Psychological Review

Process and Content in Decisions From Memory

Wenjia Joyce Zhao, Russell Richie, and Sudeep Bhatia

CITATION

Zhao, W. J., Richie, R., & Bhatia, S. (2021, September 2). Process and Content in Decisions From Memory. *Psychological Review*. Advance online publication. http://dx.doi.org/10.1037/rev0000318

Process and Content in Decisions From Memory

Wenjia Joyce Zhao[1, 2], Russell Richie[2], and Sudeep Bhatia[2]
[1] Department of Psychology, The Ohio State University
[2] Department of Psychology, University of Pennsylvania

Information stored in memory influences the formation of preferences and beliefs in most everyday decision tasks. The richness of this information, and the complexity inherent in interacting memory and decision processes, makes the quantitative model-driven analysis of such decisions very difficult. In this article we present a general framework that can address the theoretical and methodological barriers to building formal models of naturalistic memory-based decision making. Our framework implements established theories of memory search and decision making within a single integrated cognitive system, and uses computational language models to quantify the thoughts over which memory and decision processes operate. It can thus describe both the *content* of the information that is sampled from memory, as well as the *processes* involved in retrieving and evaluating this information in order to make a decision. Furthermore, our framework is tractable, and the parameters that characterize memory-based decisions can be recovered using thought listing and choice data from existing experimental tasks, and in turn be used to make quantitative predictions regarding choice probability, length of deliberation, retrieved thoughts, and the effects of decision context. We showcase the power and generality of our framework by applying it to naturalistic binary choices from domains such as risk perception, consumer behavior, financial decision making, ethical decision making, legal decision making, food choice, and social judgment.

Keywords: memory, judgment, decision making, computational model, vector semantics

Supplemental materials: https://doi.org/10.1037/rev0000318.supp

Memory is widely considered to be one of the fundamental psychological processes at play in decision making, and a large body of work has attempted to characterize the ways in which memory and decision processes interact with each other to influence behavior (Alba et al., 1991; Dougherty et al., 1999; Goldstein & Gigerenzer, 2002; Johnson et al., 2007; Marewski & Mehlhorn, 2011; Schooler & Hertwig, 2005; Shadlen & Shohamy, 2016; Tversky & Kahneman, 1973). This work spans numerous disciplines, and forms one of the main avenues through which research on human cognition guides practical applications in the social and behavioral sciences. Of course, the study of memory-based decision making is not only important for its practical applications. Memory and decision making form two of the major areas of research on the human mind (see recent reviews in Busemeyer et al., 2019 and Kahana, 2020), and integrating models from these two areas within a tractable framework is one of the major theoretical challenges in cognitive psychology today.

Addressing this challenge has been difficult. Both memory and decision making are highly complex. Each involves dynamic and stochastic processes that operate over the contents of the mind. Combining these processes within a single model whose parameters can be fit to experimental data and be used to make quantitative predictions is nontrivial. Moreover, naturalistic memory-based decisions involve the retrieval and evaluation of complex thoughts that reflect learnt knowledge, past experience, emotion, and intuition, as well as sophisticated reasons. In order to model naturalistic memory and decision processes researchers need to first observe and quantify the informational content of these thoughts. This too poses significant challenges.

Consider, for example, the decision prompt "is nuclear power safe?." In order to respond, decision makers must retrieve, from memory, information supporting each of the two response options ("yes" or "no"), typically in the form of natural language thoughts (e.g., "it is difficult to dispose of nuclear waste," "nuclear power is a substitute for harmful energy sources like fossil fuels," and "disasters like Chernobyl are always possible"). Decision processes must evaluate and aggregate these thoughts in order to make a decision. Of course, the retrieval of one thought may cue the retrieval of another, and thus memory involves nuanced dynamics that guide and constrain the eventual decision. The goal of the theorist is to model this complex process, but it is not currently clear how this can be done. How can we observe and quantify the content

of people's thoughts, formalize the mechanisms that people use to retrieve, evaluate, and aggregate these thoughts, and predict response probabilities, deliberation length, thought content, and other decision variables?

Fortunately, there have been numerous advances in psychology that can help us solve this problem. One such advance is query theory, which proposes that thought retrieval is sequential and susceptible to feedback effects (Johnson et al., 2007; Weber et al., 2007). Query theory also offers a powerful experimental paradigm for jointly eliciting memory data (in the form of retrieved thoughts) and decision data, in naturalistic decision settings. In this article, we use the conceptual structure and empirical paradigm offered by query theory as the basis for a quantitative framework for modeling memory-based decisions, such as those discussed in the preceding paragraph.

Another insight is provided by memory models like search of associative memory (SAM; Raaijmakers & Shiffrin, 1981), context maintenance and retrieval (CMR; Polyn et al., 2009), and more general cognitive architectures like adaptive control of thought-rational (ACT-R; Anderson et al., 2004; for ACT-R models of decision making, see also Dimov et al., 2020; Fechner et al., 2018; Gonzalez et al., 2003; Marewski & Mehlhorn, 2011). These theories propose that the items retrieved from memory are used to update a dynamic context representation (or alternatively a short-term memory store, declarative memory or imaginal buffer, or working memory representation), which guides subsequent retrieval and interacts with other processes in the cognitive system. We assume that this type of context also encodes all decision-relevant information, which allows us to tractably represent the interactions between memory and decision making processes and predict the eventual decision.

A third key insight for our framework involves the use of sentence vectors, obtained from recent deep-learning-based language models trained on large-scale language data (Cer et al., 2018; Devlin et al., 2018). Sentence vectors provide a powerful contemporary extension to word-based semantic models (see reviews in Bhatia et al., 2019; Lenci, 2018; Mandera et al., 2017) by accommodating nuances in meaning in natural language sentences. We illustrate the value of these vectors by using them to quantify and cluster semantically related thoughts, and subsequently model sequences of retrieved thoughts.

By combining the above insights, we are able to build a tractable computational framework for studying naturalistic memory-based decisions. Our framework integrates numerous existing models of memory and decision making, and can be used to quantitatively predict how people think and decide, as well as how various contextual factors alter these thoughts and decisions. We illustrate the power, generality, and tractability of our approach, by using it to fit 576 distinct memory and decision models to data from multiple experiments involving eight different psychological domains. Our fits reveal the precise set of memory and decision mechanisms at play in our experiments, and we illustrate the importance of these mechanisms with a wide range of qualitative and quantitative tests. In doing so we provide a useful computational approach for formally specifying and testing the psychological substrates of everyday judgment and decision making.

Theoretical Background

Judgment and decision making involve the subjective evaluation and integration of information to generate a discrete (e.g., choice) or continuous (e.g., rating) response. Models in this field try to characterize mental processes at play in the evaluation and integration of information (Busemeyer et al., 2019; Oppenheimer & Kelso, 2015). Heuristic models, for example, are algorithmic rules which simplify the evaluation and integration process (relative to rational models that use all available information;[1] Gigerenzer & Gaissmaier, 2011; Payne et al., 1988; Simon, 1956); sequential sampling models are stochastic processes that aggregate information over time, typically until a decision threshold is reached (e.g., Bhatia, 2013; Busemeyer & Townsend, 1993; Lee & Cummins, 2004; Roe et al., 2001; Turner et al., 2018; Usher & McClelland, 2004); and connectionist models are dynamical systems that describe decision making as the outcome of spreading activation processes in units of interconnected neurons (Glöckner & Betsch, 2008; Holyoak & Simon, 1999; Suri et al., 2020). Many of these models also make predictions about the specific information that is sampled and used in the decision, which can be tested using eye-tracking, mouse-tracking, or thought listing "process-level" data (Schulte-Mecklenbeck et al., 2011).

There is also a rich body of research in psychology that formally models the complexities inherent in memory search (see Kahana, 2020 for a recent review). For example, models of free recall, such as the SAM and the CMR models specify the processes at play in the retrieval of previously presented information, and can describe whether or not certain items are retrieved from memory, as well as the order in which they are retrieved and how previously retrieved items cue subsequent recall (Atkinson & Shiffrin, 1968; Hintzman, 1984; Murdock, 1982; Polyn et al., 2009; Raaijmakers & Shiffrin, 1981). Related research on semantic memory search describes retrieval dynamics when information is not presented explicitly prior to recall (but rather is stored in memory after years of past experience; Abbott et al., 2015; Hills et al., 2012). Most of the above memory models characterize representations for items in memory using semantic spaces built on word co-occurrence statistics in large-scale language data (Jones & Mewhort, 2007; Landauer & Dumais, 1997). Such spaces specify words and concepts as high-dimensional vectors, and quantify the semantic relatedness or association between pairs of words or concepts using distances between words in the semantic space (see Bhatia et al., 2019; Jones et al., 2015; Lenci, 2018; Mandera et al., 2017 for reviews).

Our goal in this article is to build a framework that can combine models of memory search (which specify the dynamics of information retrieval) with models of decision making (which specify the dynamics of information integration) in order to quantitatively model the cognitive processes and information content involved in memory-based decision making. Although there have been many insightful models of memory and decision making proposed previously (Bhatia, 2013; Dougherty et al., 1999; Goldstein & Gigerenzer, 2002; Hertwig et al., 2008; Juslin & Persson, 2002; Stewart et al., 2006; Thomas et al., 2008; von Helversen & Rieskamp, 2008), most of these do not specify the nuances of the retrieval dynamics involved in this process (as pure memory models do for free recall and memory search), and do not attempt to explain how retrieved information is dynamically aggregated by decision makers to determine choice (as pure decision models do in preferential choice). In addition, these models are typically applied to abstract decision tasks, and are unable to specify the content of the

---

[1] Although of course using all available information might not be adaptive or resource rational (Lieder & Griffiths, 2020; Payne et al., 1988).

information that is retrieved from memory in more naturalistic decisions.

One exception involves models within the ACT-R framework (Anderson, 2007; Anderson et al., 2004). ACT-R provides a powerful approach to (a) modeling retrieval processes that operate on declarative memory, and (b) specifying their interactions with control processes that determine subsequent actions. Such actions include memory probes to obtain further information, logical rules for reasoning, and motor actions to indicate responses. In the domain of decision making, ACT-R based models have been used to implement a variety of decision strategies and evaluative processes that aggregate information in different ways, addressing different paradigms and settings that require high-level judgment and decision making (Dimov et al., 2020; Fechner et al., 2016; Link et al., 2016; Link & Marewski, 2015; Marewski & Mehlhorn, 2011; Marewski & Schooler, 2011; Schooler & Hertwig, 2005). At the core of these models is the use of intermediary cognitive buffers (either the retrieval buffer for the declarative memory module or the imaginal module) to represent dynamically evolving task-relevant information. Information that is retrieved from declarative memory is placed into these buffers, and various production rules operate on the content of these buffers to guide deliberation. With this approach it is possible to describe multiattribute decision heuristics like take-the-best, (Dimov et al., 2020; Marewski & Mehlhorn, 2011) and also specify their relationship with the decision environment and other task-relevant variables (Marewski & Schooler, 2011; Schooler & Hertwig, 2005). This work is also important in its use of ecological data (e.g., natural language statistics) to model the accessibility of items in memory—this allows for the application of the models to certain types of naturalistic decision problems involving real-world concepts (Marewski & Schooler, 2011; Link et al., 2016; Link & Marewski, 2015; also see Anderson & Schooler, 1991 and Goldstein & Gigerenzer, 2002).

Another important theoretical framework relevant to this article is query theory (Hardisty et al., 2010; Johnson et al., 2007; Weber et al., 2007). Query theory attempts to describe the role of memory search in naturalistic judgment and decision tasks. It proposes that responses in these tasks are generated by sequentially querying memory for knowledge, experience, emotions, or reasons relevant to the decision. These queries yield a sequence of thoughts, which determine the decision maker's final response. Crucially, thoughts that are retrieved earlier on influence the retrieval of subsequent thoughts, and contextual cues (such as primes, emotions, or endowments) can bias the decision by altering the thoughts that are the first to be retrieved. The thoughts that lead up to a decision can be observed through query theory's experimental paradigm, which asks decision makers to list the thoughts that come to their mind as they deliberate, and to stop listing thoughts as soon as they have made a decision. For example, Johnson et al. (2007) showed that, in a standard endowment effect paradigm, sellers (as opposed to buyers) tended to (a) list more thoughts supporting high valuation of their endowment, and (b) list such thoughts earlier in the sequence, and (c) these tendencies predicted higher valuations of the endowment. Similar studies have been conducted to explain asymmetric discounting in intertemporal choice (Weber et al., 2007), and the effects of attribute framing on consumer choice (Hardisty et al., 2010).

This list of retrieved thoughts in the query theory paradigm is analogous to the list of words or concepts retrieved in standard free recall and memory search tasks.[2] Thus, in this article, we use memory models of recall dynamics to describe the cognitive processes at play in thought retrieval. We likewise apply extensions of semantic space models that are commonly used to describe memory search, to quantify the content of the thoughts over which memory processes operate. We combine these memory mechanisms with established decision models, that describe the cognitive processes involved in integrating information provided by retrieved thoughts into a response. We do this using a dynamically evolving context representation of the type used in the CMR and SAM memory models (Polyn et al., 2009; Raaijmakers & Shiffrin, 1981). As with ACT-R models of decision making (Dimov et al., 2020; Marewski & Mehlhorn, 2011) we assume that context holds both memory and decision-relevant information—retrieved thoughts enter context, and influence both the retrieval of subsequent thoughts and the formation of decision variables used in evaluation and response generation.

An important assumption of ours is that memory and decision making not only interact through context, but that they interact *only through context*. This allows us to model memory and decision processes as being conditionally independent on context, and specify these processes using distinct modeling components that can be separately fit to data. We demonstrate the value of this assumption, and the resulting tractability of our framework, by fitting 576 distinct memory-based decision models to thought listing and decision data obtained from the query theory experimental paradigm. Our switchboard analysis (Turner et al., 2018) sheds light on the best combination of memory and decision mechanisms for characterizing naturalistic memory-based decisions.

Overview of Framework

Decision Task

We examine decision prompts involving eight different naturalistic decision domains: risk perception, consumer behavior, financial decision making, ethical decision making, legal decision making, food choice, decisions about well-being, and decisions about society and culture. We deliberately selected questions that were not likely to elicit homogeneous responses across participants and questions for which individuals would not display a strong prepotent bias (as with ideologically and politically charged questions), both of which would drastically reduce the role of online memory search and decision dynamics.

Following query theory's experimental paradigm (illustrated in Figure 1), each trial presented participants with a decision prompt (e.g., "Is nuclear power safe?"), and then asked them to list the thoughts that came to their minds as they deliberated. Participants were asked to use separate text boxes for separate thoughts and to list the thoughts in the order in which they occurred. They were asked to stop listing thoughts once they had made a decision, and proceed to a subsequent screen in which they recorded their decision. This decision involved selecting one of two response options (e.g., "yes" or "no") based on their subjective preference or belief. Finally, after the decision had been made, participants were asked to rate each of their thoughts, on a 7-point scale, based on how strongly the thought supported one of the response options over the other.

---

[2] Of course, while the resulting data in free recall and query theory are structurally similar, the different tasks may lead to different retrieval dynamics and search strategies, due to different underlying goals.

Figure 1

An Illustration of the Query Theory Experimental Paradigm

Note. Participants are asked to list the thoughts that come to their mind as they think about the decision (first screen/left panel). When they have made up their minds, they are asked to stop listing thoughts and make a response (second screen/middle panel). After the decision, they are asked to rate each of their listed thoughts in terms of support for the two response options (third screen/right panel). Note that the wording and rating scale illustrated below are simplifications of those used in our experiments (see text for details).

Experiments 1a–1h elicited decisions in each of these eight domains, without attempting to manipulate participant responses. Experiments 2a and 2b used explicit primes to bias participant thoughts and eventual decisions. Here, half the participants were first asked to list a reason supporting the "yes" response (e.g., why nuclear power *is* safe) and the other half were first asked to list a reason supporting the "no" response (e.g., why nuclear power *is not* safe). After listing these reasons, participants were allowed to deliberate freely, and, as in the first eight experiments, were asked to list their thoughts in the order in which they occurred, and select their preferred response once they had made a decision. The decision prompts, associated response options, number of analyzable participants, and basic descriptive statistics of thoughts and choices in these 10 experiments are displayed in Table 1. Overall, these 10 experiments involved 2,433 participants listing a total of 13,101 thoughts. The upcoming methods section has additional details about experimental methods and procedures for these experiments. The Supplemental Materials also present the results of an additional experiment (Experiment 3), in which each participant was given all eight questions from Experiments 1a–1h. This experiment was conducted to test whether the results of Experiments 1a–1h persist on the individual level. We do not discuss this experiment and its results in detail in the main text.

In addition to a final decision, these experiments give us, in each trial, a sequence of unique natural language thoughts and quantified supports for each listed thought. We can write a thought sequence as $[q_1, q_2, \ldots, q_T]$, where q_t is the tth listed thought and T is the total number of thoughts listed by the participant. The supports for these thoughts can similarly be represented in a T length vector $[s_1, s_2, \ldots, s_T]$. We have s_t in the set $\{-3, -2, -1, 0, 1, 2, 3\}$, with larger s_t indicating that the tth thought has higher support for the first response option.

Thought Clusters

Memory processes influence decisions by determining the content of decision-relevant thoughts and the sequence in which they are listed. Thus, in order to model memory search in memory-based decision making, we need to be able to apply memory models to predict the thought sequences, $[q_1, q_2, \ldots, q_T]$, generated by participants in our experiments. However, unlike recall sequences in established memory paradigms (e.g., list recall of words or semantic memory search of concepts in a given category), the set of listed thoughts in our experiment is large and unconstrained, with each thought being a unique draw from a seemingly infinite thought space. In order to build tractable memory models of thought sampling, we (a) use pretrained computational language models that take string representations of sentences and compute fixed-length real-valued vectors, and (b) apply k-means clustering to these vector representations to group thoughts into a smaller, finite number of discrete clusters. As a simplifying assumption, we treat the memory process as operations over the discrete thought clusters, and use the thought cluster recall sequences to fit different memory models to our data. We make this simplifying assumption because it is hard to describe and fit stochastic processes over a high-dimensional continuous space.[3] In the discussion, we return to this issue and related issues concerning the representation of our thoughts, and the processes operating on these representations.

To obtain vector representations of our thoughts, our primary analysis relies on the transformer version of Google's Universal Sentence Encoder (USE; Cer et al., 2018), which is a deep neural

---

[3] It may theoretically be possible to model memory processes as operating over the finite set of distinct thoughts listed in our experiments. Practically, however, this is intractable, and also has various additional limitations. The vast majority of listed thoughts are unique, and therefore we cannot estimate base rates for each individual thought given the frequency of the thought in our data set. It is also worth noting that our clustering solutions help us pool semantically identical thoughts that nonetheless have slightly different wording (e.g., "nuclear waste is hard to dispose" and "it is hard to dispose of nuclear waste"). It will be hard to avoid this kind of issue if we consider thoughts individually without clusters. The use of clustering also allows us to model the behavior of new participants whose listed thoughts are not in the set of thoughts listed by prior participants (as is the case for Experiments 2 and 3). These new thoughts can essentially be categorized into the preexisting clusters, and base rates and other parameters for these preexisting clusters can be used to predict new participants' responses.

