence of cached on which candidates came to mind. For this analysis, we combined all twelve months from the first two experiments with the six "unequal-value" months in the third experiment. To make the Stage I values of these months comparable, we divided the values into six percentile bins; so, e.g., the lowest percentile would include each subject's two least valuable months from the first two experiments and each subject's single least valuable month from the third experiment. (The results are similar if the months are aggregated with their raw values.) We then regressed whether each month came to mind on this Stage I value percentile. As in each experiment alone, months with higher Stage I values were more likely to come to mind ($b = .055\,[.030, .080], \beta = .20\,[.10, .29], z = 4.1, p < .001$).

We then tested for an aggregate effect of Stage II values on which months came to mind. (For this analysis, since the Stage II values were identical across the three experiments, we did not need to bin the values by percentile to aggregate them.) Words with higher Stage II values were more likely to come to mind ($b = .18\,[.027, .33], \beta = .62\,[.10, 1.1], z = 2.3, p = .021$), although this result has the same caveat as before: Since Stage II values are not randomized across participants, they are confounded with the months themselves (i.e. "May" could come to mind more often because its third letter is late in the alphabet, or because it happens to just be a salient month).

Finally, we tested for an aggregate effect of Stage I value and Stage II value on selection among generated candidates, using the same mixed-effect multinomial logit analyses as above. There was a strong effect of Stage II value ($b = 2.3\,[1.9, 2.7], \beta = 3.5\,[2.9, 4.1], z = 10.5, p < .001$), and no detectable effect of Stage I value ($b = .054\,[-.050, .16], \beta = .08\,[-.075, 2.3], z = 1.0, p = .31, BF_{null} = 313$). These results reaffirm the patterns observed in each experiment individually.

5.5 Study 7

Study 7 was a follow-up study designed to test whether our results could be explained by a salience confound. The logic is described in detail in the main text. If people are more likely to generate items high in cached value, then, collapsing across the "gains" and "losses" conditions, we should see a main effect of Stage I value; if people are instead more likely to generate items high in absolute extremity (i.e. furthest from zero), then we should see an interaction between Stage I value and condition; and if people are doing a mixture of both, we should see both a main effect and an interaction (i.e. a significant simple effect in the gains condition and an attenuated or nonexistent effect in the losses condition).

We preregistered the following procedure for determining sample size:

> We will stop data collection when we either (a) obtain 2000 participants who pass the exclusion criteria (i.e. N = 2000 after exclusion), (b) obtain 4000 participants pre-exclusion, or (c) the rate at which we are collecting new participants falls below 50 per day. The first criterion requires us to "look at the data" in order to determine the stopping point, but we will not examine the effects of interest, and since the exclusion criteria are predetermined (as described above) this data collection procedure is not biased.

Option (c) occurred; the rate at which we were able to collect new participants on Amazon Mechanical Turk dropped below 50 per day, and so we stopped collecting more participants. This left us with 2745 total participants and 1787 participants who passed the preregistered exclusion criteria.

For our analysis, one detail is worth noting: We coded condition as -0.5 for the losses condition and 0.5 for the gains condition, and coded Stage I word values by their rank within the range of word values available in the participant's condition (so the lowest-valued word was coded as 1, the highest-valued word was be coded as 12, and the others as 2-11 accordingly). This coding scheme supported a straightforward interpretation of the main effect as the average effect of Stage I value across the conditions, and the interaction as the difference between slopes in the two conditions. We also centered Stage I value before entering it into the regression.