Table 1

Overview of Data

	Decision prompt	N	Ave. Thoughts (SD)	Ave. Words (SD)	Ave. Support (SD)	Choice Prob.
1a	Q: Should the American drinking age be lowered to 18? R: Yes or no	241	4.74 (2.2)	14.8 (9.1)	0.27 (2.4)	0.490
1b	Q: Imagine you received a $1,000 gift certificate that could be used for either household electronic purchases or a vacation. Which would you prefer? R: Electronics or vacation	229	5.16 (2.3)	12.3 (7.0)	0.13 (2.4)	0.511
1c	Q: Can money buy happiness? R: Yes or no	228	5.48 (3.2)	13.5 (10)	0.62 (2.4)	0.535
1d	Q: Is nuclear power safe? R: Yes or no	238	4.53 (2.2)	12.1 (7.8)	0.06 (2.3)	0.521
1e	Q: Would you prefer to eat a salad or a burger for dinner? R: Salad or burger	234	5.00 (2.2)	10.1 (5.7)	−0.1 (2.4)	0.325
1f	Q: Would you copy a piece of software without paying for it? R: Yes or no	240	4.71 (2.2)	12.5 (7.0)	0.03 (2.3)	0.558
1g	Q: Would you prefer to invest your retirement savings primarily in stocks (high risk, high return) or in bonds (low risk, low return)? R: Stocks or bonds	238	4.63 (2.0)	13.1 (7.3)	−0.2 (2.2)	0.391
1h	Q: Does technology make us more alone? R: Yes or no	234	5.35 (3.5)	14.8 (10.5)	0.42 (2.4)	0.406
2a	Q: Can money buy happiness? R: Yes or no Combined	215	7.03 (4.2)	14.9 (11)	0.58 (2.3)	0.516
	Condition 1 (Yes prime)	117	6.90 (3.9)	15.8 (13)	1.27 (2.1)	0.607
	Condition 2 (No prime)	98	7.19 (4.5)	13.8 (9.6)	−0.2 (2.4)	0.408
2b	Q: Is nuclear power safe? R: Yes or no Combined	216	6.28 (3.0)	13.6 (10)	−0.1 (2.3)	0.565
	Condition 1 (Yes prime)	112	5.84 (2.4)	13.7 (10)	0.72 (2.1)	0.723
	Condition 2 (No prime)	104	6.75 (3.6)	13.6 (9.9)	−0.9 (2.1)	0.394

Note. This table summarizes decision prompts (questions, Q, and response options, R), number of analyzable participants (N), average number of listed thoughts per participant (Ave. Thoughts), average number of words per listed thought (Ave. Words), and choice probabilities for the first response option (Choice Prob.), for our 10 experiments. In Experiments 2a and 2b participants were randomly assigned to one of two conditions.

network that learns to produce sentence vector representations so as to perform multiple natural language processing tasks, from predicting arbitrary running text to question answering. We use this model instead of alternatives, like the Deep Averaging Network version of USE or recent models like ELMO (Peters et al., 2018) and BERT (Devlin et al., 2018), as the transformer version of USE achieved state-of-the-art performance on sentence similarity benchmarks without additional fine-tuning of the USE network (at the time of writing the manuscript). We optimize for performance on sentence similarity so that clustering groups sentences that a human would find similar in meaning.

Models like USE that deliver sentence vectors can be seen as an evolution of earlier models that compute vectors for words, like LSA (Landauer & Dumais, 1997), BEAGLE (Jones & Mewhort, 2007), Word2Vec (Mikolov et al., 2013), or GloVe (Pennington et al., 2014), based on the patterns of co-occurrence among words in large corpora. The advantage of using models like USE to obtain sentence vectors, over simply, say, averaging word vectors in a sentence, is that models like USE will take into account the order and identity of all words within a sentence when computing a vector. Obviously, the order of words within a sentence, and not just their identity, is a critical component of the meaning of the sentence (cf. "Nuclear power is better than fossil fuel" vs. "Fossil fuel is better than nuclear power").

Note that our use of thought vectors and, more generally, spatial representations of semantic content, is purely practical. Such vectors do not (yet) offer a complete cognitive theory of language learning, linguistic compositionality, or semantic representation. Rather, these vectors are merely methodological tools for quantifying the rich and unconstrained set of thoughts that are sampled from memory in naturalistic decision tasks.

Our use of such models is also partially inspired by how semantic space models have been used in past research in memory and decision making. Besides modeling word similarity and relatedness judgments, word vectors can model semantic memory search, list recall, free association, and more (for review see Bhatia et al., 2019; Jones et al., 2015; Lenci, 2018; Mandera et al., 2017). Though the details of these tasks vary, generally the distance between two-word vectors in the semantic space predicts the probability that one word will cue the recall of a second. In much the same way, we will use the distance between clusters of thoughts derived from sentence vectors in the thought space to model the probability of one cluster cuing the recall of another during thought listing.

Once we have obtained vectors for each thought via the USE, we use thought support ratings—the ratings participants give to indicate which answer (if any) each thought supported—to separate thoughts into three sets, one for the first response option, one for the second, and one for neutral thoughts that support neither response. We then apply k-means clustering to the thoughts belonging to each set so that we obtain separate clustering solutions for first response thoughts, second response thoughts, and neutral thoughts. With this methodology, we are able to transform our thought listing data $[q_1, q_2, \ldots, q_T]$ into a corresponding list of thought clusters $[r_1, r_2, \ldots, r_T]$. Unlike q_t, r_t is not a natural language sentence but rather a categorical variable that indicates the specific cluster that the tth thought is drawn from. In our analysis we consider $k = 2$, $k = 3$, and $k = 4$ k-means clustering solutions for thoughts in the first and second response sets (we do not cluster thoughts in the neutral set because relatively few sentences are rated as being neutral).

This yields a total of 5, 7, or 9 thought clusters over which we assume memory processes can operate. In the main text we will present analysis with the $k = 3$ cluster solution (corresponding to seven thought clusters), and in the Supplemental Materials we repeat our analysis with the $k = 2$ and $k = 4$ cluster solutions.

Figure 2 contains word clouds of the words that are more frequent in each cluster relative to all other clusters for the experiment asking participants "Is nuclear power safe?" (Experiment 1d), for the $k = 3$ clustering solution. From these word clouds, one can often infer the decision-relevant propositions or ideas conveyed by a cluster of thoughts. For example, our clustering solution reveals a no-supporting cluster related to radiation and nuclear disasters and another one related to arguably safer clean energy alternatives like solar and wind power. In contrast, one yes-supporting cluster seems to convey the notion that nuclear power stations take lots of safety measures and other precautions, while another seems focused on the carbon emission advantage of nuclear power over fossil fuels.
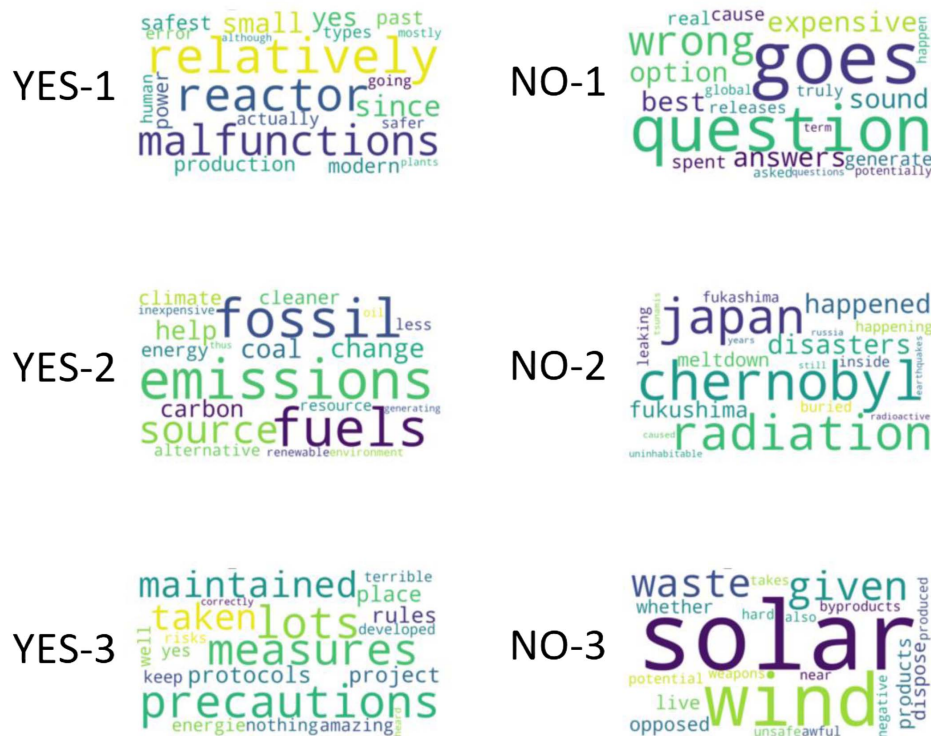
Context

By clustering thoughts into mutually exclusive and exhaustive groups we transform the unconstrained thought listing data obtained from the query theory experimental paradigm into a form that can be described using standard memory models (which are typically applied to the recall of a small number of items). Note that there are important quantitative relationships between thought clusters, such as semantic congruence (the degree to which two thought clusters have similar semantic content) and decision congruence (whether or not two thought clusters support the same response). Memory models are able to describe the effect of semantic similarity, temporal contiguity, and source congruence on the dynamics of word recall, and can, for this reason, also be used to characterize the effect of semantic congruence and decision congruence on thought cluster sampling. Likewise, decision models are commonly used to describe discrete choice data, which is identical to the type of binary response data we obtained in our experiments. Additionally, our experiments also elicit lists of support ratings for the thought sequences generated by participants. Support lists quantify the sequence of information that is integrated by the decision maker, and can be used to flexibly model the dynamic properties of the decision process (as in Busemeyer & Rapoport, 1988; Lee & Cummins, 2004; Lee et al., 2019).

Now, different memory and decision models can be built and applied separately to thought listing or decision data, but how can we combine these models within a single cohesive framework that simultaneously describes memory sampling and decision making? Our solution to this problem relies on the idea of *context*, taken from the CMR family of memory models (Howard & Kahana, 2002; Polyn et al., 2009). Particularly, we assume that all decision-relevant information known by the decision maker at time t is stored in a dynamic context variable C_t. C_t evolves with each memory sample, and determines both memory sampling and decision making. Note that we are using the term "time" to refer to the discrete memory sampling steps that lead up to the decision, and not the continuous response time data that memory and decision models are sometimes fit on. Due to the complexity inherent in the thought sampling and thought listing process, we will not be modeling continuous response time data in this article.

Figure 2

Word Clouds of the Words that Are More Frequent in Each Cluster Relative to All Other Clusters for the First Experiment ("Is Nuclear Power Safe?"), for the k = 3 Clustering Solution

Note. Here we display only clusters for yes-supporting thoughts (left panels) or no-supporting thoughts (right panels). See the online article for the color version of this figure.

The crucial feature of our data is that we are able to observe the set of sampled thoughts and supports, and are thus able to specify C_t for any point in time during deliberation. As C_t is the cognitive intermediary between memory sampling and decision making, we can use C_t to derive the probabilities of the observed memory sample or decision at t, using existing memory or decision models. Additionally, as the memory samples and decisions do not interact through any other cognitive variables, the probabilities of the observed memory sample and the observed decision at t are conditionally independent given C_t. Conditional independence allows us to separate the probabilities predicted by the memory and decision models, and thus fit memory and decision processes to our data individually, facilitating a tractable modeling solution.

Context is not only a practical assumption. It also offers a theoretically principled approach to describing the effect of situational variables, such as primes, cues, emotions, choice sets, and reference points. Such variables have been shown to bias judgments and decisions by altering memory (see e.g., Weber & Johnson, 2009 for a review), and are easily understood through the lens of memory context. In Experiment 2 of this article, we explicitly prime participants with thoughts supporting or opposing a decision (as in Johnson et al., 2007 and Weber et al., 2007). That is, we ask subjects to first list a thought supporting one choice option, and then allow them to list thoughts and choose of their own volition.

By assuming that these thought primes influence memory context at the start of the decision, we can quantitatively model their effect on thought recall, and subsequently characterize their downstream influence on the decision.

Note again that although our implementation of this idea involves the CMR model, and the use of the term context, a nearly identical implementation could involve the SAM model (Raaijmakers & Shiffrin, 1981), in which memory and decision-relevant information would be stored in a short-term memory store. Closely related proposals have also been made within the ACT-R framework (Anderson et al., 2004; Dimov et al., 2020; Fechner et al., 2016; Link et al., 2016; Link & Marewski, 2015; Marewski & Mehlhorn, 2011; Marewski & Schooler, 2011; Schooler & Hertwig, 2005). In such models, memory and decision-relevant information are stored in the buffers of different modules (e.g., declarative memory or imaginal), and such an architecture could also be used to mimic our framework. Context, more generally, can be seen as a type of short-term storage or working memory buffer, widely used in models of memory and related cognitive processes (see, Oberauer, 2009 for a discussion). To keep things simple, and to ensure maximum compatibility between our approach and leading models of memory dynamics, we use the terminology and modeling architecture of the CMR rather than the more elaborate architecture of approaches such as ACT-R.

Illustration

Our framework uses memory models to study how participants generate thought cluster sequences $[r_1, r_2, \ldots, r_T]$, and how this depends on previously sampled thought clusters and the semantic and decision-congruence relationships between them. It also uses decision models to predict the final responses, and how they depend on the sequence of supports, $[s_1, s_2, \ldots, s_T]$, emitted by the sampled thoughts.

Figure 3 provides an illustration of our proposed framework, represented as a neural network. There are three layers: The first layer ("Memory") contains the set of K thought clusters stored in memory, that could be sampled during deliberation. The second layer ("Decision") contains the set of three decision possibilities available at each point in time: choosing the first response option, choosing the second response option, or continuing deliberation by sampling another item from memory. The third layer ("Context") contains all sampling-relevant and decision-relevant information stored in context. This information is divided into two separate parts, with the first representing previously sampled thought clusters, and the second representing the aggregated supports (i.e., preferences or beliefs) generated through these samples. In the illustration shown in Figure 3, decision makers are asked "is nuclear power safe?." We assume that they have a total of $K = 5$ thought clusters and that the two response options are "yes" and "no."

The decision begins, at $t = 1$, with some context C_1 and the (implicit) decision to sample memory based on this context. This results in the second thought cluster being sampled, which produces a thought sample with support $s_1 = +2$ (indicating moderate support in favor of the first response, "yes"). This process generates a context representation C_2 that encodes the thought cluster and the degree of support offered by the thought sample. At $t = 2$, the decision maker uses C_2 to make a decision: Choose one of the two response options or keep deliberating. In our example, the decision maker decides to continue deliberating. Thus, at $t = 2$, C_2 determines the next memory sample, Cluster 3, which offers support of $s_2 = 0$ (indicating neutral support for both responses). Once again, context updates, forming a new representation, C_3. C_3 encodes a (potentially decayed, as they are in Figure 3) representation of the two previously sampled thought clusters and supports of the sampled thoughts. The updated context determines the decision, at $t = 3$, to continue deliberating. C_3 subsequently generates the memory sample at $t = 3$. In our illustration, the first thought cluster is sampled, which produces a thought sample with support $s_3 = +3$. The context is updated to C_4 to encode the cluster index and support ratings of the new thought sample. The aggregated supports then cause the decision maker to choose the first response option, "yes," at $t = 4$, ending deliberation.

Preview of Models and Data

Memory models specify the memory-relevant information stored in context, and how this information determines the items that are retrieved from memory. In the settings modeled in this article, this information takes the form of previously sampled thought clusters. Below we will allow for sampled thought clusters to influence their own sampling probabilities, as well as the sampling probabilities of other semantic and decision congruent thought clusters (thought clusters will also be allowed to have different baseline activation strengths, which influence recall independently of context). We will

also allow for different assumptions regarding the decay of information in context, including full decay, no decay, and partial decay. These assumptions determine how thought clusters sampled earlier on in the decision influence memory sampling later on in the decision.

Decision models specify the decision-relevant information that is stored in context and how this information determines responses. In the settings modeled in this article, this information takes the form of aggregated supports. Below, we will allow for the representation of separate aggregated supports for each of the two response options, or for the representation of a single relative aggregated support favoring the first option over the second. We will also allow for the aggregation of continuous supports or discrete supports (which indicate whether or not a response is supported by a thought, but not the magnitude of this support). Decisions will also be allowed to depend on internal decision thresholds or exogenous time limits. Finally, as with the memory models, we will allow for different assumptions regarding the decay of support information in context, including full decay, no decay, and partial decay.

Together the above assumptions will help us explain patterns in our experimental data, such as the finding that retrieved thought clusters cue the retrieval of other semantically related thought clusters as well as other thought clusters that support the same response (findings that, to preview, are visualized in Figure 6). They will also help us model how and when aggregated support is used to make a decision (Figures 7 and 8; to preview, time limit models, e.g., cannot account for the clear observed dependence of stopping on aggregate decision support). With the correct combination of modeling assumptions (revealed through fits of our memory and decision models to participant data) we will be able to quantitatively predict response probabilities, deliberation length (i.e., number of thoughts sampled), thought content (i.e., the specific thought clusters sampled), and other key memory and decision variables. Crucially, best-fit models would also allow for the prediction of priming effects—that is, how the manipulation of starting thoughts alters the dynamics of retrieval and the eventual decision.

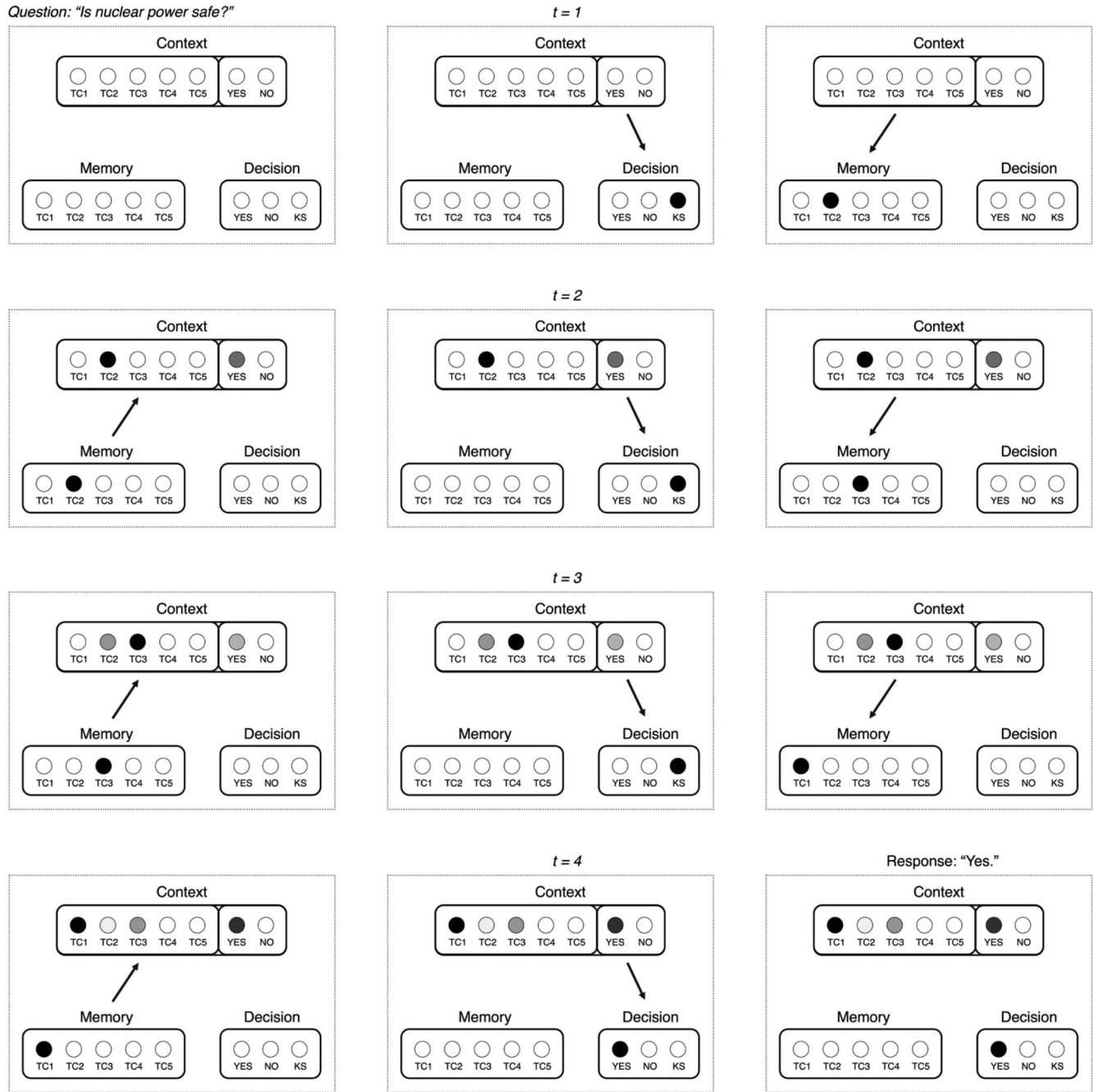## Model Specification

General Structure

At each point in time t the decision maker has the choice of selecting one of the two response options and terminating deliberation, or of sampling memory and continuing deliberation. These decisions are made based on context C_t and we write the probability that the decision made at time t is x_t, given context C_t, as $\Pr^D[x_t \mid C_t]$. If memory is sampled, the decision maker can sample one of K thought clusters. This again depends on context, and we write the probability that the thought cluster sampled at t is r_t, given context C_t, as $\Pr^M[x_t \mid C_t]$. Note that $\Pr^M[x_t \mid C_t]$ is itself conditional on the decision maker deciding to sample memory at t, though we omit this conditional from our notation for expositional convenience.

Now, for each subject we observe a sequence of memory samples $[r_1, r_2, \ldots, r_T]$. We also observe a final decision. The final decision can equivalently be seen as a sequence of T decisions to keep sampling memory, each of which yields a thought cluster sample, followed by a decision to end deliberation at $T + 1$. This sequence can be written as $[x_1, x_2, \ldots, x_{T+1}]$ where x_t is the decision to keep sampling if $t < T + 1$, and x_{T+1} is the response option eventually chosen by the participant. With this

Figure 3

An Illustration of a Hypothetical Decision in which the Individual is Asked "Is Nuclear Power Safe?"