The results supported our hypothesis (Fig. S**??**). When regressing whether each word came to mind on condition, Stage I value, and their interaction, people showed a main effect of Stage I value ($b = .013\ [.0036, .022], \beta = .090\ [.022, .16], z = 2.6, p_{\text{one-tailed}} = .0045$), as well as a marginal interaction ($b = .017\ [-.0085, .042], \beta = .060\ [-.028, .15], z = 1.3, p_{\text{one-tailed}} = .09$). As predicted by the "mixture" account, there was a significant simple effect of Stage I value in the gains condition ($b = .021\ [.0065, .036], \beta = .15\ [.048, .25], z = 2.9, p = .0040$) and a weak-to-nonexistent effect in the losses condition ($b = .0038\ [-.012, .019], \beta = .027\ [-.083, .14], z = .48, p = .63$). (There was no significant effect of Stage II value on whether a word came to mind ($b = .03\ [-.0062, .068], \beta = .44\ [-.096, .98], z = 1.6, p = .11$) or interaction between Stage II value and condition ($b = -.005\ [-.021, .011], \beta = -.035\ [-.15, .076], z = -.62, p = .54$), although, as described above, this result is uninformative.) As in Studies 4-6, which option was selected out of the consideration set was determined almost entirely by Stage II value ($b = 1.9\ [1.7, 2.1], \beta = 2.9\ [2.6, 3.3], z = 17, p < .001$), not Stage I value ($b = -.01\ [-.069, .049], \beta = -.018\ [-.12, .082], z = -.35, p = .73, BF_{null} = 1700$).

6 Negation experiments (Studies 8-9)

Here, we report the details of the two experiments where we pitted cached and context-specific values against each other (the "negation" experiments; Fig 4 in the main text). In both experiments, we asked people to make decisions where the online, context-specific value of each candidate was exactly anticorrelated with the cached value, and compared these to decisions where the cached and context-specific values were aligned. We describe each experiment in turn.

6.1 Study 8

The first experiment involved choosing a food that people would either most or least want to eat for dinner. Specifically, we asked half of participants ("think of best" condition):

> Imagine that someone has offered to cook you dinner tonight. What meal would you most want for the dinner? (Please limit your answer to normal meals that someone would reasonably cook.)

And we asked the other half of participants ("think of worst" condition):

> Imagine that someone has offered to cook you dinner tonight. What meal would you least want for the dinner? (Please limit your answer to normal meals that someone would reasonably cook.)

We then gave people our consideration set measure, and asked them to rate the general value of each food that came to mind. As in the earlier dinner experiments, we asked, for each food that came to mind: "How much do you like each of these foods?" We gave participants a scale from 1-7, where 1 was labeled as "this food is among my **least favorite** dishes", 4 as "this food is **average** for me", and 7 as "this food is among my **favorite** dishes".

Analysis Study 8 had $N = 500$ participants, and we excluded 113 (using the same exclusion criteria as in the earlier dinner experiments), leaving 387 for analysis.

To test whether general cached values persisted in influencing candidate generation in this setting, we examined how often people generated foods that were the "opposite" valence from the current decision. Specifically, we examined how often people thought of above-average foods (greater than 4 on the 1-7 scale) when they were supposed to be thinking of bad ones, compared to how often people thought of below-average foods when they were supposed to be thinking of good ones. We estimated a mixed-effects logistic regression model, regressing whether each generated food was the opposite valence from the current decision on each participant's decision condition. In both experiments, people were more likely to generate opposite-valence foods in the "think of worst" condition ($b = 3.1\ [2.5, 3.7], \beta = 3.8\ [3.1, 4.5], z = 11, p < .001$). (Here in the Supplement, we also visualize the same results, but without dichotomizing value into above-average and below-average. Specifically, we plot the total number of candidates of each value generated by subjects (Fig. S6A). The same pattern holds; people in the "think of worst" condition generated many more candidates that were low in context-specific value, i.e. foods that were generally good.)

Moreover, this effect was not driven by increased confusion in the "think of worst" condition. People in the "think of worst" condition were actually slightly better at selecting the correct context-specific food from among the foods they generated (Fig

S6A; 73% [67%, 79%] chose the best food in their consideration set in the "think of best" condition, while 76% [70%, 82%] chose the worst food in their consideration set in the "think of best" condition).