Note. There are five thought clusters that can be sampled sequentially from memory in this example. Thought Clusters 1–2 support the choice of "yes," thought Cluster 3 is neutral, and thought Clusters 4–5 support the choice of "no." Each row corresponds to one time point in the memory-based decision. At any time point t beyond $t = 1$, the context (C_t) is updated to reflect how often (and how recent) different thought clusters have been sampled (m_t), as well as (time-weighted) aggregated supports in favor of "yes" (u_t^1) and "no" (u_t^2). The updated context drives decision making (middle panel of each row; see the section *Decision Mechanisms* for choice probability computation) and memory sampling (right panel of each row; see the section *Memory Mechanisms* and Equations 2–3 for sampling probability computation). Here the hypothetical decision takes four time periods and ends with the selection of the "yes" response. KS = Keep sampling; TC = Thought cluster.

structure, $\prod_{t=1}^{T}(\mathrm{Pr}^D[x_t \mid C_t] \cdot \mathrm{Pr}^M[r_t \mid C_t])$ captures the probability of observing the T thought clusters that are sampled during deliberation, each of which follows a decision to keep sampling memory. Likewise, $\mathrm{Pr}^D[x_{T+1} \mid C_{T+1}]$ captures the probability of ending deliberation and selecting the chosen response after the Tth thought cluster sample. Thus together, we can write the probability of our observation as

$$\prod_{t=1}^{T}(\mathrm{Pr}^D[x_t \mid C_t] \cdot \mathrm{Pr}^M[r_t \mid C_t]) \cdot \mathrm{Pr}^D[x_{T+1} \mid C_{T+1}].$$

This, in turn, can be rewritten as:

$$\prod_{t=1}^{T} \mathrm{Pr}^M[r_t \mid C_t] \cdot \prod_{t=1}^{T+1} \mathrm{Pr}^D[x_t \mid C_t]. \tag{1}$$

Equation 1 describes the probability of our observation as a function of separate memory and decision components, which are predicted by separate memory and decision models. Specifically, $\prod_{t=1}^{T} \mathrm{Pr}^M[r_t \mid C_t]$ is the probability of observing $[r_1, r_2, \ldots, r_T]$, predicted by a given memory model. These predictions depend on how the memory model specifies the probability of each thought cluster being sampled, and how context (consisting of the set of previously sampled thoughts) influences these probabilities. Good memory models should give high values of $\prod_{t=1}^{T} \mathrm{Pr}^M[r_t \mid C_t]$. Likewise, $\prod_{t=1}^{T+1} \mathrm{Pr}^D[x_t \mid C_t]$ is the probability of observing a sequence of T decisions to keep sampling memory followed by a decision to end deliberation at $T + 1$ and select the response option eventually chosen by the participant, predicted by a given decision model. These predictions depend on how the decision model specifies the probability of each choice, and how context (consisting of the set of previously sampled supports) influences these probabilities. Good decision models should give high values of $\prod_{t=1}^{T+1} \mathrm{Pr}^D[x_t \mid C_t]$.

Note that the above model allows for the possibility that participants do not list any thoughts at all (i.e., they choose to end sampling at $t = 1$). We did not allow participants to do this, and thus when we fit the models, we assume that x_1 is the decision to continue sampling and that $\mathrm{Pr}^D[x_1 \mid C_1] = 1$.

Overall, the predictions of the memory and decision models depend only on context. Thus, memory and decision making are conditionally independent given context. In other words, once we know about the value of context at a given time $t(C_t)$, knowledge about recall at that time (r_t) provides no further information about the decision at that time (x_t) and vice versa. For this reason, memory and decision models can be fit to our memory and decision data separately, as long as they appropriately specify how context influences memory or decision making. As these assumptions pertain to the elements of the sampled thoughts and support that are retained in context, our empirical paradigm allows us to observe the context specified by a given decision model. Thus, despite C_t being a random variable (that depends on the sampled thoughts and supports prior to t), we ultimately know the realization of C_t that guides recall, r_t, and decision, x_t, at time t. This is why we omit $\mathrm{Pr}[C_t]$ from the above equations.

Memory Mechanisms

We assume that context encodes a representation of previously sampled thought clusters, and that this representation determines subsequent memory samples. The most general way of quantifying this representation at time t is with a $K \times (t - 1)$ matrix M_t of indicator variables, with K specifying the total number of thought clusters, and cell (k, l) of M_t specifying whether or not the kth thought cluster was sampled at time l (1 if it was, 0 otherwise). As time progresses, columns can be added to M_t, with values based on the sampled thought clusters. As only one thought cluster can be sampled at any given time, the values in each column must sum to 1.

Although such a general representation may be useful for implementing complex memory search rules, we assume, for simplicity, that the M_t can be replaced with a K-length vector $m_t = M_t \cdot \alpha_t$. Here α_t is a $t - 1$ length vector $[(1 - \delta_M)^{t-2}, (1 - \delta_M)^{t-3}, (1 - \delta_M)^{t-4}, \ldots, (1 - \delta_M)^0]$ of decay weights, with δ_M in range $[0, 1]$. m_t can be seen as a vector of activations reflecting how recently a particular thought cluster was sampled. This vector guides thought cluster sampling at time t by sending feedback into the thought layer, which determines thought cluster activation based on decision and semantic congruence between pairs of thoughts. Our use of this specification for context implies that the effect of a sampled thought on subsequent memory is assumed to increase additively with multiple samples of the same thought cluster and weaken with time at an exponential rate.

Decision congruence captures whether or not a given pair of thought clusters support the same response option. For a pair of thought clusters, j and k, we write decision congruence as $D(j, k)$, with $D(j, k) = 1$ indicating that the two clusters support the same response, or if they are both neutral; otherwise $D(j, k) = 0$. The set of pairwise decision-congruence measures between the K thought clusters are represented in the $K \times K$ matrix D. Semantic congruence is the degree to which two thoughts clusters have similar semantic content. For a pair of thought clusters, j and k, we write semantic distance as $S(j, k) = \|c_j - c_k\|$, where c_j and c_k are the vector centroids of clusters j and k. $S(j, k)$ is the Euclidean distance of the centers of clusters j and k and thus smaller values of $S(j, k)$ indicate higher semantic congruence between j and k. The set of pairwise semantic congruence measures are represented in the $K \times K$ matrix S. D and S are deterministically specified by our data and by the clustering solution on our semantic space. We assume that the context vector m_t influences thought cluster activation in the memory layer based on a linear combination of D and S. We write the K-length vector of thought cluster activation in memory as:

$$A_t = \gamma + \omega_R \cdot m_t + \omega_S \cdot S \cdot m_t + \omega_D \cdot D \cdot m_t. \tag{2}$$

Here γ is a vector of baseline activation strengths, which reflects a prior distribution over the clusters (and does not update or decay during deliberation). ω_S and ω_D are scalar weights on semantic and decision congruence, respectively. If ω_S is negative and ω_D is positive, then decision makers are likely to sample thought clusters that are semantically related or support the same responses as the thought clusters sampled previously (note that S captures semantic *distances* between thought clusters). ω_R specifies an additional weight that can be directly placed on the activations captured by m_t. Intuitively ω_R indicates the degree to which clusters stored in context are likely to be resampled (controlling for the high degree of semantic and decision congruence that a cluster has with itself). If $\omega_R < 0$, then cluster resampling is less likely, reflecting a type of recall inhibition or repetition suppression. Although we do observe

recall inhibition and repetition suppression in simple memory (and attention) tasks, our more complex tasks may also reveal $\omega_R > 0$, reflecting a tendency to fixate on a thought once it has been recalled. Thus, we use a more general term, cluster resampling, to refer to ω_R.

Finally, the probability of sampling a given thought cluster is obtained by passing A_t through a softmax function, which is a link function used in multinomial logistic regressions and neural networks. Thus, if the jth thought cluster is sampled at t (i.e., $r_t = j$), we obtain:

$$\Pr^M[r_t \mid C_t] = \frac{e^{A_t(j)}}{\sum_{k=1}^{K} e^{A_t(k)}}. \tag{3}$$

By using this formula to calculate $\prod_{t=1}^{T} \Pr^M[r_t \mid C_t]$, which is the memory component of Equation 1, we can analytically derive the probability of observing a given sequence of thought cluster samples, $[r_1, r_2, \ldots, r_T]$. This probability depends only on the memory decay weight parameter, δ_M, the semantic and decision-congruence parameters, ω_S and ω_D, the cluster resampling parameter, ω_R, and the K baseline cluster sampling parameters $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_K]$.

Decision Mechanisms

We assume that context encodes a representation of previously sampled thought cluster supports, and that this representation determines the decision made at a given point in time. The most general way of quantifying this representation at time t is with a $t - 1$ length row vector of support U_t, whose lth entry specifies the support emitted by the thought cluster sampled at time l. These supports are in the set $\{-3, -2, -1, 0, 1, 2, 3\}$, with positive values indicating support for the first response. As time progresses, cells are added to U_t, with values based on the sampled thought clusters.

Although such a general representation may be useful for implementing complex decision rules, we consider a number of different ways of simplifying the information stored in U_t. One way is to assume that decision makers do not encode continuous supports, but rather encode discrete supports $V_t = \text{sign}(U_t)$. This corresponds to a heuristic decision strategy that ignores the magnitudes of supports offered by each sampled thought.

Another way of simplifying contextual representations involves aggregating the components of U_t or V_t into a single measure of the support for a given response option. Particularly, we can assume that decisions are made based on a scalar $u_t = U_t \cdot \beta_t$ or a scalar $v_t = V_t \cdot \beta_t$. Here β_t is a $t - 1$ length vector $[(1 - \delta_D)^{t-2}, (1 - \delta_D)^{t-3}, (1 - \delta_D)^{t-4}, \ldots, (1 - \delta_D)^0]$ of decay weights, with δ_D in range $[0, 1]$. u_t and v_t are representations of the relative support for the first response option compared to the second response option, that additively combines the supports emitted by all sampled thoughts, with higher weights for recently sampled thoughts. u_t and v_t evolve dynamically as thought clusters are sampled. If the support of the thought cluster sampled at t is s_t then u_t and v_t can equivalently be written as $u_t = (1 - \delta_D) \cdot u_{t-1} + s_t$ and $v_t = (1 - \delta_D) \cdot v_{t-1} + \text{sign}(s_t)$.

A variant of the above assumption involves using separate representations for each of the response options that only aggregate sampled supports that are in favor of the corresponding response option. In order to implement this assumption, we first generate $U^1_t = \max(U_t, 0)$, $U^2_t = \max(-U_t, 0)$, $V^1_t = \max(V_t, 0)$, and $V^2_t = \max(-V_t, 0)$. Here each maximum operator is applied to each element of U_t, V_t, $-U_t$, or $-V_t$, and places a floor of 0 on that element's value. We then define $u^1_t = U^1_t \cdot \beta_t$, $u^2_t = U^1_t \cdot \beta_t$, $v^1_t = V^1_t \cdot \beta_t$, and $v^2_t = V^2_t \cdot \beta_t$ to specify the aggregated supports in favor of each response option. u^1_t, for example, corresponds to the (time-weighted) sum of all sampled supports that favor response 1. If the support of the thought cluster sampled at t is s_t then u^1_t can equivalently be written as $u^1_t = (1 - \delta_D) \cdot u^1_{t-1} + \max(s_t, 0)$.

Finally, we need to make assumptions regarding the rules used to make a decision at each point in time. There are two types of rules that we consider. The first uses a threshold mechanism to select one of the two response options if the aggregated supports (plus an additive error) exceed a threshold value. We write this threshold value as τ, and the normally distributed error at time t as $\varepsilon_t \sim N(0, \sigma)$. Thus, the decision maker selects the first response option at time t if $u_t + \varepsilon_t > \tau$ (or $v_t + \varepsilon_t > \tau$) and selects the second option at time t if $u_t + \varepsilon_t < -\tau$ (or $v_t + \varepsilon_t < -\tau$). If neither of these two conditions is satisfied, the decision maker decides to continue deliberation and sample another thought cluster from memory. This threshold mechanism can also be applied to u^1_t and u^2_t, with $\varepsilon_t \sim N(0, \sigma)$ specifying the additive noise applied to these two variables. Here the first response option is selected if $u^1_t + \varepsilon_t > \tau$ and $u^2_t + \varepsilon_t < \tau$, and the second response option is selected if $u^1_t + \varepsilon_t < \tau$ and $u^2_t + \varepsilon_t > \tau$. If neither of these conditions is satisfied the decision maker continues deliberation and samples another item from memory. An analogous decision rule can be used for v^1_t and v^2_t.

Another type of decision rule involves sampling for a fixed number of time steps, T, that is, sampling until a time limit T is reached. We implement this time limit assumption by assuming that there is a fixed probability of continuing deliberation, λ, after each thought sample. With this assumption, T follows a Geometric distribution. At T, the decision maker selects the first response option if $u_T + \varepsilon_T > 0$ or $v_T + \varepsilon_T > 0$, and selects the second response option if $u_T + \varepsilon_T < 0$ or $v_T + \varepsilon_T < 0$. In the case of separate support representations, the decision maker selects the first response option if both $u^1_T + \varepsilon_T > 0$ and $u^2_T + \varepsilon_T < 0$ (or both $v^1_T + \varepsilon_T > 0$ and $v^2_T + \varepsilon_T < 0$). It selects the second response option if both $u^1_T + \varepsilon_T < 0$ and $u^2_T + \varepsilon_T > 0$ (or both $v^1_T + \varepsilon_T < 0$ and $v^2_T + \varepsilon_T > 0$). If these two conditions are not satisfied (i.e., if the accumulators are both positive or both negative) then the decision maker selects each option with a 50% probability.

Note that the assumptions regarding the noise terms above resemble those used in standard accumulator models, like drift-diffusion processes or leaky accumulators. However, unlike these models, we do not aggregate the errors over time in u_t, v_t, and the other decision variables. Thus u_{t+1} does not depend on ε_t (and is instead a deterministic function of the supports sampled prior to t), and ε_t can be interpreted as the noise in retrieving aggregate supports (or equivalently, the noise in setting a decision threshold). This greatly simplifies our modeling task without restricting the explanatory scope of our model. Overall, this and the other assumptions discussed above allow us to analytically derive the probability of the decision maker choosing one of the two response options, or deciding to sample an additional thought cluster, at each point in time. For example, if we assume decision makers aggregate continuous relative supports, u_t, to a threshold τ, then decision probabilities are simply given by the probability of $u_t + \varepsilon_t$ exceeding τ or $-\tau$. Specifically, if the decision maker selects the first response option at time t, we obtain $\Pr^D[x_t \mid C_t] = 1 - F(\tau - u_t)$, where F is the cumulative distribution function for a normal distribution with standard deviation σ. In contrast, if the decision maker selects

the second response option at t, we obtain $\Pr^D[x_t \mid C_t] = F(-\tau - u_t)$, and if the decision maker selects neither response option (and instead continues sampling thoughts), we obtain $\Pr^D[x_t \mid C_t] = F(\tau - u_t) - F(-\tau - u_t)$. In the case of the accumulation of discrete relative supports to a threshold, the above equations simply replace u_t with v_t. When we have the accumulation of absolute supports, these equations are modified to calculate the probability that one accumulator surpasses the threshold while the other does not. Finally, in the case that decision makers use a time limit instead of a threshold, the decision probabilities are given by the probability that the time limit has or has not been reached by a particular point in time (which depends on the probability of continuing deliberation at each time point, λ), and the probability of the accumulators being in favor of the first or second response option when the time limit is reached (which again is given by the CDF of the normal distribution, with threshold 0).

By using the above steps to calculate each of the probabilities that enter $\prod_{t=1}^{T+1} \Pr^D[x_t \mid C_t]$, the decision component of Equation 1, we can analytically derive the probability of observing a sequence of T decisions to keep sampling memory followed by a decision to end deliberation at $T + 1$ and select the response option chosen by the participant, given the support ratings associated with each sampled thought. This probability depends only on (a) the decision decay weight parameter, δ_D, (b) whether or not the decision maker accumulates continuous supports or discrete supports, (c) whether or not the decision maker uses a relative accumulator or a discrete accumulator, (d) the threshold (τ) or time limit (λ) parameters describing the decision rule, and (e) the standard deviation of the noise (σ) at each point in the decision.

Note that the reason why we allow different decay rates for memory and decision making is for computational tractability. If we constrained these decay parameters to be the same, then the conditional independence assumption would no longer hold. Of course, by separately fitting the two decay rates we are also permitting the special case in which they are identical. Our best-fit parameters (shown below), however, reveal that these decay rates are not identical, implying that our assumption is not only useful for computational tractability, but also facilitates superior model fits.

Switchboard Analysis

In the above sections, we have introduced a number of different mechanisms for modeling memory and decision processes. For example, memory context can influence thought cluster sampling based on semantic congruence or decision congruence, or both. Likewise, decision making can involve a single representation for the relative support of the two options or separate representations for the absolute support for each option, as well as decisions based on a threshold or a time limit mechanism. Both memory and decision context can be susceptible to decay. These and other mechanisms are largely separate from each other and can be individually added to or removed from the models. These mechanisms are summarized in Table 2.

Our goal in this article is to analyze memory and decision models with each different combination of underlying mechanisms. As there are a total of $2 \times 2 \times 2 \times 3 = 24$ different memory models generated through the combination of the four memory mechanisms, and a total of $2 \times 2 \times 2 \times 3 = 24$ different decision models generated through the combination of the four decision mechanisms, our article implicitly involves the analysis of $24 \times 24 = 576$ memory-based decision models. Each model features a unique set of mechanisms, and generates a unique set of predictions for our recall and decision data. By fitting these 576 models to our data we can accurately characterize the set of memory and decision mechanisms necessary to describe memory-based decision making.

The type of analysis we perform here is known as a *switchboard analysis* and was recently used by Turner et al. (2018) to analyze the

Table 2
Summary of Memory and Decision Mechanisms Used in Our Framework

Memory mechanisms	Decision mechanisms
Semantic congruence: Sampled thought clusters influence sampling probabilities based on semantic similarity ($\omega_S \neq 0$).	*Continuous support*: Accumulators aggregate continuous supports, s_t, in accumulators u_t or u^1_t and u^2_t.
No semantic congruence: Cluster sampling is independent of semantic similarity between thought clusters ($\omega_S = 0$).	*Discrete support:* Accumulators aggregate discrete supports, $\mathrm{sign}(s_t)$, in accumulators v_t or v^1_t and v^2_t.
Decision congruence: Sampled thought clusters influence sampling probabilities based on whether they support the same or different responses ($\omega_D \neq 0$).	*Relative accumulation*: Decision makers use single relative accumulators, u_t or v_t, encoding relative support for the two response options.
No decision congruence: Cluster sampling is independent of decision-congruence relations between thought clusters ($\omega_D = 0$).	*Absolute accumulation:* Decision makers use two separate accumulators, u^1_t and u^2_t or v^1_t and v^2_t, encoding absolute support for each response option.
Cluster resampling: Sampled thought clusters influence their own sampling probabilities, controlling for semantic and decision congruence ($\omega_R \neq 0$).	*Threshold decision rule*: Decisions are based on a threshold τ so that an option is chosen if accumulator plus error crosses τ or $-\tau$.
No cluster resampling: Sampled thought clusters do not influence their own sampling probabilities, controlling for semantic, and decision congruence ($\omega_R = 0$).	*Time limit decision rule*: Decisions are based on time limit T (based on probability of continuing deliberation, λ) so that the option with the larger accumulator plus error at T is chosen.
Full context decay: $\delta_M = 1$, so that only the previously sampled thought cluster influences subsequent cluster sampling probabilities.	*Full context decay:* $\delta_D = 1$, so that only the previously sampled support is encoded by the accumulators (e.g., $u_t = s_t$).
No context decay: $\delta_M = 0$, so that all sampled thought clusters influence subsequent cluster sampling probabilities.	*No context decay:* $\delta_D = 0$, so that all sampled supports are encoded and given the same weight by the accumulators (e.g., $u_t = u_{t-1} + s_t$).
Partial context decay: $0 < \delta_M < 1$, so that recently sampled thought clusters have a larger effect on subsequent cluster sampling probabilities.	*Partial context decay:* $0 < \delta_D < 1$, so that all sampled supports are encoded but recently sampled supports are given more weight by the accumulators, for example, $u_t = (1 - \delta_D) \cdot u_{t-1} + s_t$.

set of decision mechanisms responsible for multiattribute context effects. A switchboard analysis allows researchers to investigate not only existing models but also hybrid models that can be generated by combining various model mechanisms, and thus provides a more nuanced understanding of the psychological properties at play in the task at hand. In our case, the switchboard analysis also allows us to pair different types of memory models with different types of decision models, and subsequently characterize the very large set of distinct cognitive models that could be responsible for memory-based decision making.

Relationship With Existing Models

The model mechanisms summarized in Table 2 are inspired by existing theories, and appropriate combinations of model mechanisms are capable of perfectly instantiating or closely approximating these theories. As discussed earlier, our assumption of context as the key cognitive intermediary guiding memory and decision making is taken from established memory models, such as CMR, SAM, and ACT-R (Anderson et al., 2004; Polyn et al., 2009; Raaijmakers & Shiffrin, 1981). Subsequently allowing for semantic congruence ($\omega_S < 0$) and partial context decay ($0 < \delta_M < 1$) allows our framework to mimic the semantic contiguity properties of these models. Such models also often allow for some decay in context (CMR), baseline or base-level activation of items in memory (SAM and ACT-R), as well as the possibility of noisy retrieval (CMR, SAM, and ACT-R), as is the case above. Conversely, assuming full context decay ($\delta_M = 1$) restricts the memory process so that only thought clusters sampled at $t - 1$ influence cluster sampling probabilities at t. This is capable of mimicking the behavior generated by a random walk in a semantic network (e.g., Abbott et al., 2015).[4] Decision congruence is a key component of query theory, and thus $\omega_D > 0$ allows our framework to capture the essential memory mechanisms underlying this theory (Johnson et al., 2007; Weber et al., 2007)[5]. Of course, turning all the mechanisms off and setting ($\omega_S = \omega_D = \omega_R = \delta_M = 0$) generates independent sampling over time, which is the implicit assumption in many memory-based decision making models such as the associative accumulation model (Bhatia, 2013).

On the decision side, we assume that supports are aggregated in either relative (u_t or v_t) or absolute ($u^1{}_t$ and $u^2{}_t$ or $v^1{}_t$ and $v^2{}_t$) accumulators. When combined with the threshold decision rule and no context decay ($\delta_D = 0$), the relative accumulation assumption yields a standard discrete relative evidence accumulator (one of the baseline models in Busemeyer & Rapoport, 1988; Gluth et al., 2012; Lee & Cummins, 2004) and with some modifications, a discrete version of the popular drift-diffusion model (Ratcliff & Rouder, 1998). In the case in which we use discrete supports, v_t, our framework mimics the core assumptions of the dynamic decision-by-sampling model (Noguchi & Stewart, 2018; also see Stewart et al., 2006). Conversely, threshold decision making and partial context decay ($0 < \delta_D < 1$), combined with the absolute accumulation assumption, generate a discrete leaky accumulator (such as that used in Usher & McClelland, 2001, 2004; also see Bhatia, 2013 and Roe et al., 2001). Note that many practical tests of these models have used an exogenous time limit mechanism (instead of threshold decision making). These tests can be captured by replacing the threshold decision rule with the time limit decision rule in our framework.

If we assume full context decay ($\delta_D = 1$) our models no longer accumulate supports. Rather, at each point in time, decisions are made based only on the support sampled in the previous time period. This is a feature of a number of heuristic models of decision making. Thus, for example, we obtain decision processes similar to the lexicographic heuristic (Fishburn, 1967; Gigerenzer & Goldstein, 1996) with discrete support, relative accumulation, and a very small threshold, as long as we have full context decay. Continuous support with a slightly larger threshold in the presence of full context decay gives us decisions resembling the lexicographic semiorder heuristic (Tversky, 1969; see Lee & Cummins, 2004 for an early application of this idea). Of course, both the lexicographic and lexicographic semiorder heuristics assume that attributes are sampled in a particular order, and thus a perfect implementation of these heuristics within our framework would require appropriate restrictions to the thought-generating process as well (this could be accomplished with appropriate baseline sampling biases and strong negative cluster resampling effects, combined with no context decay and no semantic/decision-congruence effects).

Finally, permitting a long enough exogenous time limit or a large enough threshold, along with no context decay, allows us to model the weighed additive decision rule, in which the decision maker has enough time to sample every thought (Keeney & Raiffa, 1993). Here, using discrete supports instead of continuous support allow us to capture simplifications of this rule, such as the tallying heuristic (which only considers which option is better or worse on the sampled attribute, without taking into account the strength of each piece of evidence; Gigerenzer & Goldstein, 1996; Russo & Dosher, 1983).

Of course, we can also combine various memory and decision mechanisms to generate combinations of existing memory and decision models for describing memory-based decision making. For example, allowing for semantic congruence, threshold decision making, absolute accumulation, as well as partial memory and decision context decay, give us a CMR-style memory model whose output is integrated by a leaky accumulator. In contrast, allowing for decision congruence with a threshold decision rule, no context decay, and the relative accumulation of discrete supports, give us a query theory memory model whose output is integrated by a decision-by-sampling style model. Similarly, allowing for only semantic congruence in memory, relative accumulation of discrete supports to a threshold in decision making, and full context decay in both memory and decision making, give us a Markov random walk through a semantic network whose output is integrated by lexicographic semiorder heuristic. Overall, our framework instantiates these and 573 other combinations of memory and decision models, and is capable of testing which of these combinations best describes memory-based decision making through switchboard analysis.

---

[4] In general, our specific specification of working memory capacity limits, in terms of context decay, stems from our application of the CMR memory model to memory-based decision making. Decay is also part of leaky integration models, which we consider on the decision side of our analysis. Thus both the memory and decision components of our model share the same capacity limits, which we believe offers some amount of theoretical cohesion. Of course alternate capacity limits, such as limitations to the number of items in working memory as in the ACT-R or SAM frameworks, could be feasible. We revisit this possibility in the discussion section of this article.

[5] We can also generate, as a special case, a memory model that samples only based on decision congruence and does not consider the semantic content of thoughts, by setting $\omega_S = 0$. Intuitively, this would resemble a model with only three thought clusters: one for positively supporting thoughts, one for negatively supporting thoughts, and one for neutral thoughts.

Our framework is related to theories that integrate retrieval dynamics with response dynamics in domains such as categorization (Cox & Shiffrin, 2017; Hockley & Murdock, 1987; Nosofsky & Palmeri, 1997; Ratcliff, 1978), as well as ACT-R-based models of memory and decision making (Dimov et al., 2020; Fechner et al., 2016; Link et al., 2016; Link & Marewski, 2015; Marewski & Mehlhorn, 2011; Marewski & Schooler, 2011; Schooler & Hertwig, 2005). Theoretically, many of the high-level architectural details are shared with these other models, though of course we do not implement evaluation mechanisms that are more relevant to categorization than decision making (e.g., exemplar-based rules in Nosofsky & Palmeri, 1997) or mechanisms that would be necessary to relate our processes to other cognitive systems (e.g., production rules that use visual data to update working memory, or use working memory to output motor responses, as in Dimov et al., 2020). Empirically, our evaluation of decision models is related to prior work that systematically varies the time course of evidence (e.g., Trueblood et al., 2021; Tsetsos et al., 2012; Zhou et al., 2009), though we consider a somewhat simpler set of models and additionally rely on a memory model to control the time course of evidence. Unlike all prior work, we attempt to characterize the retrieval and evaluation of complex thoughts in high-level judgments and decisions, and are able to observe (and fit) the sequence of retrieved thoughts using participant data observed with the query theory experimental paradigm.

Method

Participants

We recruited a total of 2,433 participants (Age: $M = 32$ years $SD = 12$, range = 18–76; 47% female) from Prolific Academic for Experiments 1a–1h and Experiment 2. We limited our data collection to participants who were from the U.S. and fluent in English, and had an approval rate above 80%. Participants were only allowed to participate once, and earned a base payment of $1.35. 50% of participants were given a bonus worth 50% of the base payment for listing clear thoughts. Participants received the bonus if they were in the top half of participants listing the most thoughts, although, critically, participants were not explicitly told that this was the criterion for bonus payment (subjects were instead incentivized to provide clear thoughts).

Procedure

For Experiments 1a–1h, participants were randomly assigned to two of the eight questions[6] such that an approximately equal number of participants completed each question. For Experiments 2a and 2b, participants were assigned to only one question ("Can money buy happiness?" or "Is nuclear power safe?"), and were randomly assigned to first list a thought supporting Choice 1 *or* a thought supporting Choice 2. See Table 1 for all eight questions and possible responses. Figure 1 illustrates the order of tasks in our survey. Participants first engaged in the thought listing task. They were told to list all thoughts that came to mind as they considered a particular question, regardless of their eventual answer. They were also told that they should list thoughts in the order that they come to mind, and that their thoughts be complete English sentences understandable to a third party. We required participants to list at least one thought consisting of at least two words (two strings of characters separated by a space). After this, participants made their response,

and then rated each of their listed thoughts in terms of support for the two response options, on a 7-point Likert scale from "*Supported {choice 1} a great deal*" to "*Supported {choice 2} a great deal*." Finally, participants indicated their age and gender. The experiment instructions are provided in Appendix A.

Data Preprocessing

Due to an error in programming the survey for Experiments 1c and 1h, 13 and 10 subjects, respectively, did not record thought supports for one or more of their thoughts in these two experiments. We excluded all such participants (in alternative analyses we do not report; we simply deleted such thoughts but retained the participants; this does not qualitatively alter our results). We also excluded all participants in Experiments 2a and 2b who did not follow the priming instructions (e.g., first listed a thought supporting Choice 1 even though they were told to first list a thought supporting Choice 2), which led to 37 participants (of 252) and 37 (of 253) being dropped from Experiments 2a and 2b, respectively.

As is standard when using USE and other sentence encoders, we did not remove stop words, nor did we lemmatize or case standardize the words in our thoughts. Raw thoughts were thus passed through the transformer version of the USE to extract 512-dimensional, real-valued vectors for each thought. For each of the eight questions in Experiment 1, we then separated thoughts into three groups, one each for those supporting the first choice, the second choice, and neither choice. We performed separate k-means clustering on thoughts supporting the first choice, and thoughts supporting the second choice, using Scikit-learn (Pedregosa et al., 2011). Thus, if $k = 2$, we have five clusters (two for thoughts supporting the first Choice 1, two for the second choice, and one for neutral thoughts).[7] These clustering solutions obtained from Experiment 1 were reused in Experiment 2 (and Experiment 3; see Supplemental Materials). Additionally, to enable us to quantify the effect of semantic congruence in memory sampling, for a given question and value of k, we calculated the Euclidean distance between every pair of cluster centroids. The centroids were extracted through functionality provided by Scikit-learn. In general, the centroid of a cluster should of course simply be the mean of all the thought vectors assigned to that cluster (with minor differences between the mean and the centroid if the clustering algorithm stops before fully converging).

Qualitative Analysis

Overview of Data

The choice probabilities of the first option are summarized in Table 1. Participant responses were evenly distributed across the two responses options (the most uneven distribution is in Experiment 1e, in which 67% of participants chose a burger over salad).

---

[6] Note that although we obtained two responses per participant in Experiment 1, this is too little data to appropriately model individual heterogeneity. Thus, our analysis of Experiment 1 involves only group-level fits. To examine the issue of individual-level variance we conducted Experiment 3 (which is reported in Supplemental Materials).

[7] Note that in this article, we use k to represent the number of clusters that support the choice of an option (obtained using k-means clustering). On the other hand, K is the total number of clusters. Because there are k clusters supporting the first choice, k supporting the second choice, and one supporting neither, $K = 2k + 1$.

Note that in Experiments 2a and 2b participants were randomly assigned to one of two conditions. In Condition 1, participants were asked to begin by listing a thought supporting the first option, whereas in Condition 2, participants were asked to start with a thought supporting the second option. We combined participants of both priming conditions for most of the following analyses, except for the last result section, where we report results regarding priming effects.

Each participant listed as many thoughts as they wished before a decision was made (although participants were forced to list at least one thought). Distributions of the total number of thoughts sampled by a participant are illustrated in the top row of Figure 4. Each panel corresponds to an experiment. The median participant listed between 4 and 6 thoughts, and the modal number of thoughts ranged from 3 to 6, depending on the experiment. We elaborate on the model results visualized in Figure 4 in the section *Describing Data*. For now, we just point out that the primary model assumes absolute accumulation toward a threshold with partial context decay, the relative model assumes relative accumulation toward a threshold with partial context decay, the time limit model assumes a time limit mechanism instead of threshold accumulation, the full decay model assumes absolute accumulation with full context decay, and the no decay model assumes absolute accumulation with no context decay.
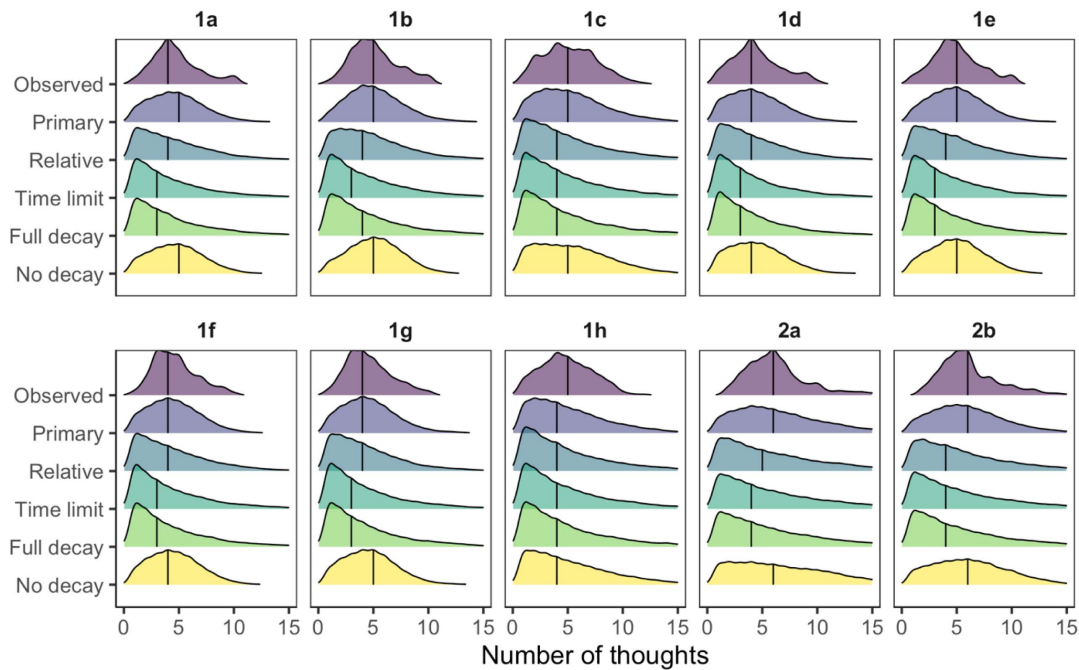
Thoughts generated by all participants in an experiment were clustered based on their semantic similarity and supporting response type. Here we report the results of the $k = 3$ clustering solution (in the Supplemental Materials we repeat our analysis with the $k = 2$ and $k = 4$ cluster solutions). With three clusters for thoughts in the first option set, three clusters for thoughts in the second option set, and one cluster for all neutral thoughts in an experiment, this solution leads to seven clusters in total for each experiment. The top row of Figure 5 illustrates the proportion of each thought cluster, with panels corresponding to different experiments. There was some variability in the shares of the seven thought clusters. Some clusters were highly infrequent, making up less than 10% of the data, whereas other clusters described more than 25% of participant thoughts (if each cluster was equally likely to be sampled, we would observe cluster frequencies of $1/7 = 14.3\%$ in our data). Again, we save greater discussion of the model results in Figure 5 for the section Describing Data, and for now just point out that the primary model is one with semantic and decision congruence and partial context decay but no cluster resampling, and the others are the same but with one or both of semantic and decision congruence turned off.
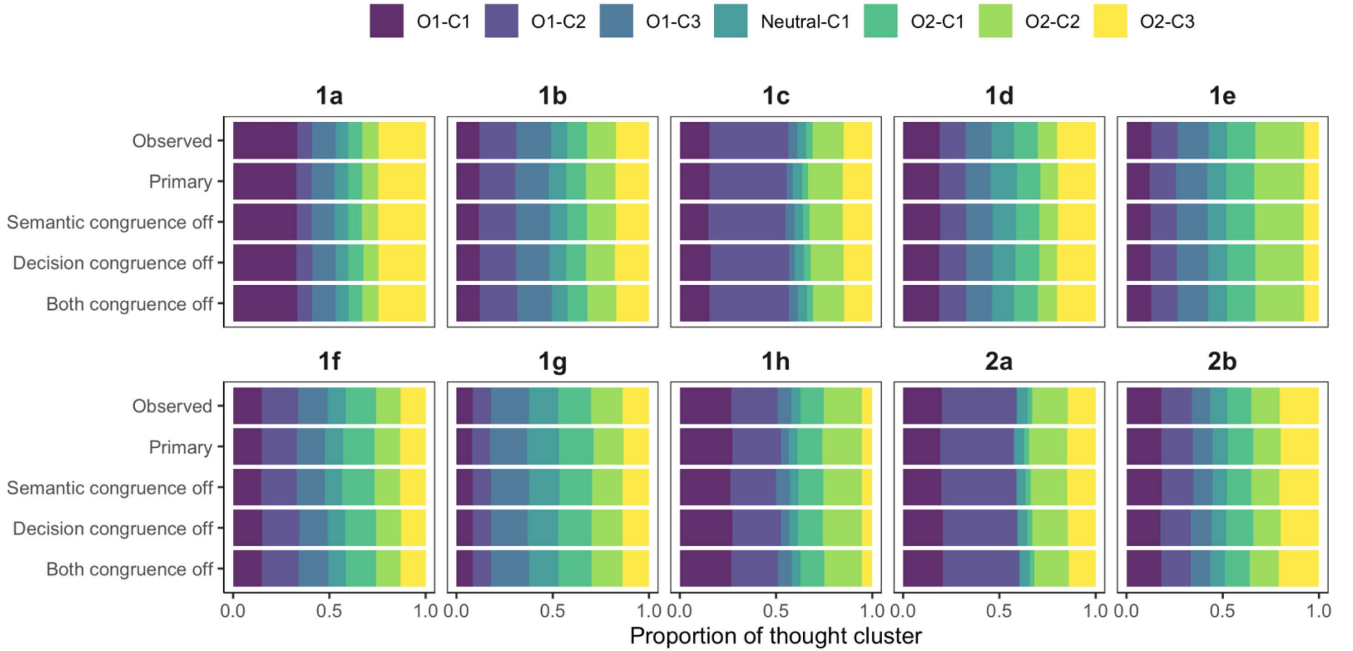
Memory Patterns

The memory models described above predict a number of different qualitative patterns in thought cluster recall. For example, nonzero values of ω_S, ω_D (and ω_R) yield thought sampling that is non-IID (independent and identically distributed) over time;

Figure 4
Distribution of Total Number of Thoughts Sampled Across Participants

Note. Each panel corresponds to an experiment. The solid vertical lines represent medians. Across 10 experiments, the primary model, assuming absolute accumulation of discrete supports toward a threshold with partial context decay, provides the closest predictions regarding the median and mode of the observed thought lengths. We also display the predictions of a model with relative accumulation, a model with a time limit mechanism instead of threshold accumulation, a model with full decay, and a model with no decay. See the online article for the color version of this figure.

Figure 5
Distributions of Thought Clusters From the k = 3 Clustering Solution in the 10 Experiments

Note. Here O1–C1, O1–C2, O1–C3 represent clusters for thoughts supporting the first option. Neutral-C1 represents the cluster for all neutral thoughts. O2–C1, O2–C2, O2–C3 represent clusters for thoughts supporting the second option. The primary model, with semantic and decision congruence and partial context decay but no cluster resampling, provides almost perfect predictions of the observed cluster proportions. However, so do models that turn decision and semantic congruence off. Note that all models have flexible baseline cluster activation parameters, allowing them to accommodate aggregate thought cluster distributions. See the online article for the color version of this figure.

thought clusters that are congruent with the previously sampled thought, semantically or decision-wise, are more/less likely to be recalled. If $\omega_S = \omega_D = 0$ (and $\omega_R = 0$) then we would not observe the memory dynamics predicted by these models. Note that semantic congruence is a feature of many previous memory models, such as CMR, SAM, and ACT-R, as well as models that describe memory search as a random walk in a semantic network. Likewise, decision congruence is a feature of query theory. Our experiments test for the magnitude of these effects in a memory-based judgment task.

Figure 6 illustrates how sampled thought clusters at time t relate to the clusters sampled at $t - 1$. Again, these figures use the $k = 3$ k-means thought clustering solutions for each experiment. In the left panel, the asterisks describe the mean Euclidean distance between the cluster centroids of two sequentially sampled thought clusters in our observed data, averaged across participants. In the right panel, the asterisks describe the probability of sampling a thought that supported the same response option as the previously sampled thought, averaged across participants. These are our measures of semantic and decision congruence, respectively.