6.2 Study 9

In the second negation experiment, we applied the same design to our "months" paradigm. People went through the same Stage I training as in the other month experiments, with one change: The month values were now stochastic (each drawn from a Gaussian with a randomly assigned mean of 1-12, each with a standard deviation of 1.75). Then, in Stage II, we asked half of participants, "What was the best month to choose in Part 1?", and we asked the other half, "What was the worst month to choose in Part 1?". We then administered the same consideration set measure as in the other month experiments.

Analysis As mentioned in the main text, we collected data in the paradigm of Study 9 on two separate instances. The first instance was to establish the effect reported in this paper; the second was as part of piloting a larger follow-up design which we did not ultimately pursue. The effect in the first instance was large and significant ($N = 135, b = 1.1\ [.59, 1.6], \beta = 1.2\ [.67, 1.8], z = 4.1, p = .000036$); the second instance showed the same directional effect as the first, but was not significant ($N = 260, b = 0.22\ [-.11, .55], \beta = .25\ [-.12, .63], z = 1.3, p = .18$). (This non-significant result would not be too surprising, assuming the effect size obtained by aggregating the two data sets; a post-hoc bootstrap power analysis showed that a sample of that size, with the effect size from the aggregated data, would only have a 60% chance of obtaining a significant result.) For transparency, we include all data we had collected in this paradigm in order to report our best estimate of the effect. For the remainder of the analysis, we report the results from the aggregated data, and note that these results should be viewed as suggestive but preliminary and needing further replication.

Study 9 had $N = 457$ total, and we excluded 62, leaving 395 for analysis. For Study 9, we used the same exclusion criteria as in the months experiments above, with two changes. First, since the Stage I training was now stochastic, we dropped the criterion that people had to make the correct Stage I choice in over 70% of trials. Second, we added a question asking whether people had done an experiment very similar to this one before, and excluded anyone who answered yes.

The analysis was similar to Study 8. We characterized a month as "opposite-valence" if it was above-average in the "think of the worst month" condition (i.e. had a Stage I mean value greater than 6) or below-average in the "think of best" condition (i.e. had a mean value less than 7). People were again more likely to think of opposite-valence words in the "think of worst" condition ($b = 0.50\ [.23, .77], \beta = .57\ [.25, .89], z = 3.5, p < .001$). (We again visualize the undichotomized results in Fig. S6.) And again, this effect was not driven by increased confusion in the "think of worst" condition: People were slightly better at selecting the correct context-specific month in the "think of worst" condition (Fig S6B; 48% [41%, 55%] chose the best month from their consideration set in the "think of best" condition, versus 57% [50%, 64%] chose the worst month from their consideration set in the "think of worst" condition).