Did participants sample congruent clusters at an above-chance level? To answer this question, we need to control for participant heterogeneity, that is, the possibility that thoughts generated by a participant may be more congruent with each other, compared to those generated by different participants. Participant heterogeneity can lead to the appearance of semantic and decision congruence on the aggregate level, and it needs to be controlled using
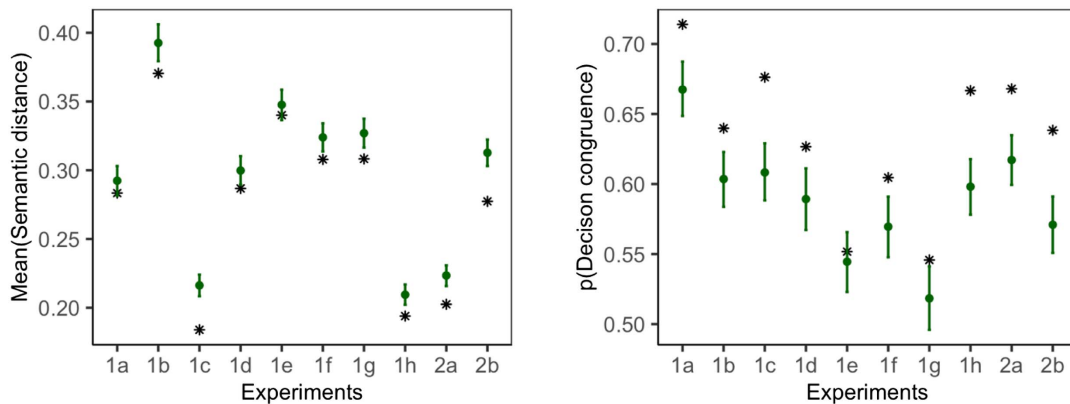
permutation analysis. For our permutation analysis, we randomly reordered the thought clusters retrieved by each participant and computed the average of the mean Euclidean distance between two sequentially sampled clusters among participants. We repeated this 10,000 times to calculate the average semantic congruence measures, and 95% confidence intervals, that would be observed if participant heterogeneity was the sole determinant of semantic congruence. We performed a similar analysis for decision congruence. The solid dots and error bars in Figure 6 illustrate these statistics. As can be seen, most asterisks lie below the 95% confidence intervals around the average permuted semantic congruence statistic, indicating that on a participant level, thought clusters sampled next to each other were more semantically related (had less semantic distance) than thought clusters sampled with longer time in between. Likewise, most asterisks lie above both the error bars for the average permuted semantic decision-congruence statistic, indicating that, with individual heterogeneity controlled for, contiguously sampled thoughts were more likely to support the same option, compared to thoughts sampled further apart.

Note that the above patterns also emerge with the $k = 2$ and $k = 4$ cluster solutions, as shown in the Supplemental Materials.

Decision Patterns

The decision models discussed in the previous section also predict a number of different qualitative patterns in our support rating and

Figure 6

Permutation Analyses for Semantic Congruence and Decision Congruence Effects

Note. The asterisks represent observed mean Euclidean distance between the cluster centroids of contiguously sampled thought clusters (the left panel), and observed probabilities of sampling a thought supporting the same option as the previously sampled thought (the right panel). Solid dots and error bars are means and 95% confidence interval based on 10,000 permutations in which we randomly reordered thought clusters generated by each participant so that individual heterogeneity in cluster base rates was preserved but sequential effects were broken. These figures display significant semantic and decision-congruence effects in most experiments. See the online article for the color version of this figure.

decision data. For example, models that assume that decision makers use threshold decision rules predict that the probability of terminating the decision and choosing a given option increases as the accumulated (absolute or relative) support for that option increases. Conversely, models that assume that decision makers accumulate information until a time limit is reached predict that the probability of terminating the decision is insensitive to the accumulated support (though the response option that is eventually selected does depend on accumulated support). Models that do not assume any accumulation at all (i.e., have full context decay), such as the lexicographic semiorder heuristic, would predict that response probabilities depend only on the support sampled in the previous time period (and not on the supports sampled in earlier time periods).

Some of the above patterns can be observed in Figures 7 and 8. Figure 7 shows the probability of continuing deliberation (upper panel) or selecting the first option (lower panel), as a function of the cumulative support for the first relative to the second option until that point in time. Here we can see that the probability of continuing deliberation decreases in the magnitude of cumulative support, consistent with the predicted patterns of threshold models and contrary to the predicted patterns of time limit models. Likewise, the probability of selecting the first response option also depends on cumulative support, even as cumulative support increases above +3 (which is the maximum support that can be emitted by a single thought sample). This suggests that decision makers are aggregating support over time, contrary to the predicted patterns of models that assume full context decay (these models would predict that cumulative support beyond +3 would not influence response probabilities, as decision makers cannot aggregate the supports emitted by more than a single thought cluster).

Figure 8 shows the probabilities of these decisions as a function of the number of thoughts generated. That is, we calculated participants' probability of continuing deliberation (upper panel),

or choosing the first option (lower panel), after sampling a certain number of thoughts. Again, we see that the probability of continuing deliberation decreases, and the probability of selecting an option increases, as the number of generated thoughts increases (the bottom panel of Figure 8 shows that the probability of selecting the first option increases, but note that the probability of selecting the second option increases as well, since the probability of continuing deliberation declines). This pattern contradicts the simulated patterns of full context decay models. Without accumulating supports, the probability of making any decision should be independent of the number of thoughts previously sampled.
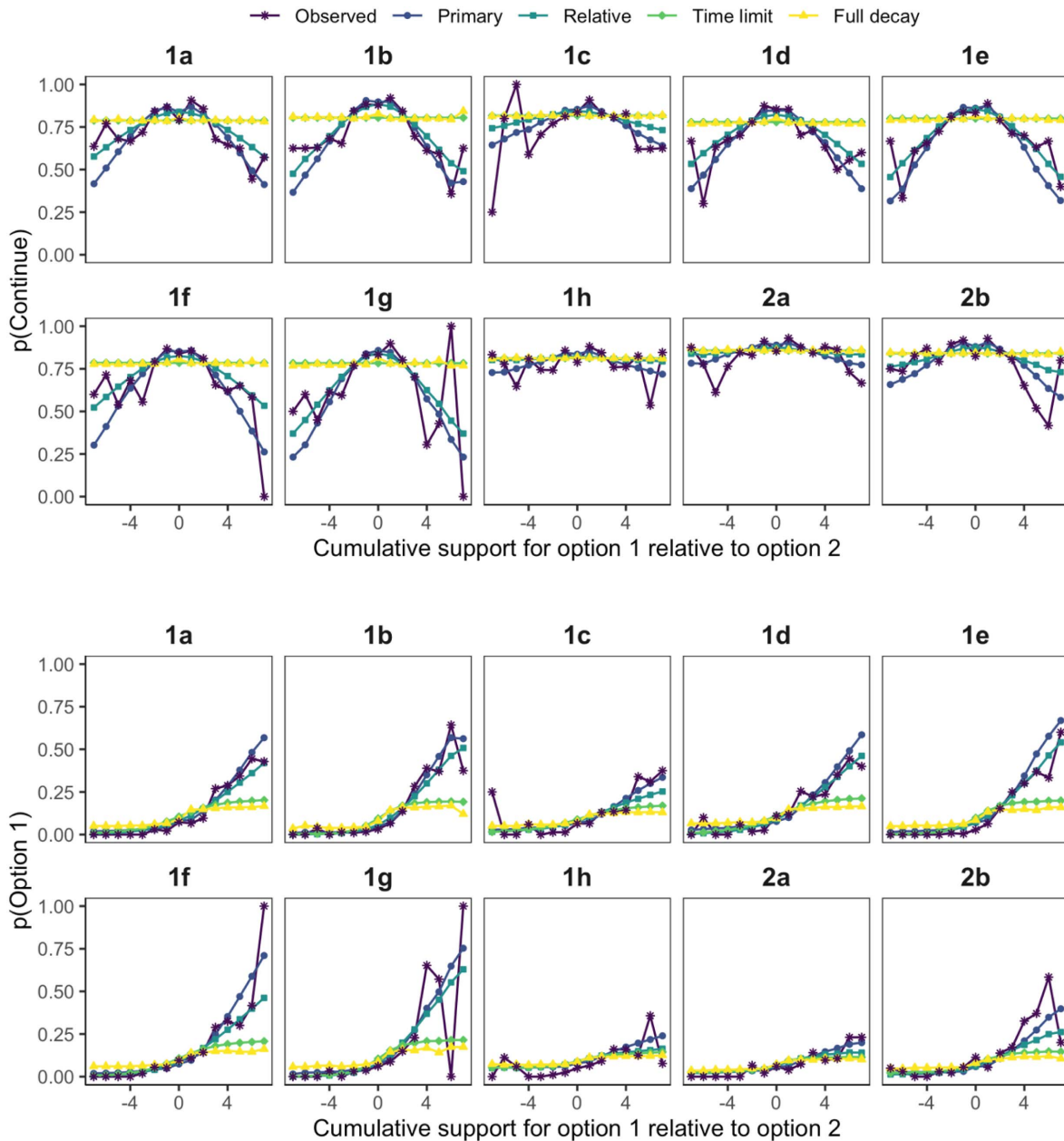
Quantitative Analysis

Best-Fit Models

The qualitative patterns in Figures 6–8 do provide some evidence for semantic congruence and decision congruence in memory, and the accumulation of supports to a threshold in decision making with partial or no context decay. To rigorously understand these patterns in our data, and to disentangle the above mechanisms from other related mechanisms that could be responsible for the observed patterns, we fit 576 distinct memory-based decision making models (24 memory models × 24 decision models) to the data from each of the 10 experiments using Bayesian model fitting. Because of the conditional independence hypothesis, we were able to separately fit the 24 memory models to thought listing data and the 24 decision models to decision data. As outlined above, these models are obtained by combining different memory and decision mechanisms, and a switchboard analysis comparing each of these models against each other can shed light on the combination of memory and decision mechanisms necessary to describe the data. Note that as in the qualitative analysis above, we pool the data from each of the two conditions in Experiments 2a and 2b. However, as these

Figure 7
Probability of Continuing Sampling (i.e., One Minus the Probability of Ending Deliberation; Upper Panel), and Probability of Selecting the First Response (Lower Panel), as a Function of the Cumulative Support for the First Relative to the Second Option, in the 10 Experiments
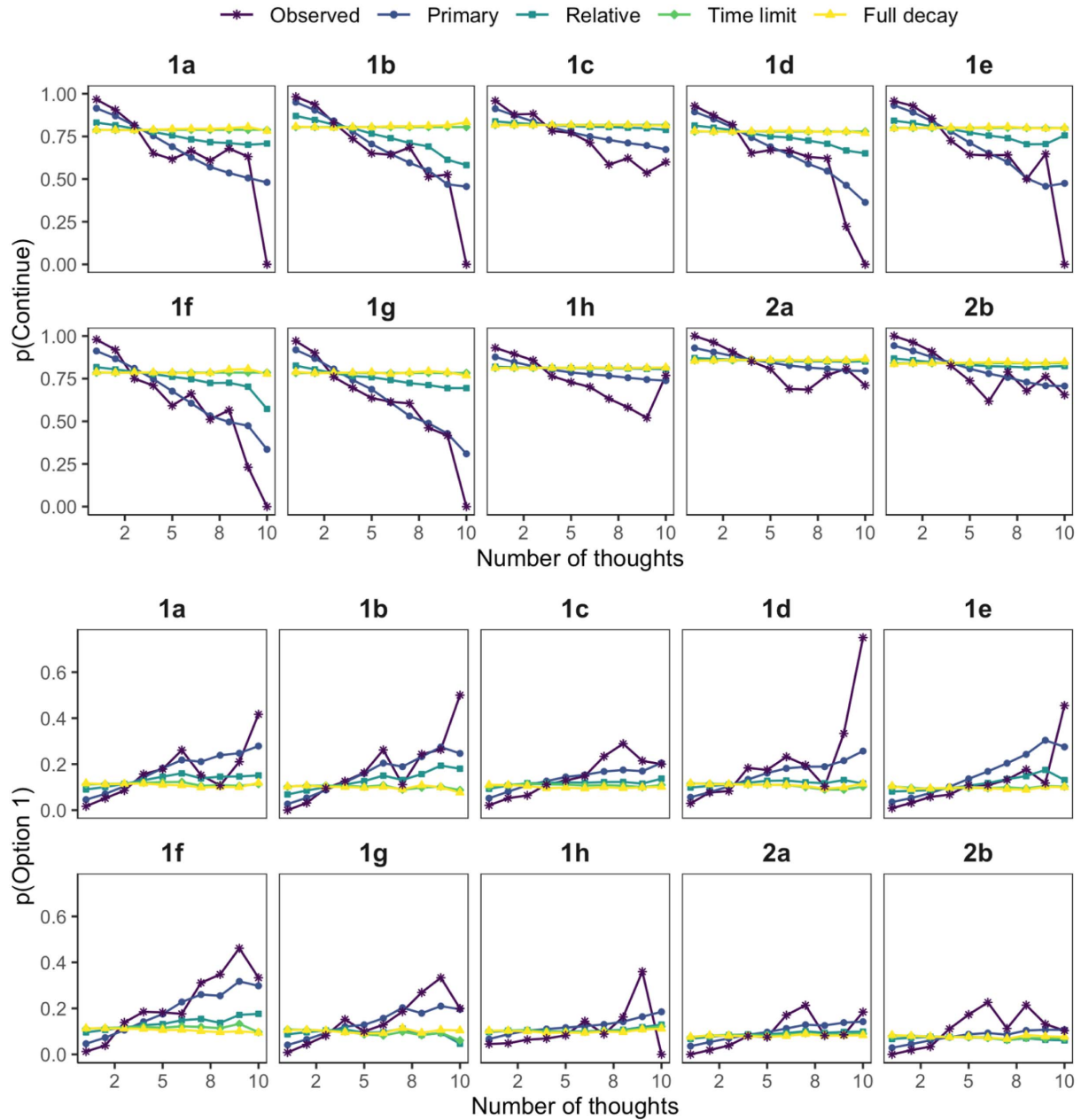

Note. The probability of continuing sampling thoughts decreases in absolute values of relative support, whereas the probability of choosing the first option increases in cumulative support. These patterns are consistent with the predictions of threshold models with partial or no context decay, such as the primary model that assumes absolute accumulation of discrete supports toward a threshold with partial context decay, and a similar variant that assumes relative accumulation toward a threshold. On the contrary, the pattern cannot be captured by the model variant that assumes full decay or a model that assumes an external time limit mechanism instead of threshold accumulation. See the online article for the color version of this figure.

experiments asked participants to begin the thought listing task by retrieving a thought that supports either the first or the second response option, we constrained the sampling probabilities of neutral and opposite thought clusters to be 0 (in the next section

we analyze the effects of the experimental manipulation in these two experiments in detail). Also note that as we did not elicit multiple decisions from each subject in Experiments 1 and 2, all our fits were performed on the group level. Additional details regarding model

Figure 8

Probability of Continue Sampling (Upper Panel), and Probability of Selecting the First Response (Lower Panel), as a Function of the Number of Thoughts Generated, for the Ten Experiments

Note. The probability of continuing sampling decreases, and the probability of selecting an option increases, as the number of generated thoughts increases. Our primary model, which assumes absolute accumulation of discrete supports toward a threshold with partial decay, predicts the observed pattern better than a similar model with relative accumulation. Model variants assuming a time limit instead of threshold accumulation, or assuming full context decay, do not capture behavioral patterns. See the online article for the color version of this figure.

fitting, such as parameter priors and model convergence, are provided in Appendix B. The Supplemental Materials repeat our analysis for the data from Experiment 3, which allows for the hierarchical modeling of participant heterogeneity.

In Table 3 we show which of these mechanisms are present in the best-fit model in each of our 10 experiments, evaluated based on Watanabe–Akaike information criteria (WAIC; Watanabe, 2010) of the joint thought listing and decision data. On the memory side, we

Table 3
Memory and Decision Mechanisms in Best-Fit Model (Based on WAIC) in Each Experiment

Experiment		1a	1b	1c	1d	1e	1f	1g	1h	2a	2b	# Best fit
Memory mechanisms												
Cluster resampling	On	✗	✓	✓	✗	✓	✓	✗	✓	✗	✗	5
	Off	✓	✗	✗	✓	✗	✗	✓	✗	✓	✓	5
Semantic congruence	**On**	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	9
	Off	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	1
Decision congruence	**On**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
	Off	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
Context decay in memory	Full	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
	Partial	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
	None	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
Model WAIC		3,637	4,187	3,600	3,733	4,251	4,045	4,047	3,816	3,899	4,461	
Decision mechanisms												
Support	Continuous	✗	✓	✗	✗	✗	✗	✓	✓	✗	✗	3
	Discrete	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	7
Accumulation	Relative	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
	Absolute	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
Decision rule	**Threshold**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
	Time Limit	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
Context decay in decision	Full	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	0
	Partial	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	8
	None	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	2
Model WAIC		1,271	1,130	1,433	1,284	1,232	1,267	1,234	1,511	1,404	1,281	

Note. Memory and decision mechanisms that are present in a majority of experiments are bolded. The memory models and decision models were fit to data separately. WAIC = Watanabe–Akaike information criteria.

can see that the best-fit models display both semantic and decision congruence in memory, except in Experiment 1e, whose best-fit model has semantic congruence turned off ($\omega_S = 0$).[8] All the best performing models display partial context decay. We do not find consistent evidence of cluster resampling. On the decision side, all the best-fit models exhibit absolute accumulation of evidence to a threshold, with context maintenance (eight have partial context decay and two have no decay).[9] There is no consistent evidence of the use of continuous versus discrete supports across the experiments.

The above results persist with the $k = 2$ and $k = 4$ cluster solutions, as shown in the Supplemental Materials. Note that the results reported in the main text do not account for heterogeneity across individuals and/or across questions. Again, in the Supplemental Materials, we report the methods and results of an additional experiment (Experiment 3), in which 51 participants each responded to all eight decision prompts. We conducted our switchboard analyses using hierarchical memory models and hierarchical decision models, and obtained similar conclusions. The Supplemental Materials also provide results of parameter recovery analyses showing that the parameters of the models can be adequately recovered from data.

Importance of Mechanisms

In Table 4 we show the overall importance of the different mechanisms in our experiments by displaying the mechanisms present in the top-10 models based on aggregate WAIC, that is, the sum of the WAICs for the 10 experiments. The best-fit memory model, in terms of aggregate WAIC, displays cluster resampling, semantic and decision congruence, as well as partial context decay. The second, third, and fourth best-fit models retain partial context decay, but turn off cluster resampling, decision congruence, or semantic congruence,

respectively. Notably, turning off cluster resampling leads to a very small WAIC increase (ΔWAIC = 86, aggregated over 10 experiments), demonstrating that this nested (and thus more parsimonious) model actually fits data almost as well as the best-fit full model. This is not surprising, as the effect of cluster resampling overlaps with that of semantic and decision congruence (resampling from a same thought cluster corresponds to sampling from a decision-congruent cluster with zero semantic distance). In fact, all the qualitative memory patterns we identified in previous sections can be generated by the more parsimonious, nested model without cluster resampling. Henceforth we refer to this model, with semantic and decision congruence and partial context decay but no cluster resampling, as the *primary memory model*.

As Table 4 shows, the best-fit decision model displays absolute accumulation of discrete support to a threshold in decision making, with partial context decay. Here, accumulation toward a threshold is an important feature, shared by all the top-eight decision models. Absolute, rather than relative, accumulation is favored by all models ranked in the top-four. Partial or no context decay, rather than full context decay, is exhibited by all the top-10 models. There is no consistent evidence regarding discrete versus continuous support accumulation: The best model has a discrete support accumulation, whereas the close-performing 2nd best model has a continuous support accumulation (ΔWAIC = 62, aggregated over 10

---

[8] The decision prompt for Experiment 1e is "Would you prefer to eat a salad or a burger for dinner?" In our experiments, participants display weaker semantic congruence in questions regarding consumer preferences, compared to other decision domains. For detailed discussions, please refer to the section on Parameters in the Hierarchical Modeling and Participant Heterogeneity section of the Supplemental Materials.

[9] Overall, the magnitude of context decay parameters in the decision models are much smaller (closer to 0), compared to those in the memory models in all experiments. We elaborate on this point in the Parameters section below in the main text.

Table 4

Top-10 Memory and Decision Models Based on Aggregate WAICs Across All 10 Experiments

Model rank		1	2	3	4	5	6	7	8	9	10
Memory mechanisms											
Cluster resampling	**On**	✓	✗	✓	✓	✓	✓	✗	✗	✓	✗
	Off	✗	✓	✗	✗	✗	✗	✓	✓	✗	✓
Semantic congruence	**On**	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓
	Off	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Decision congruence	**On**	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗
	Off	✗	✗	✓	✗	✗	✗	✗	✗	✓	✓
Context decay in memory	Full	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗
	Partial	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓
	None	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗
Aggregate model WAIC		39,685	39,770	40,179	40,264	40,403	40,480	40,519	40,551	40,573	40,643
Decision mechanisms											
Support	Continuous	✗	✓	✗	✓	✓	✓	✗	✗	✓	✗
	Discrete	✓	✗	✓	✗	✗	✗	✓	✓	✗	✓
Accumulation	Relative	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
	Absolute	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
Decision rule	**Threshold**	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
	Time limit	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
Context decay in decision	Full	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
	Partial	✓	✓	✗	✗	✓	✗	✓	✗	✓	✓
	None	✗	✗	✓	✓	✗	✓	✗	✓	✗	✗
Aggregate model WAIC		13,065	13,127	13,248	13,264	13,629	13,702	13,759	13,828	14,036	14,133

Note. The model mechanisms that are present in the best model are bolded. Here Rank 1 corresponds to the best performing model. The memory models and decision models were fit to data separately. WAIC = Watanabe–Akaike information criteria.

experiments). Out of the top-10 models, half of them integrate supports in a discrete manner whereas the others integrate supports in a continuous manner. In the following analyses, we consider the model with absolute accumulation of discrete supports to a threshold, with partial context decay, as the *primary decision model*.

In Figure 9 we display changes in WAICs in the 10 experiments as each of the mechanisms in the primary memory and decision models are altered. The top-left panel shows that turning cluster resampling has a negligible effect on WAIC. The remaining panels on the top row show that turning semantic congruence, decision congruence or partial decay off leads to notable increases in WAICs. The bottom-left panel shows how WAICs change when we assume discrete support rather than continuous support in the primary model. Depending on different experiments, discrete or continuous support results in better model fits, but overall, the impact of this mechanism is very small. The remaining panels on the bottom row illustrate the effects of switching the accumulation rule (from absolute to relative), the decision rule (from threshold based to time limit), and the context decay (from partial or no decay to full decay): These changes all lead to much worse model fits.

Figure 9 also shows that altering the different mechanisms has a different effect on fits across our 10 experiments. Thus, for example, turning off decision congruence increases the WAIC by almost 200 in Experiment 1a but increases WAIC by only 16 in Experiment 1c. In contrast, replacing absolute accumulation with relative accumulation increases the WAIC by 102 in Experiment 1f, but only 39 in Experiment 1h. Thus, even though all decision settings display semantic and decision congruence in memory, the absolute accumulation of evidence to a threshold in decision making, and context maintenance in both memory and decision making, not all settings require these mechanisms to the same extent.
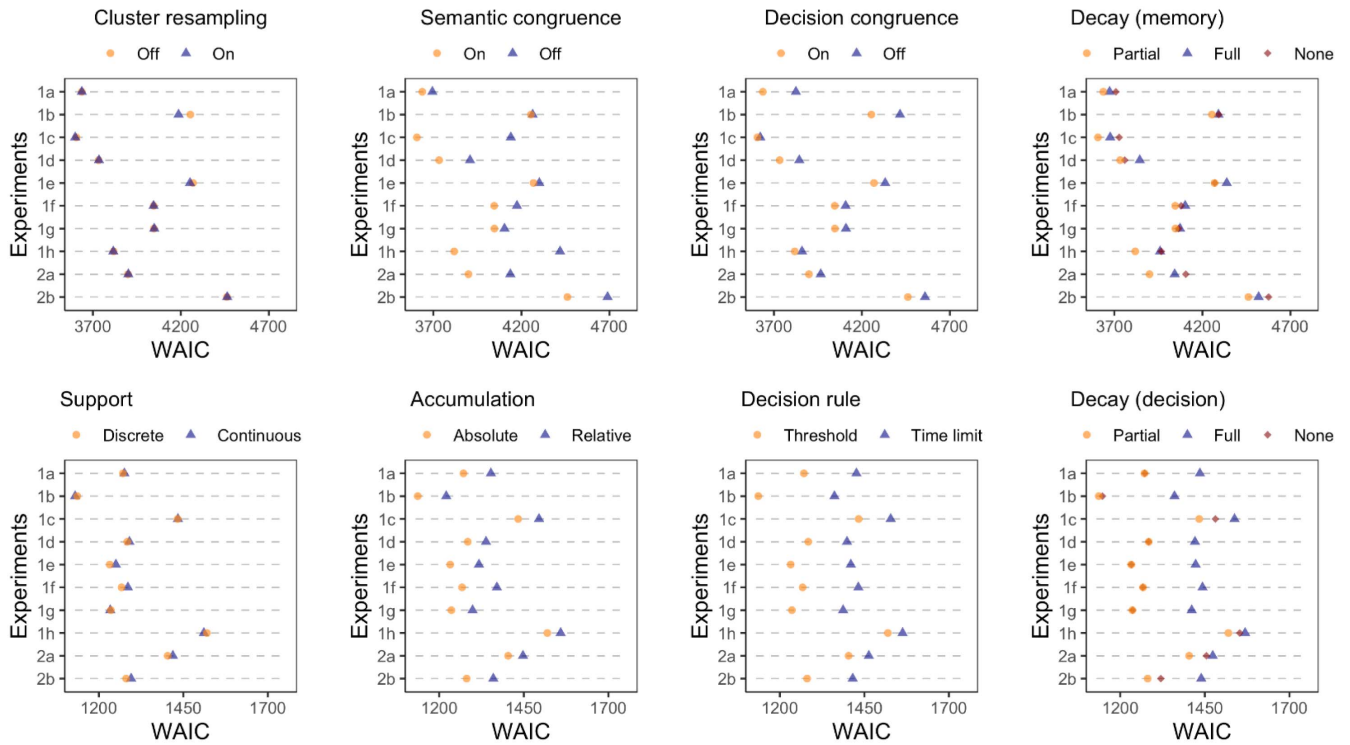
The above results persist with the $k = 2$ and $k = 4$ cluster solutions, as shown in the Supplemental Materials. All the results

are replicated using hierarchical model analyses of Experiment 3 data (which are reported in the Supplemental Materials), except that the best-fit hierarchical decision model exhibits no context decay in decision making (rather than partial context decay).

Parameters

Finally, Figure 10 displays the best fitting parameter values in our models. The first row shows posterior means and 95% credible intervals of parameters from the best-fit, full memory model. Most of the 10 experiments have small negative values of ω_R, large negative values of ω_S, large positive values of ω_D, and moderate values of δ_M, except for Experiment 1b and 1e, which have positive ω_R and positive ω_S. These parameters again show that most experiments display small (nonsignificant amounts of) recall inhibition and repetition suppression, and large (significant) amounts of decision and semantic congruence, as well as partial context decay. The second row shows parameters from the primary decision model. Here we see that the decision decay parameters (δ_D) are closer to 0 and are smaller than the memory decay parameters (δ_M). This suggests that memory and decision making have different degrees of decay in context.[10] The observed patterns in best-fit parameters persist with the $k = 2$ and

---

[10] The inconsistency between estimated decay rates in the memory and decision models requires some justification at the level of cognitive processing. Assuming context decay in both memory and decision making offers a theoretically cohesive perspective on capacity limits in our article (in addition to integrating the assumptions of CMR memory models and leaky integration decision models), however differences in parameters between the two decays indicate that despite their architectural similarities, memory and decision context may involve separate implementations. This could be due to differences in the types of evidence that these mechanisms store, and the types of computational goals that these mechanisms optimize or adapt to.

Figure 9

WAIC Changes as We Alter the Primary Model Mechanisms in the 10 Experiments

Note. The primary memory model assumes semantic and decision congruence and partial context decay but no cluster resampling. Turning cluster resampling on does not change WAICs much; turning the other three mechanisms off increases WAICs substantially. The primary decision model assumes absolute accumulation of discrete supports toward thresholds, with partial context decay. Switching to continuous supports does not influence WAICs much; switching the other three mechanisms leads to large WAIC increases. WAIC = Watanabe–Akaike information criteria. See the online article for the color version of this figure.

$k = 4$ cluster solutions, as well as the hierarchical model analyses, as shown in the Supplemental Materials.
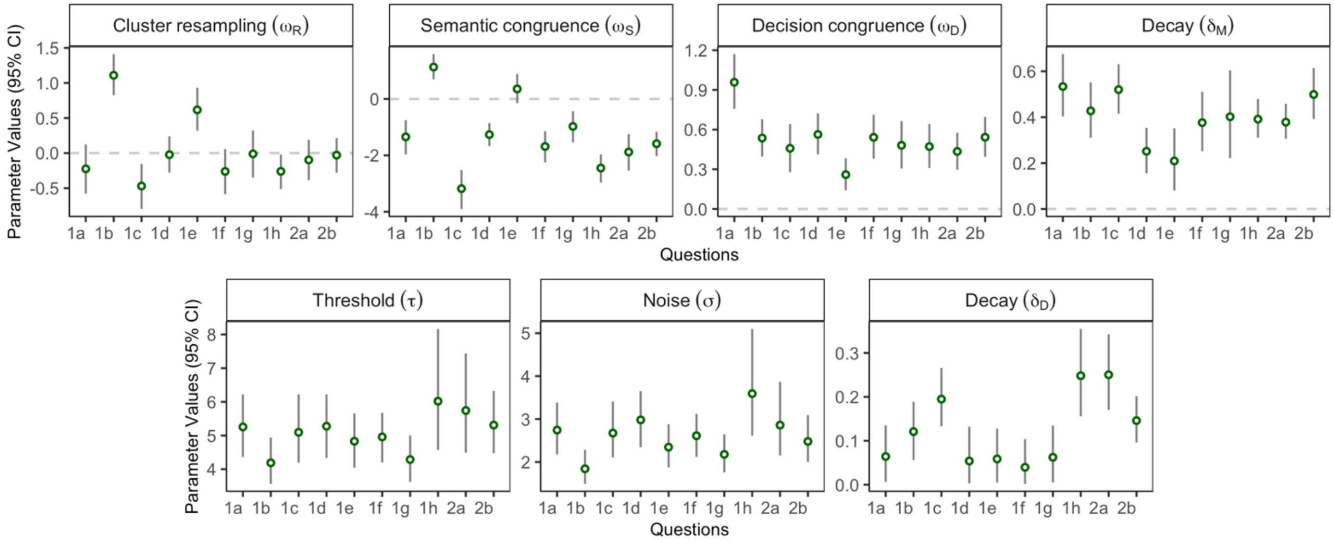
Describing Data

Choice Probability

Our framework provides quantitative descriptions regarding participant responses in the types of naturalistic memory-based decision tasks in Experiments 1a–1h and 2a–2b. We begin by examining choice probability. Here we used the posterior distributions of the decision parameters, as well as the support ratings generated by each participant. Note that we do not have support ratings for the thoughts that would have been generated by participants had they continued deliberating past their decision, and thus cannot model the dynamics of the (hypothetical) decision process past the point of choice. We address this limitation by considering our predicted posterior probability of the participant's choice at the time of choice. If this probability is greater than .5, we say that our model predicts that the corresponding option is chosen. If not, then we predict a 50–50 choice of the two options. Figure 11 displays the accuracy rates of our model predictions with this assumption. In the upper panel, it displays the accuracy rate for the first option and in the lower panel it displays the accuracy rate for the second option.

Across the 10 experiments, our primary decision model, which assumes absolute accumulation of discrete supports toward a threshold with partial context decay, achieved an averaged accuracy rate of 84.6% (minimum accuracy of 74.7% and maximum accuracy of 93.4%) when predicting the choices of participants that selected the first option. We compared these accuracy rates to those generated by four model variants. The first variant assumes relative accumulation toward a threshold, the second assumes relative accumulation with time limits (this model variant fits observed data better than a model that assumes absolute accumulation with time limits as shown in Table 4), the third assumes full context decay, and the fourth assumes no context decay. The remaining assumptions of these models are identical to those of the primary decision model. The accuracy rates of these models are 84.5%, 84.0%, 79.8%, and 86.1%, respectively. Likewise, for participants that selected Option 2, our primary model achieved an accuracy rate of 74.7% (minimum 59.3%, maximum 87.9%), and the four model variants achieved accuracy rates of 75.3%, 75.4%, 72.8%, and 73.1%, respectively. Note that accuracy rates are slightly lower for participants that selected the second option, because a small portion of participants selected the second option without listing any thoughts supporting that option. We speculate that this may be due to a combination of order effects that influence thought listing (i.e., by cuing thoughts that support the first option) and decision noise (that subsequently cause the second option to be chosen).

Figure 10

Parameter Values for the Full Memory Model and the Full Decision Model for the 10 Experiments

Note. The full memory model (first row) assumes flexible cluster resampling, semantic congruence, and decision congruence with partial context decay. The full decision model (second row) assumes absolute accumulation toward threshold with partial context decay. The dots represent posterior means, and the error bars indicate 95% credible intervals. See the online article for the color version of this figure.

Number of Thoughts

Figure 4 displays the posterior predictive distribution of the number of thoughts generated by each participant in each experiment, based on our model fits. Note that to simulate the number of thoughts sampled before a decision, we need parameter values from both the memory model and the decision model. Here we used posterior distributions of the primary memory model parameters to simulate the cluster sampling process. At Time 1, the primary memory model samples a thought according to the baseline activation strengths of the seven clusters (γ); at Time 2 onwards, it samples a thought cluster according to the baseline activations, as well as decision and semantic congruence parameters (ω_D and ω_S), and context decay parameters (δ_M). At each time point, we simulate a decision (select one of the two options or continue sampling) based on the primary decision model parameters. This way, our memory model and decision model, when combined together, can simulate human-like memory search and decision. Here we present results based on 12,000 rounds of simulations using posterior distributions of parameters. Like before, we compare the performance of our primary decision model, to four other model variants. Recall that these variants are the same as the primary model, except that they assume relative accumulation toward a threshold, relative accumulation with a time limit decision rule, full context decay, or no context decay, respectively.

From Figure 4 we can see that the predicted density plots from the primary model resemble the observed plots rather closely, showing that our primary model is able to capture not only choice probabilities but length of deliberation (analogous to response time measures in cognitive decision research) as well. Across the 10 experiments, the primary model makes the most accurate description of modal number of thoughts, as evaluated by mean absolute deviation across experiments (MAD; primary: 0.9; relative: 3; time limit: 3.2; full decay: 3.2; no decay:
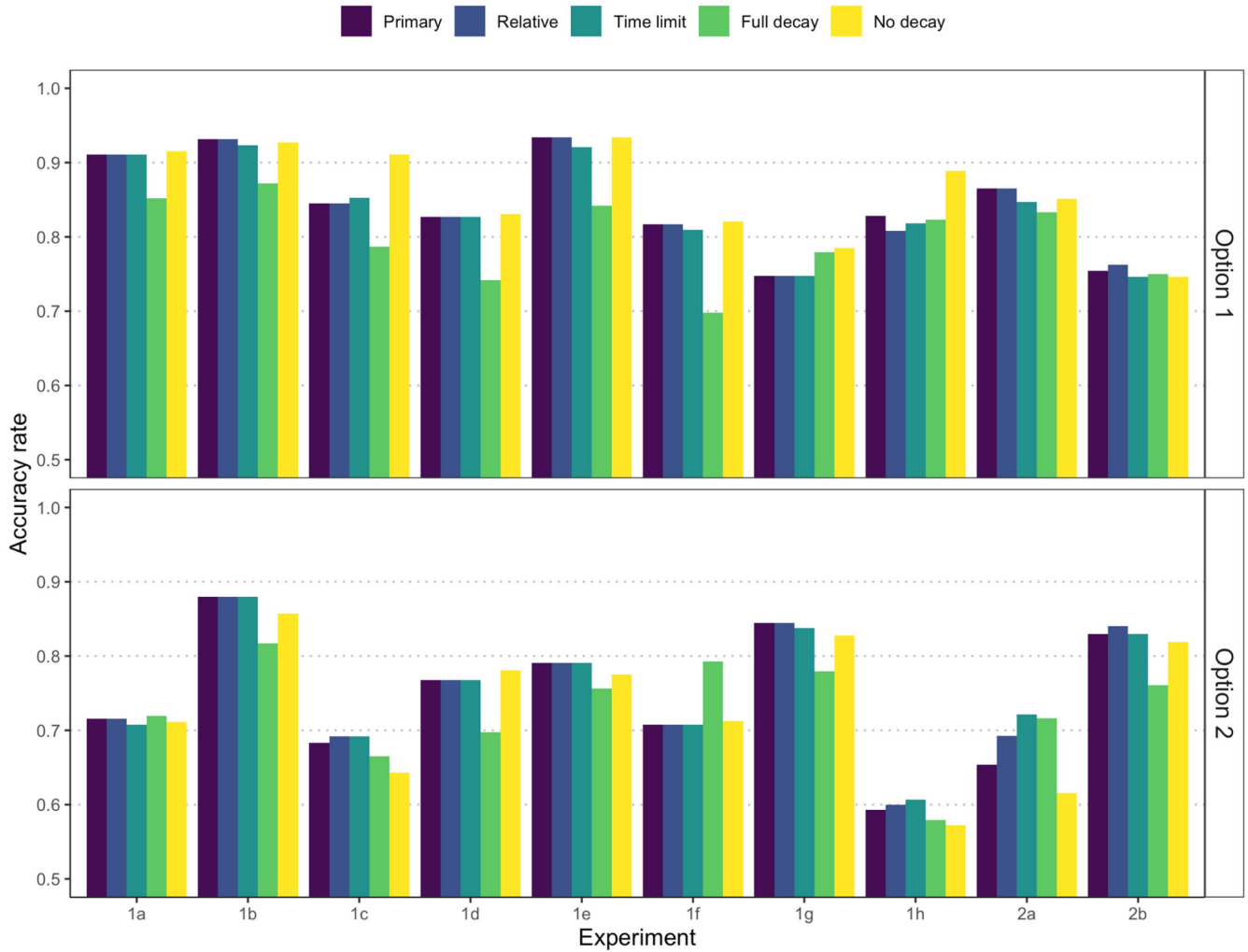
1.3). Compared to the other variants, the primary model also provides more accurate description of the median, as evaluated by MAD (primary: 0.2; relative: 0.6; time limit: 1.0; full decay: 1.1; no decay: 0.3). Overall, the models with relative accumulation or no decay provide slightly worse descriptions than the primary decision model; whereas the models with time limits or full decay provide substantially worse descriptions (these models are likely to underpredict the total number of thoughts when decision makers sample both positive and negative supports and thus do not accumulate strong enough preferences to make a decision).

Thought Clusters

Using simulated decisions from the combined memory-decision model, we can predict the distribution of thought clusters for each experiment. For this, we used the parameter values estimated from the primary decision model, and compared the performance of four memory models, including the primary memory model, a constrained model with semantic congruence turned off, one with decision congruence turned off, and the final one with both turned off. Cluster distributions based on 12,000 simulated decisions are presented in Figure 5. Here we observe that the simulated cluster distributions of the primary memory model closely match observations; the MAD of the proportions of clusters is only 0.002 over our 10 experiments. However, models without decision congruence, semantic congruence, or any congruence effects also do very well at predicting thought cluster distribution (MAD of 0.004, 0.003, 0.005, respectively). This is likely due to the fact that these models all permit flexible baseline cluster sampling probabilities, $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_7]$, which can adjust to capture cluster sampling frequencies even when decision and semantic congruence are turned off. We obtain similar results with the $k = 2$ and $k = 4$ cluster solutions, as shown in the Supplemental Materials.