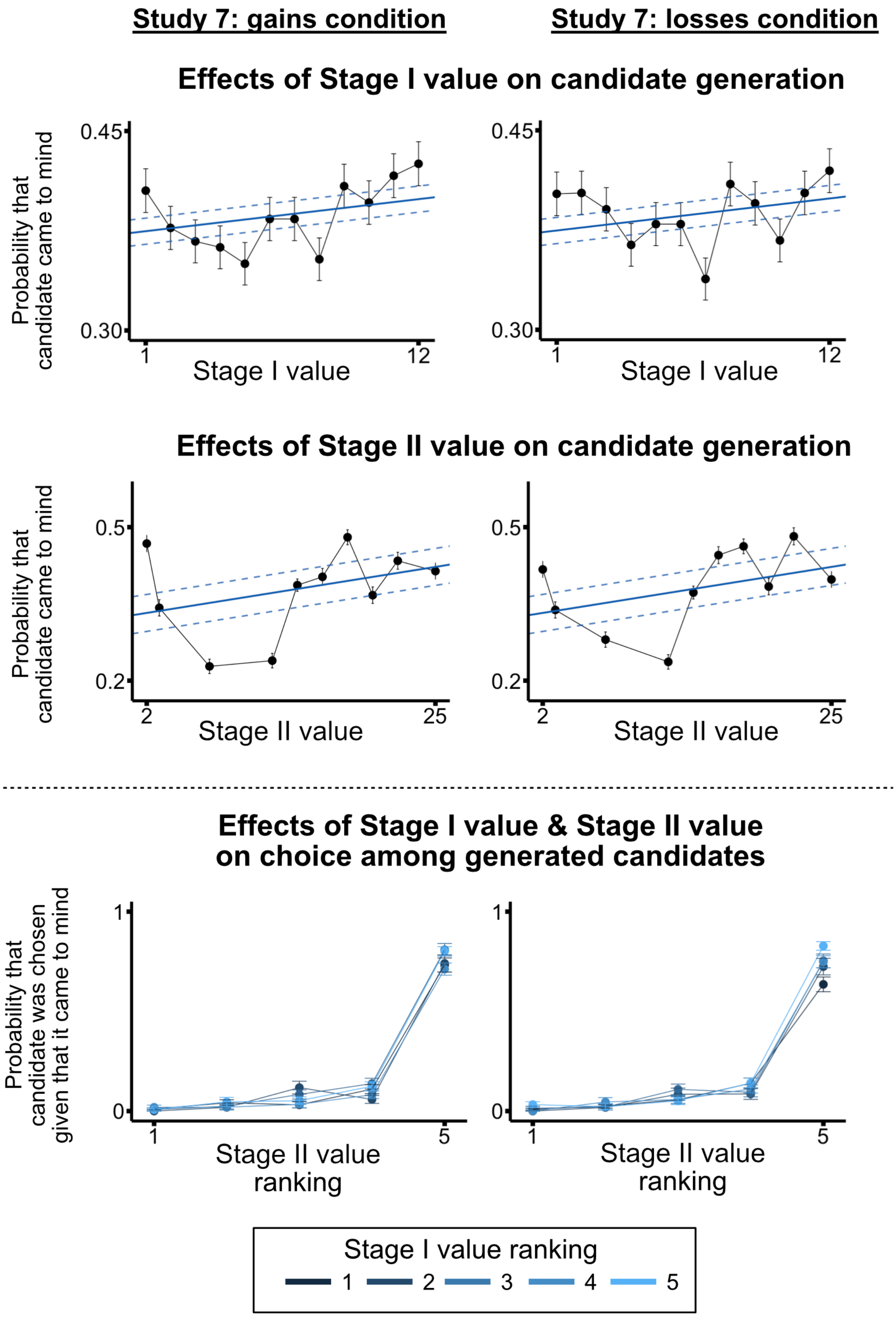

Figure S 5. Results from Study 7, with best-fit lines (dashed lines represent ±1 standard error).