Figure 11

Choice Prediction Accuracy for Participants Selecting the First Response Option (Upper Panel) and the Second Response Option (Lower Panel), for the 10 Experiments

Note. Note that the *y*-axis starts from 50%, which is the chance level of prediction accuracy. The primary model assumes absolute accumulation of discrete supports toward a threshold with partial context decay, though most other models achieve a similar degree of accuracy. See the online article for the color version of this figure.
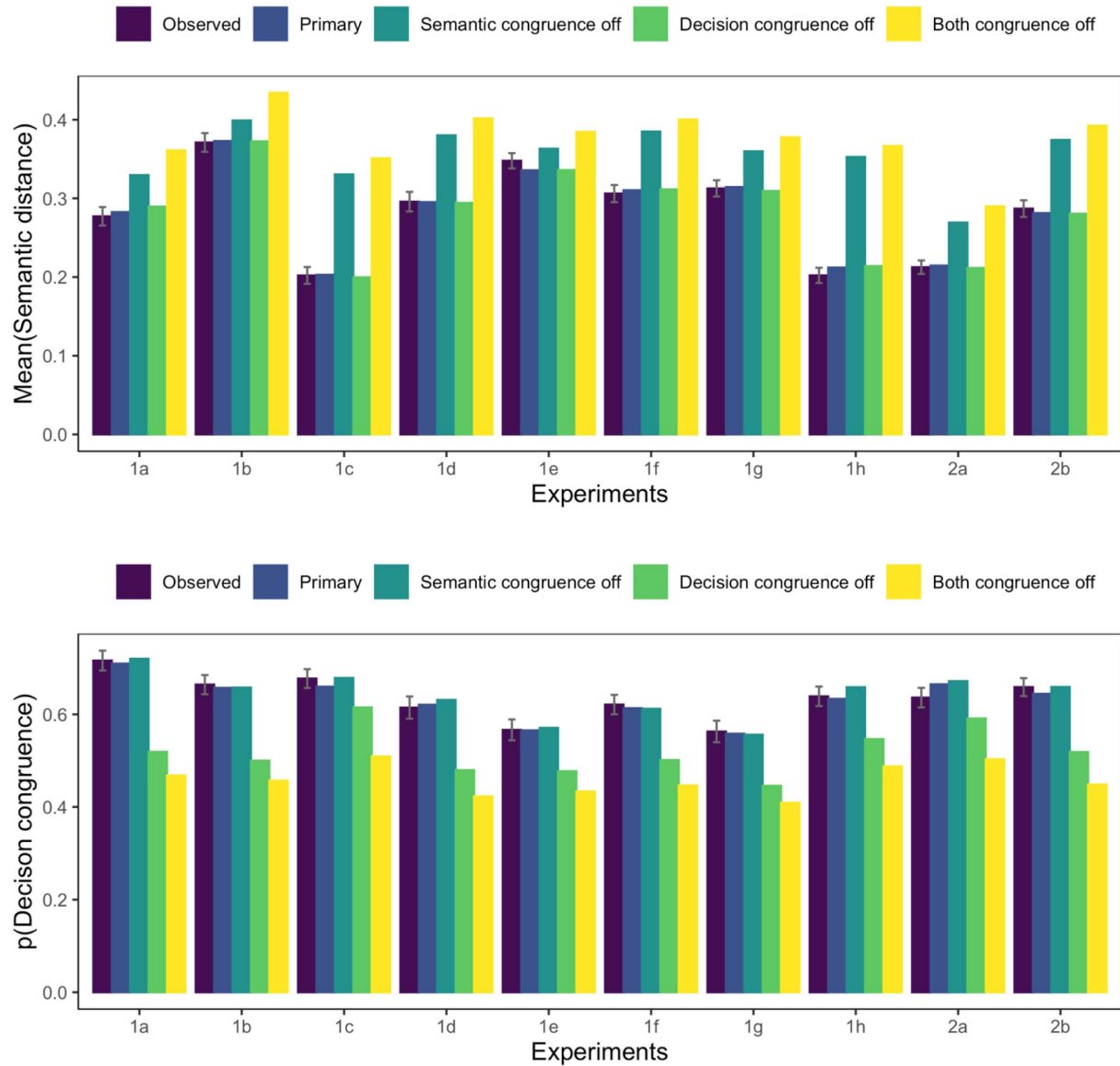
Memory Patterns

The results in Figure 5 suggest that semantic and decision-congruence assumptions of the models may not be needed for capturing observed qualitative patterns of thought cluster distributions. However, in Figure 12 we show that this conclusion is premature—a full memory model which permits semantic and decision congruence is necessary for describing dynamic dependencies in thought cluster sampling. Such a model predicts the probability of sampling a semantic or decision congruent cluster to be higher than that generated by an IID sampling model, a pattern reflected in human data. We simulated 12,000 rounds of the decision process using parameter posterior distributions from the primary decision model and the four memory models mentioned above. Our primary memory model, with both semantic and decision congruence on, achieves a MAD of 0.004 for predicting the mean semantic distance between contiguous thoughts (upper panel), and a MAD of

0.010 for predicting the probability of sampling a decision-congruent thought (lower panel). In contrast, a model without semantic congruence has a MAD of 0.073 for predicting the semantic congruence pattern and a MAD of 0.010 for predicting the decision-congruence pattern. Likewise, a model without decision congruence has a MAD of 0.006 for predicting the semantic congruence pattern but a MAD of 0.116 for predicting the decision-congruence pattern. Finally, a model without any congruence effects fails to predict both patterns, with a MAD of 0.095 for semantic congruence and a MAD of 0.177 for decision congruence. We obtain similar results with the $k = 2$ and $k = 4$ cluster solutions, as shown in the Supplemental Materials.

Decision Patterns

In Figures 7 and 8 we show that the assumptions in our primary decision model are also necessary for describing the qualitative

Figure 12

Dynamic Dependencies in Thought Cluster Sampling Across the 10 Experiments

Note. Note that the primary model has both semantic and decision congruence turned on. All the models presented here have cluster resampling off and partial context decay on. Models without semantic congruence parameters overpredict semantic distances between neighboring thoughts (upper panel). Models without decision-congruence parameters underpredict the probability of sampling a thought supporting the same option as the previous thought (lower panel). Error bars indicate standard errors across participants. See the online article for the color version of this figure.

patterns in our decision data. Specifically, Figure 7 shows the predicted response probabilities (keep deliberating, top panel, or choosing the first option, bottom panel) as a function of the cumulative support for the first option. Here the predicted response probabilities were computed using the posterior distributions of the decision parameters, as well as observed support ratings generated by participants. We can see that the predicted probabilities from the primary decision model reflect the qualitative patterns in our data, with the probability of continuing deliberation decreasing as the magnitude of cumulative support increases, and the probability of selecting a response increasing as its support increases. Here the

primary decision model achieves a (MAD of 0.083 for predicting continuation probabilities, and a MAD of 0.050 for predicting probabilities of selecting the first option. Performance for the models that replace absolute accumulation with relative accumulation is almost the same as our primary model (MAD continuation: 0.080; MAD first option choice: 0.050). The models that force context decay to be completely off are comparatively worse (MAD continuation: 0.091; MAD first option choice: 0.057). The models that replace threshold decision making with a time limit, or force context decay to be completely on, are especially bad (MAD continuation: 0.130 and 0.128; MAD first option choice: 0.079 and 0.096), as they

are unable to capture the effect of cumulative support on the probability of sampling.

Figure 8 shows the predicted probabilities of choice as a function of the number of thoughts generated. Here our primary model simulations capture the observed data pattern, in which the probabilities of continuing deliberation decrease and the probabilities of selecting an option increase, as the number of generated thoughts increase. Across all the models, our primary model achieved the smallest MAD for predicting probabilities of continuing deliberation (0.081) and probabilities of selecting the first option (0.054). The models that switch context decay completely off perform almost equally well (MAD continuation: 0.086; MAD first option chosen: 0.054). Choice simulations of the models with relative accumulation are less sensitive to the number of thoughts (as reflected by flatter slopes in Figure 8) and result in larger MADs (MAD continuation: 0.147; MAD first option chosen: 0.079). Again, the models that replace threshold decision making with a time limit, or force context decay to be completely on, are especially bad (MAD continuation: 0.184 and 0.185; MAD first option chosen: 0.092 and 0.093).

Predicting Manipulation Effects

If the parameters of our best-fit model capture underlying memory and decision processes, then these parameters should be able to predict the effects of experimental manipulations that alter participant thoughts at the start of the experiment. Inspired by previous experimental research using such a methodology (e.g., Johnson et al., 2007; Weber et al., 2007), we attempted such manipulations in Experiments 2a and 2b. These experiments used the decision prompts from Experiment 1c ("Can money buy happiness?") and Experiment 1d ("Is nuclear power safe?"), but asked half the participants to begin deliberation by listing a thought that supports the first option ("yes") and asked the other half of the participants to begin deliberation by listing a thought that supports the second option ("no"). By using the same decision prompts from Experiments 1c and 1d, we were able to evaluate the generalizability of our model, by testing whether it can make accurate out-of-sample predictions for new data sets.

As expected, our manipulations did bias choices in favor of the prime, with participants asked to list a "yes"-supporting thought being more likely to respond with "yes," and participants asked to list a "no"-supporting thought being more likely to respond with "no." In Experiment 2a, the choice probability of "yes" is 60.7% in Condition 1, and 40.8% in Condition 2, $\chi^2(1) = 7.65$, $p = .006$. Similar priming effects can also be observed in Experiment 2b, choice probability for "yes" in condition 1: 72.3%; condition 2: 39.4%; $\chi^2(1) = 22.43$, $p = .001$.

Can our memory and decision models make out-of-sample predictions regarding the choice probability change between the two priming conditions? To test this, we used the posterior samples of parameters from Experiments 1c and 1d and predicted choice probabilities for trials in which the first thought cluster supported "yes" (Condition 1) and "no" (Condition 2). The point estimations for choice probability differences were based on 12,000 posterior predictive samples from Experiments 1c and 1d.

As shown in Figure 13, the primary memory model, the model with semantic congruence turned off, and the model with decision congruence turned off, predict choice probability differences between the two conditions in a manner that is very similar to

the observed differences. In Experiment 2a, where the observed choice probability difference is 19.9%, the three models predict choice probability differences of 24.5%, 26.5%, and 22.1%, respectively. Similarly, in Experiment 2b, where the observed choice probability difference is 32.9%, the predicted choice probability differences are 35.7%, 38.9%, and 24.6%, respectively. The model with both decision and semantic congruence turned off underpredicts the differences between conditions, with predictions of 13.7% in Experiment 2a, and predictions of 23.4% in Experiment 2b. Consistent with query theory, our modeling results illustrate two mechanisms responsible for the effect of thought priming on decision: (a) the manipulated first thought; (b) congruence effects, which reinforce the manipulation by further activating thoughts congruent with the prime.

In Figure 14 we further illustrate the congruence effects on thought sampling in Experiments 2a and 2b. Here, after excluding the first thoughts (which were exogenously manipulated), we still observe higher sampling proportions for thought clusters supporting the first response option for participants in Condition 1, and higher sampling proportions for thought clusters supporting the second response option for participants in Condition 2. We then used the parameters estimated from Experiments 1c and 1d to predict this effect. Crucially, our primary models closely capture these thought sampling biases. Overall, there is a MAD of 0.027 between the predictions of the primary model and observed thought cluster proportions. The predictions of the other memory model variants have larger MADs (decision congruence only: 0.029; semantic congruence only 0.039; no congruence effects: 0.071). Overall, these results show that our memory and decision models predict both the change in choice probabilities and the change in thought sampling probabilities, between the two priming conditions. These conclusions are unchanged if we use the $k = 2$ and $k = 4$ cluster solutions, instead of the $k = 3$ solutions, as shown in the Supplemental Materials.
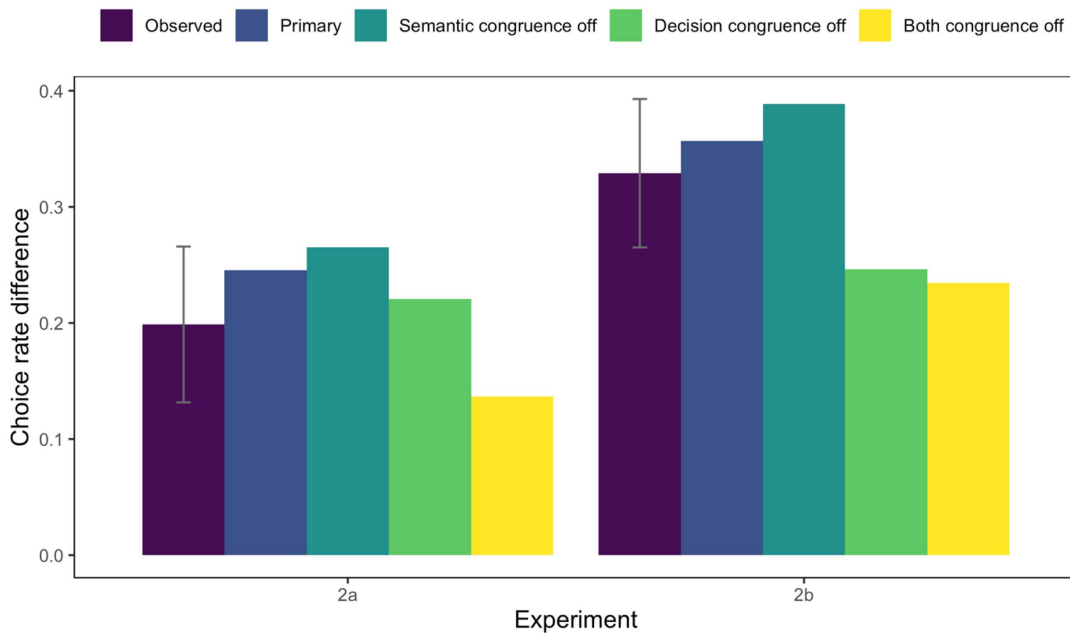
The results in Figures 13 and 14 were based on out-of-sample predictions (parameters estimated using Experiments 1c and 1d and predictions made for Experiments 2a and 2b). We obtained similar conclusions using within-sample predictions (parameters estimated using Experiments 2a and 2b). The results are shown in Figures 17S and 18S.

Discussion

A Tractable Framework

We have presented a framework for modeling naturalistic memory-based decision making, in the empirical paradigm introduced by query theory. Our approach uses computational language models to preprocess and quantify the content of thoughts retrieved from memory. These thoughts are combined into discrete clusters, and a context-dependent memory process specifies the retrieval of thoughts from these clusters. Context-dependent decision processes are similarly used to model how retrieved thoughts guide final responses.

Overall, our framework is highly general as it subsumes numerous existing memory and decision models within a single computational system. It is also highly tractable, as the parameters of these models can be retrieved from participant data. We illustrate the generality and tractability of our model by fitting 576 memory and

Figure 13

Difference in Probability of Choosing the First Option for Condition 1 Versus 2, in Experiments 2a–2b

Note. In Condition 1 (2) participants were asked to list a thought supporting the choice of the first (second) option first. In both experiments, participants in Condition 1 are more likely to indicate the first option as their final answer. The model with both semantic and decision congruence turned off cannot capture this pattern as well as the other models. All the models presented here have cluster resampling off and partial context decay on. Posterior samples of parameters from Experiments 1c–1d (where the same decision prompts are shown to participants as in Experiments 2a–2b) are used to simulate model predictions. Error bars indicate standard errors. See the online article for the color version of this figure.

decision models to participant thought retrieval and decision data from multiple experiments in domains such as risk perception, consumer behavior, financial decision making, ethical decision making, legal decision making, food choice, and well-being, society, and culture.

The models fit data in our switchboard analysis (Turner et al., 2018) include variants of the CMR, Markov random walk, query theory, relative accumulation to threshold, decision-by-sampling, leaky accumulation to threshold, lexicographic heuristic, lexicographic semiorder heuristic, tallying heuristic, and the weighted additive models. We are also able to fit hybrid models generated by combining the distinct mechanisms assumed in the above models. Likewise, we are able to pair the various memory models with the various decision models, in order to comprehensively examine the very large set of distinct memory-based decision models. In doing so, our work adds precision and a deeper understanding of cognitive mechanisms underlying behavior in naturalistic decisions elicited through the query theory experimental paradigm.

Dynamics of Memory and Decision Making

Our fits reveal that memory processes typically display semantic and decision congruence, and decision processes typically display the accumulation of absolute evidence to a threshold. Memory displays moderate context decay (rather than full or no decay), and we find mixed evidence for the role of context decay in decision making (our fits to group-level data show moderate decay, whereas
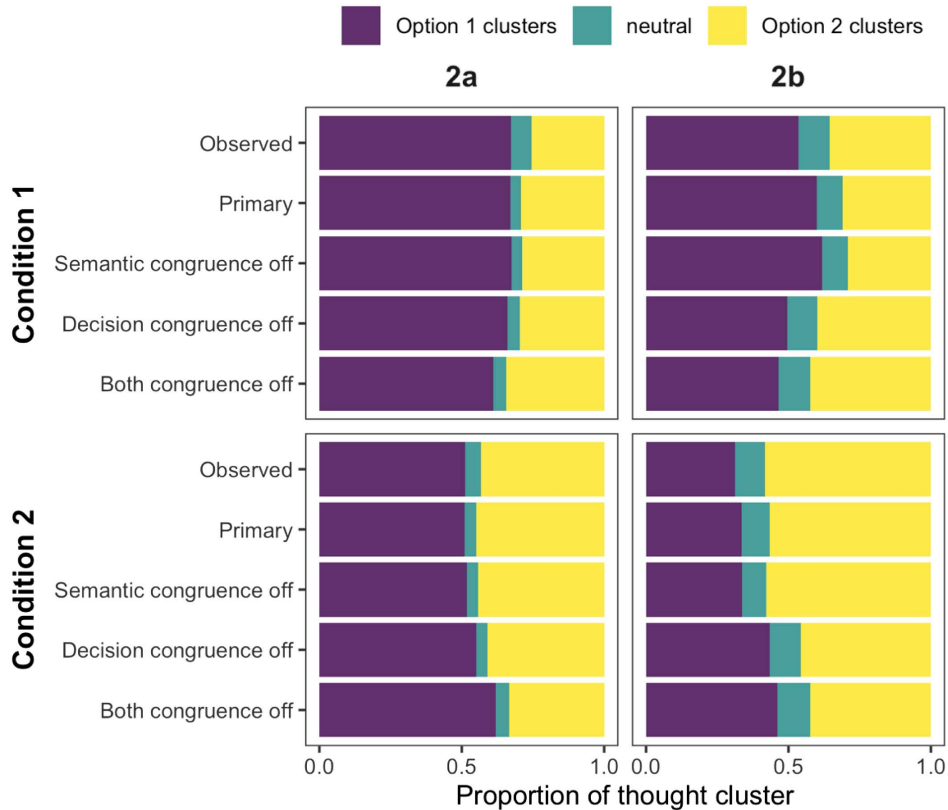
our fits to individual-level data show weak or no decay). Overall, retrieved thoughts increase the retrieval probability of other semantically related thoughts and thoughts that support the same response option, with recently retrieved thoughts playing a larger role in thought retrieval. Additionally, as thoughts are retrieved, their supports are aggregated in separate decision variables (one for each response option), which evolve dynamically until one of the variables reaches a threshold value.

Models equipped with the above mechanisms are able to quantitatively predict the likelihood of listing different thought clusters, observed thought cluster transitions, total length of deliberation (i.e., number of retrieved thoughts), and response probabilities. Such models are also necessary to account for core qualitative patterns in these memory and decision variables, such as the relationship between choice probability and the set of sampled thoughts, as well as the effect of sampled thoughts on subsequent thoughts. Finally, models with the above mechanisms are able to quantitatively predict the effect of experimental manipulations on thought listing and choice behavior. Thus, in line with human data, these models predict that priming a thought that supports one response option increases the likelihood of sampling other thoughts supporting that option (due to decision congruence), and in turn, increases the choice probability of that option. Some of the above patterns have been documented in memory research and decision research, but we are the first to quantitatively model the emergence of these patterns in memory-based decisions, and use these patterns to evaluate a large number of distinct models proposed in prior work.

Figure 14

Distributions of Thought Clusters Supporting Different Options in Experiments 2a and 2b, After Excluding the First Thought

Note. We observe higher sampling proportions for thought clusters supporting first option for participants in Condition 1 (start with a thought supporting the first option; top panel). The opposite is true for participants in Condition 2 (start with a thought supporting the second option; bottom panel). This pattern is best captured by models with decision congruence on. Note that the primary model has both semantic and decision congruence turned on. All the models presented here have cluster resampling off and partial context decay on. Posterior samples of parameters from Experiments 1c–1d (where the same decision prompts are shown to participants as in Experiments 2a–2b) are used to simulate model predictions. See the online article for the color version of this figure.

Note that these successes depend on the joint modeling of both memory and decision making. In particular, by specifying a full memory-based decision model, we are able to use the model to simulate human-like decisions without relying on experimental data like thought support ratings. At each time point, we can use the memory model to generate a thought, and use (discrete/binary) ratings of the thought cluster to update the decision accumulators, allowing us to predict, out of sample, the behavior of participants on the group level (as we do in Experiments 2a and 2b).

Conditional Independence and Thought Vectors as Useful Modeling Tools

Our approach can accommodate nuanced memory search and decision dynamics, as memory and decision processes are assumed to interact only via context. Thus, these processes are conditionally independent on context, and can be fit separately to thought listing or decision making data. Of course, this assumption can also be applied to other types of process data, such as eye-tracking or mouse-tracking

data (see e.g., Schulte-Mecklenbeck et al., 2011). As with the memory search task modeled here, these kinds of data involve the sequential sampling of discrete pieces of information, which guide the formation of beliefs and preferences. If we assume that the processes governing attentional dynamics in eye-tracking and mouse-tracking tasks interact with decision making through context (and only through context), we can formulate and test joint models of attention and decision making for which attention and decision processes are conditionally independent on context. These models could involve complex search dynamics as well as complex decision rules, but would be tractable given the conditional independence property.[11]

---

[11] Intuitively, such models would retain the decision components of our approach, but would replace the memory components with assumptions better suited to modeling attentional dynamics. In such implementations, context would keep track of the regions of the display (rather than individual thoughts) that have been previously sampled. Likewise, eye-movement transitions could be sensitive to physical proximity rather than semantic congruence, and baseline activation biases could be due to item positioning on the display.

Another crucial component of our modeling framework is our ability to quantify thought content using sentence vectors (Cer et al., 2018), and more generally, to model the types of knowledge representations that influence decisions. It is clear that a complete model of decision making needs access to what people know about judgment and decision targets, and thus needs to be equipped with realistic knowledge representations about the world. In recent work, we have argued that such representations can be approximated using vector semantic models trained on large-scale natural language data (Bhatia & Walasek, 2019; Bhatia, 2019; Bhatia et al., 2019; Richie et al., 2019). In this article we further illustrate the value of semantic vectors by showing how they can be used to model the content of thoughts during naturalistic decision making. We are especially excited about novel technical advances in this area that facilitate the development of more sophisticated models of judgment and decision making, that are equipped with both human-like knowledge representations as well as human-like cognitive processes for aggregating knowledge representations to form beliefs and preferences.

Further Work on Complex Mental Processes

For the sake of tractability, we have made simplifying assumptions about the representation or content of thoughts, as well as the memory search processes that operate over them. For example, inspired by the context and maintenance retrieval model (Polyn et al., 2009) in list recall research, we have assumed that limits in working memory capacity derive from decay in context. Of course, this is not the only way to implement limits in working memory. SAM (Raaijmakers & Shiffrin, 1981), for example, assumes that working memory is limited due to probabilistic deletion of older items. Interference-based mechanisms are also possible and may actually best explain capacity limits in certain settings (Oberauer et al., 2016). We expect that our framework can be modified to accommodate such alternative implementations of working memory capacity limits.[12] It may ultimately be that such alternative mechanisms better account for capacity limits in the present kinds of memory-based decision making.

Moreover, we have assumed that every thought belonged to one of a handful of discrete clusters—even though this discards some of the power of distributed vector representations—and that thoughts influenced search of other thoughts only through cluster resampling, decision congruence, and semantic congruence. Although thoughts for a given question may tend to revolve around a limited number of themes or topics represented by our clusters, in reality, thoughts—and the sentences we use to communicate them—do not belong to a finite number of clusters. Rather, there is an infinite number of thoughts and sentences we can express, and moreover, there is a discrete, compositional structure to these thoughts and sentences (Fodor, 1975; Fodor & Pylyshyn, 1988), and it is this structure which licenses more complex forms of (logical) reasoning than those modeled here. Finally, thought vectors are likely to vary from subject to subject, something that we cannot address in the article (even with our fits to individual-level data). This is, of course, a limitation of all distributional semantics models, which are unable to capture heterogeneity in participant representations. Accommodating participant heterogeneity and more complex forms of thinking and reasoning using quantitative representations is an important topic for future work.

It is also possible that the act of listing thoughts (as in our experiments) alters deliberation processes, so that our models do not capture memory-based decision making without the thought listing protocol. Fortunately, previous work on the query theory paradigm has found that eliciting thoughts does not alter choice frequencies (see e.g., Johnson et al., 2007), though future work could attempt to rigorously test this using the modeling framework introduced in this article. It is also possible that decisions rely on information not revealed during thought listing. For example, people may have certain strong (negative) affective associations with nuclear power that they are not entirely aware of and thus do not explicitly list in their thoughts. This is certainly possible and perhaps even likely, and it should be possible for at least some of these sorts of biases to show up in future decision models as a starting point effect, which we did not account for here. Of course, it is possible that participants do explicitly retrieve and evaluate certain decision-relevant thoughts during deliberation, but choose to omit these in their listed thoughts for various reasons, including social desirability. However, we suspect that few, if any, of our questions are so controversial that subjects feel pressure to censor themselves.

The decision prompt or type of decision being made may also alter certain properties of memory search and decision making. In particular, our second and fifth questions—which entail choices between gift certificate purchases and dinner options, respectively—differ from the remaining questions on the memory side. These two questions have a more positive cluster resampling parameter than other questions, and have positive semantic congruence parameters where other questions have negative semantic congruence parameters (see Figure 10). Interestingly, when we fit the hierarchical models to the data collected in Experiment 3, we replicated these results (see Figure 10S in the Supplemental Materials), indicating that these patterns are not spurious, but rather systematic. We suspect that these two questions are different because of the nature of their decision domain, as they are the only two questions on preferential decision making. We are unsure why preferential decision making would involve these differences, and therefore suggest that this is a topic for future work (which could, for example, systematically compare model parameters across many variants of the questions used in the current article). More generally, Experiment 3 in the Supplemental Materials largely replicated the results of Experiment 1, with the exception of decision decay: Our group-level fits support moderate decision decay, but decision decay is close to zero in our hierarchical fits. We speculate on the source of this discrepancy in the Supplemental Materials, but better understanding this difference in the group-level and hierarchical fits could be a direction for future work.

There are also certain processes of memory search in word recall and semantic fluency tasks that we have not extended to the present setting, which may still be applicable. Perhaps chief among these are optimal foraging processes (Hills et al., 2012; see also Abbott et al., 2015) which have been shown to operate in semantic memory search. A similar process may operate when searching for thoughts

---

[12] Although certain implementations of capacity limits, such as probabilistic forgetting, may make the likelihood function no longer analytically tractable, meaning that the Hamiltonian Monte Carlo algorithm used to fit our models no longer apply, and other Bayesian computation methods may be necessary (e.g., Turner et al., 2016).

relevant to the present kinds of naturalistic decisions, and future work should attempt to model optimal foraging in the context of memory-based decision making. Additionally, memory could also involve more superficial verbatim representations of information (in addition to the more abstract gist representations we have attempted to model), which could be integrated into the current framework by adapting some of the insights of fuzzy-trace theory (see e.g., Reyna, 2012 for a discussion).

There may also be additional processes operating on the decision side. For example, decision makers may decrease their decision threshold as they sample more thoughts (often captured by collapsing boundaries in the sequential sampling model literature, e.g., Hawkins et al., 2015). They may adopt a hybrid decision rule that combines threshold-based and time limit elements (e.g., Reutskaja et al., 2011), or make choice with nonnormally distributed noise (e.g., Voss et al., 2019). Additionally, a fuller understanding of the priming manipulation in Experiment 2 may require including a flexible starting point parameter in the decision model (e.g., Mulder et al., 2012).[13] Given the complexity of these decision mechanisms, and possible heterogeneity across individuals, we leave it to future studies to design experiments and expand our modeling framework to better capture the nuances in the decision process. We suspect asking subjects to complete multiple primed questions may lead to an unnatural decision setting, and that multiple primes may not have persistent effects.

We have also assumed that memory and decision processes do not directly interact. Thus, even though retrieved thoughts influence decisions, they do so only via the support ratings (which we are able to observe through our experimental paradigm). Conversely, the state of the decision context variables (e.g., accumulated evidence favoring the two response options) does not influence thought retrieval. By separating memory and decision processes in this manner we have gained considerable tractability, yet evidence in reasoning and decision making research suggests that accumulated preferences can have a direct influence on thought activation and retrieval. Specifically, decision makers have been shown to reason through coherence-maximizing processes that result in thought activation that is consistent with the evolving decision variables (Bhatia, 2016; Glöckner & Betsch, 2008; Holyoak & Simon, 1999; Simon et al., 2004). Such mechanisms could also be a product of goal-directed deliberation processes. Despite these computational concerns, such a specification is worth testing in future work, as coherence-based reasoning and goal-based decision making is a crucial component of naturalistic decision processes.

Perhaps many of the extensions of our approach, from accounting for additional outcome and process measures like response times or eye- and mouse-tracking data, to adding additional memory and decision processes like logical reasoning, could be implemented by joining our models with mechanisms implemented in ACT-R's modules (Anderson, 2007; Anderson et al., 2004; Dimov et al., 2020; Marewski & Mehlhorn, 2011). One advantage of this approach would be that additional free parameters might not be necessary, and instead one could use ACT-R's default parameters that have been validated across many paradigms and tasks of varying complexity. Given the architectural similarities between our frameworks—for example, context in our framework is akin to the module buffers in ACT-R—we believe such a union of approaches would be both natural and fruitful.

Ultimately, memory-based decision making is likely to contain a mix of complex thinking, reasoning, and decision processes. Some will likely be those we have modeled, like similarity and decision congruence in memory and absolute accumulation to threshold in decision making, and some we have not, like logical reasoning, optimal search, and coherence-based reasoning in memory, and starting point biases in decision making. Although our framework is by no means complete, we believe it offers a valuable statistical tool for describing the underlying dynamics of many of these more complex thought processes. By vectorizing thoughts to quantify their content, and by placing these thoughts into a context representation that guides decisions, we allow for a model capable of displaying complex transitions between thoughts, as well as predicting choice probabilities and lengths of deliberation. Attempting to integrate more complex forms of memory and decision making into such a model will likely drive our research for many years to come.

---

[13] Currently such a starting point effect is hard to empirically distinguish from the decision congruence effect of our memory model, and cannot be parametrically recovered without many repetitions of the same question for an individual.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558–569. https://doi.org/10.1037/a0038693

Alba, J. W., Hutchinson, J. W., & Lynch, J. G., Jr. (1991). Memory and decision making. In H. H. Kassarjian & T. S. Robertson (Eds.), *Handbook of consumer theory and research* (pp. 1–49). Prentice-Hall.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195324259.001.0001

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*(4), 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*(3), 522–543. https://doi.org/10.1037/a0032457

Bhatia, S. (2016). The dynamics of bidirectional thought. *Thinking & Reasoning*, *22*(4), 397–442. https://doi.org/10.1080/13546783.2016.1187205

Bhatia, S. (2019). Semantic processes in preferential decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 627–640. https://doi.org/10.1037/xlm0000618

Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, *29*, 31–36. https://doi.org/10.1016/j.cobeha.2019.01.020

Bhatia, S., & Walasek, L. (2019). Association and response accuracy in the wild. *Memory & Cognition*, *47*(2), 292–298. https://doi.org/10.3758/s13421-018-0869-6

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences*, *23*(3), 251–263. https://doi.org/10.1016/j.tics.2018.12.003

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*(2), 91–134. https://doi.org/10.1016/0022-2496(88)90042-9

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dimov, C., Khader, P. H., Marewski, J. N., & Pachur, T. (2020). How to model the neurocognitive dynamics of decision making: A methodological primer with ACT-R. *Behavior Research Methods*, *52*(2), 857–880. https://doi.org/10.3758/s13428-019-01286-2

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. https://doi.org/10.1037/0033-295X.106.1.180

Fechner, H. B., Pachur, T., Schooler, L. J., Mehlhorn, K., Battal, C., Volz, K. G., & Borst, J. P. (2016). Strategies for memory-based decision making: Modeling behavioral and neural signatures within a cognitive architecture. *Cognition*, *157*, 77–99. https://doi.org/10.1016/j.cognition.2016.08.011

Fechner, H. B., Schooler, L. J., & Pachur, T. (2018). Cognitive costs of decision-making strategies: A resource demand decomposition analysis with a cognitive architecture. *Cognition*, *170*, 102–122. https://doi.org/10.1016/j.cognition.2017.09.003

Fishburn, P. C. (1967). Methods of estimating additive utilities. *Management Science*, *13*(7), 435–453. https://doi.org/10.1287/mnsc.13.7.435

Fodor, J. A. (1975). *The language of thought*. Thomas Y. Crowell.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Glöckner, A., & Betsch, T. (2008). Modeling Option and Strategy Choices with Connectionist Networks: Towards an Integrative Model of Automatic and Deliberate Decision Making. *Judgment and Decision Making*, *3*(3), 215–228. https://doaj.org/article/f04d76a54930444e90cd6823a26cbf9f

Gluth, S., Rieskamp, J., & Büchel, C. (2012). Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(31), 10686–10698. https://doi.org/10.1523/JNEUROSCI.0727-12.2012

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635. https://doi.org/10.1207/s15516709cog2704_2

Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010). A dirty word or a dirty world?: Attribute framing, political affiliation, and query theory. *Psychological Science*, *21*(1), 86–92. https://doi.org/10.1177/0956797609355572

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(6), 2476–2484. https://doi.org/10.1523/JNEUROSCI.2410-14.2015

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1191–1206. https://doi.org/10.1037/a0013025

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440. https://doi.org/10.1037/a0027373

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*(3), 341–358. https://doi.org/10.1037/0033-295X.94.3.341

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3–31. https://doi.org/10.1037/0096-3445.128.1.3

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 461–474. https://doi.org/10.1037/0278-7393.33.3.461

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Jones, M. N., Willits, J. A., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer & J. T. Townsend (Eds.), *Oxford handbook of mathematical and computational psychology* (pp. 232–254). Oxford University Press.

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PRO-BEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*(5), 563–607. https://doi.org/10.1207/s15516709cog2605_2

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, *71*(1), 107–138. https://doi.org/10.1146/annurev-psych-010418-103358

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174084

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. https://doi.org/10.3758/BF03196581

Lee, M. D., Gluck, K. A., & Walsh, M. M. (2019). Understanding the complexity of simple decisions: Modeling multiple behaviors and switching strategies. *Decision*, *6*(4), 335–368. https://doi.org/10.1037/dec0000105

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Lieder, F., & Griffiths, T. L. (2020). Advancing rational analysis to the algorithmic level. *Behavioral and Brain Sciences*, *43*, 1–60. https://doi.org/10.1017/S0140525X19002012

Link, D., & Marewski, J. N. (2015). Populating ACT-R's declarative memory with internet statistics. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.) *Proceedings of the 13th International Conference on Cognitive Modeling (ICCM)* (pp. 69–70). University of Groningen.

Link, D., Marewski, J. N., & Schooler, L. J. (2016). An ecological model of memory and inferences. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 1883–1888). Cognitive Science Society.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marewski, J. N., & Mehlhorn, K. (2011). Using the ACT-R architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making*, *6*(6), 439–519. http://journal.sjdm.org/11/101112/jdm101112.pdf

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*(3), 393–437. https://doi.org/10.1037/a0024143

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 3111–3119). Curran Associates Inc.

Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review*, *125*(4), 512–544. https://doi.org/10.1037/rev0000102

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, *66*(1), 277–294. https://doi.org/10.1146/annurev-psych-010814-015148

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-Learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Pennington, J., Socher, R., & Manning, C. (2014, October). Glove: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*. https://doi.org/10.18653/v1/N18-1202

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, *101*(2), 900–926. https://doi.org/10.1257/aer.101.2.900

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making*, *7*(3), 332–359. http://journal.sjdm.org/11/111031/jdm111031.pdf

Richie, R., Zou, W., & Bhatia, S. (2019). Predicting high-level human judgment across diverse behavioral domains. *Collabra. Psychology*, *5*(1), Article 50. https://doi.org/10.1525/collabra.282

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*(2), 370–392. https://doi.org/10.1037/0033-295X.108.2.370

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 676–696. https://doi.org/10.1037/0278-7393.9.4.676

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*(3), 610–628. https://doi.org/10.1037/0033-295X.112.3.610

Schulte-Mecklenbeck, M., Kühberger, A., & Johnson, J. G. (Eds.). (2011). *A handbook of process tracing methods for decision research: A critical review and user's guide*. Psychology Press. https://doi.org/10.4324/9780203875292

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, *90*(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*(5), 331–336. https://doi.org/10.1111/j.0956-7976.2004.00678.x

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Stan Development Team. (2020). *RStan: the R interface to Stan* (R package version 2.21.2). http://mc-stan.org/

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Suri, G., Gross, J. J., & McClelland, J. L. (2020). Value-based decision making: An interactive activation perspective. *Psychological Review*, *127*(2), 153–185. https://doi.org/10.1037/rev0000164

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*(1), 155–185. https://doi.org/10.1037/0033-295X.115.1.155

Trueblood, J. S., Heathcote, A., Evans, N. J., & Holmes, W. R. (2021). Urgency, leakage, and the relative nature of information processing in

decision-making. *Psychological Review*, *128*(1), 160–186. https://doi.org/10.1037/rev0000255

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*, Article 79. https://doi.org/10.3389/fnins.2012.00079

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329–362. https://doi.org/10.1037/rev0000089

Turner, B. M., Sederberg, P. B., & McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, *72*, 191–199. https://doi.org/10.1016/j.jmp.2014.10.001

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*(1), 31–48. https://doi.org/10.1037/h0026750

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769. https://doi.org/10.1037/0033-295X.111.3.757

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*(1), 73–96. https://doi.org/10.1037/0096-3445.137.1.73

Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin & Review*, *26*(3), 813–832. https://doi.org/10.3758/s13423-018-1560-4

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, *11*, 3571–3594. https://dl.acm.org/doi/10.5555/1756006.1953045

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*(1), 53–85. https://doi.org/10.1146/annurev.psych.60.110707.163633

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, *18*(6), 516–523. https://doi.org/10.1111/j.1467-9280.2007.01932.x

Zhou, X., Wong-Lin, K., & Philip, H. (2009). Time-varying perturbations can distinguish among integrate-to-threshold models for perceptual decision making in reaction time tasks. *Neural Computation*, *21*(8), 2336–2362. https://doi.org/10.1162/neco.2009.07-08-817

Appendix A

Experiment Instructions

In this study, we are interested in what people think about when making their decisions. In the subsequent screens you will be presented with {Exp1: two; Exp2: one Exp3: eight} question{s} which present two options (e.g., yes or no, or cake or pie). For each question you will:

1. List the thoughts that come to your mind as you decide your answer to the question. Please report your thoughts as complete sentences, in English. Your listed thoughts should be understandable by a third party. You can list thoughts supporting both choices (yes or no, cake or pie), regardless of your eventual choice. For example, if the question was "Are school uniforms a good idea?," you could write "Self-expression is important but limited by school uniforms" and/or "Uniforms can prevent bullying of less rich kids who can not afford as nice clothes." You would not write just "student expression" or "bullying," or just "yes" or "no." NOTE: The top 50% of participants

with the clearest thoughts will be given a 50% bonus (e.g., a $2 study would pay a bonus of $1).

2. Provide your answer to the question. The questions you will be asked were designed to be subjective and potentially difficult to answer. You may not feel that your answer easily fits into either choice, but please nevertheless select one answer.

3. For each thought you listed in (1), indicate which choice (e.g., yes/no, cake/pie) the thought supported.

At the end you will be asked a couple of questions about your background.

Thank you for participating! When you are ready, click next. A blank page will momentarily appear, and then the first thought listing page will appear.

Appendix B

Model Fitting Details

The decision and memory models were fit to data using RStan (Stan Development Team, 2020). We generated four chains of 3,500 samples, where the first 500 samples of each chain were burn-ins. The results were based on the combined 12,000 samples for each model. Overall, we used weakly informative prior distributions for the parameters. The prior distributions for the baseline activation strengths of the clusters (γ) and the scalar weights in the memory models (ω_R, ω_S, and ω_D) were Normal(0, 2^2) and Normal(0, 5^2),

respectively. To avoid nonidentifiability, we constrained the activation strength of the last cluster to be 0 in all the memory models.

On the decision side, the prior distribution used for the threshold value (τ) was Half-Normal(0, 35^2). Instead of fitting the standard deviation parameter (σ) of the normally distributed error at time $t(\varepsilon_t)$ directly, we fit σ/τ in the model. The prior for this parameter is Half-Normal(0, 0.5^2). In the time limit models, the probability of continuing deliberation (λ) is the same after each thought cluster

sample, and thus the deliberation time T follows a Geometric distribution. The prior distribution used for λ was Uniform(0, 1). The prior distribution used for the standard deviation parameter (σ) of the normally distributed error at time $t(\varepsilon_t)$ is Half-Normal(0, 35^2). In both the memory and the decision models, the prior distribution for the decay parameters (δ_M, δ_D) was Uniform(0, 1).

To assess model convergence, $\hat{R}$ parameters were calculated for each model. Among all the parameters of all the models, the largest $\hat{R}$ was 1.010, indicating successful convergence of the chains. The smallest number of effective sample size was 1,082. In the article, when reporting results regarding the memory models, we use the $k = 3$ k-means thought clustering solutions (i.e., seven distinct thought clusters) for each experiment. Results for $k = 2$ and $k = 4$ k-means clustering are provided in the Supplemental Materials).