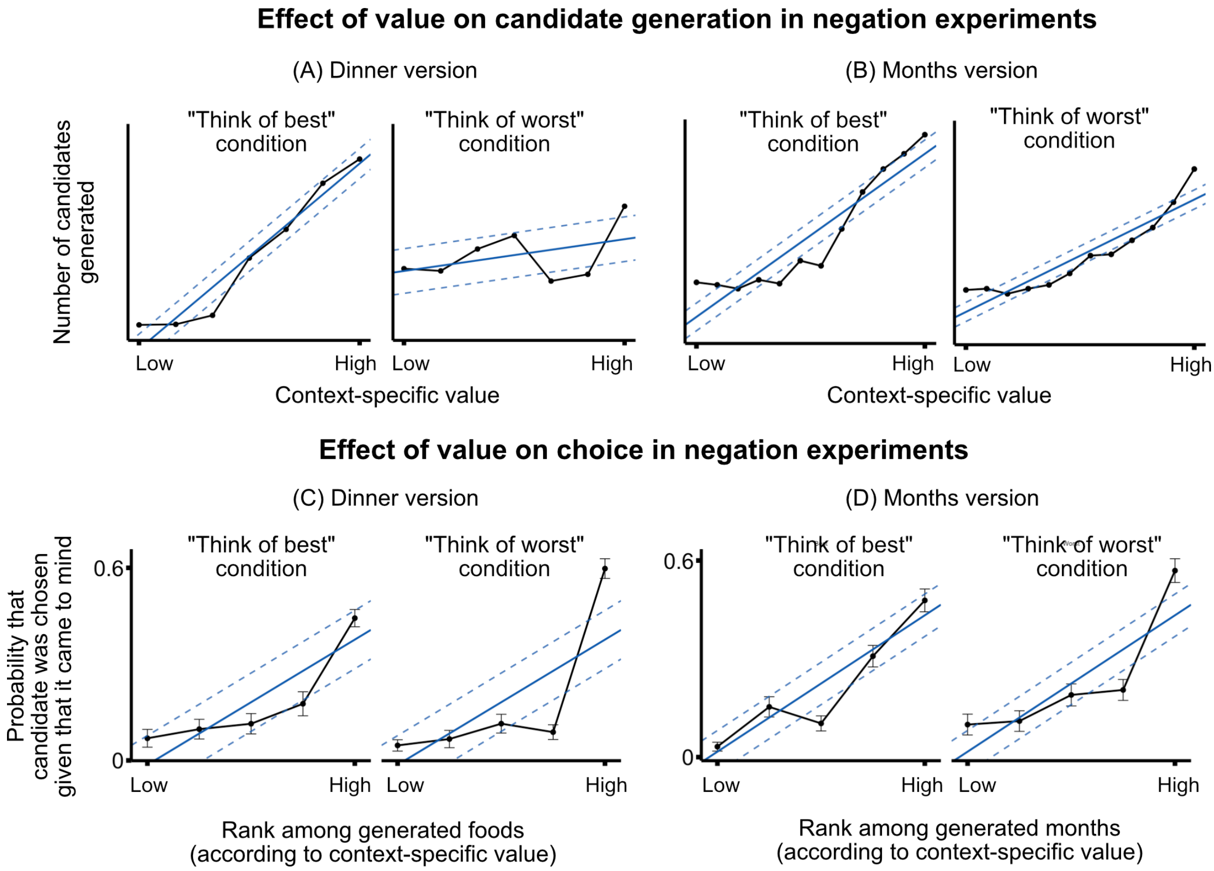

Figure S 6. (A-B) The total number of candidates of each value generated by subjects in Studies 8 and 9. Here, the x-axis denotes context-specific value—so in the "think of best" condition, it was good foods/months that were high in context-specific value, and in the "think of worst" condition it was bad foods/months that were high in context-specific value. In the "think of worst" condition, candidates with low context-specific value (i.e. foods/months with good cached values) were much more likely to intrude on people's minds compared to the "think of best" condition. (C-D) Probability of a candidate (food or month) being chosen given that it came to mind, in Studies 8 (C) and 9 (D), as a function of its context-specific rank in the decision-maker's consideration set. For example, if a person was in the "think of best" condition, then the best food in their consideration set would be ranked first and the worst food would be ranked last; if a person was in the "think of worst" condition, then the worst food in their set would be ranked first, and the best food would be ranked last. Unsurprisingly, people were likely to choose the highest-ranked word in their consideration set. The important result is that people were equally good at selecting candidates in the "think of worst" conditions, suggesting that the results in (A-B) are not because people were more confused in those conditions.

7 References

Croissant, Y., et al. (2012). Estimation of multinomial logit models in r: The mlogit packages. *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf*.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* Oxford University Press, USA.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of consumer research*, *16*(4), 393–408.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1–26.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). MIT press Cambridge.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Van der Loo, M. P. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*(1), 111–122.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive science*, *38*(4), 599–637.