Encoding and Inhibition of Arbitrary Episodic Context with Abstract Concepts

Charles P. Davis,[1,2] Pedro M. Paz-Alonso,[3] Gerry T. M. Altmann,[1,2] & Eiling Yee[1,2]

[1] Department of Psychological Sciences, University of Connecticut

[2] CT Institute for the Brain and Cognitive Sciences, University of Connecticut

[3] BCBL. Basque Center on Cognition, Brain and Language

Abstract

Context is critical for conceptual processing, but a mechanism underpinning its encoding and reinstantiation with concepts is unclear. Context may be especially important for abstract concepts—we investigated whether *episodic* context is recruited differently when processing abstract compared to concrete concepts. Experiments 1 and 2 presented abstract and concrete words in arbitrary contexts at encoding (red/green colored frames in Experiment 1; male/female voices in Experiment 2). Context recognition was worse for abstract concepts. Again using frame color and voice as arbitrary contexts, respectively, Experiments 3 and 4 presented words from encoding in the same or different context at test to determine whether there was a greater recognition memory benefit for abstract versus concrete concepts when the context was unchanged between encoding and test. There was instead a *disadvantage* for abstract concepts: they were *less* likely to be identified when context was retained. These findings suggest that episodic detail, when arbitrary, is attended less, and may even be inhibited, when processing abstract concepts. In Experiment 5, we utilized a context—spatial location—which (as we show) tends to be relevant during real-world processing of abstract concepts. We presented words in different locations, preserving or changing location at test. Location retention conferred a recognition advantage for abstract concepts. Thus, episodic context may be encoded with abstract concepts when context is relevant to real-world processing. The systematic contexts necessary for understanding abstract concepts may lead to arbitrary context inhibition, but greater attention to contexts that tend to be more relevant during real-world processing.

Keywords: concepts, semantic memory, episodic memory, abstract concepts, concreteness

Encoding and Inhibition of Arbitrary Episodic Context with Abstract Concepts

So-called “abstract” concepts[1] like *decision* are central to the human experience, yet relatively little is understood about how they are processed. Contextual information is important for understanding all concepts (e.g., Yee & Thompson-Schill, 2016), but particularly important for more abstract concepts (e.g., Barsalou & Wiemer-Hastings, 2005; Schwanenflugel, 1991). For example, while a *river* in New England shares many properties with a river in Papua New Guinea, consider the case of *decision*: your decision on which beverage to buy at a café late at night differs greatly from the decision a judge might make in determining sentencing for a felon. It is the context which determines the antecedents, outcomes, and consequences in these two instantiations of *decision*. Thus, the specific meaning of *decision* varies more depending on context than does the meaning of *river*. Here, we investigate how a particular type of context, *episodic* context, is remembered in the presence of abstract and concrete concepts.

Much work on abstract concepts has focused on their relation to different types of contextual information. In free-association style tasks, abstract concepts tend to elicit fewer object-property-related associations (e.g., is colorful) and more situation-related associations (e.g., something to talk about; Barsalou & Wiemer-Hastings, 2005; see also Crutch & Warrington, 2005). This difference is likely because the components of abstract concepts are distributed over multiple aspects of events, or multiple events, across space and time (Barsalou, 1999; see also Barsalou et al., 2018; Binder et al., 2016; Davis et al., 2020, for discussion). Moreover, abstract concepts less reliably activate particular semantic contexts, and therefore rely

[1] While we use the terms “abstract and “concrete” concepts, we do not intend to imply that there is a clear dichotomy between abstract and concrete concepts. Rather, we use these terms for succinctness, as shorthand for “more abstract” and “more concrete” concepts. In fact, although, the concepts we test were rated as highly abstract and highly concrete, a multitude of factors contribute to relative “abstractness” or “concreteness.” We discuss this further in the General Discussion (see also Davis et al., 2020).

more on currently available sentential contexts for understanding. That is, the more abstract the concept, the more difficult it is to spontaneously think of a context or circumstance in which it could occur (see context availability theory; e.g., Schwanenflugel & Shoben, 1983; Schwanenflugel, 1991).

A reason that our understanding of abstract concepts may rely particularly heavily on their current contexts is that they tend to be more semantically diverse. That is, they can occur in many semantically distinct contexts (e.g., an *idea* to take up a career creating balloon animals and an *idea* to drink another coffee, whereas *pencil* would occur in a more circumscribed range of contexts; Hoffman et al., 2013). A resulting need to select the appropriate meaning of the concept given the context may explain why abstract concepts rely more than concrete concepts do on brain regions involved in semantic control, that is, on brain regions that help select the appropriate meaning of a concept given the context (e.g., Hoffman et al., 2015; for semantic control, see e.g., Badre & Wagner, 2005; Thompson-Schill et al., 1998; Thompson-Schill, 2003). Related, abstract concepts are more reliant on semantic knowledge of *situations* (i.e. schema knowledge; Bartlett, 1932) for their recognition than are concrete concepts (Davis et al., 2020; see discussion below).

In sum, the extant evidence suggests that, on account of their distributed and diverse nature, abstract concepts rely heavily on readily available semantic context for their processing. However, the mechanism by which context is encoded and re-instantiated with the concept remains unclear. In this work, we test the hypothesis that the episodic memory system (specifically, episodic context) is differentially recruited in processing abstract vs. concrete concepts.

Episodic memory is classically defined as explicit memory for unique events (Tulving, 1983, 2002). We take the *episodic context* in which an event occurs to be the objects and their relations that co-occur in contiguous space and time with the participants in the event, but which are not a part of the event itself. They form the contemporaneous context in which the event is grounded and make that event unique. Sitting in a chair is just the same as sitting in a(nother) chair, unless there are specific details which differentiate these events of sitting. These details could be *arbitrary* (e.g., the color of the wall behind the chair) or they could be *systematically related* to the event or its participants (e.g., the configuration of objects on the dining room wall being predictive of the chairs and of events such as sitting down and eating).

The encoding of these details, as a part of the episodic experience, relies on *relational binding* (Cohen & Eichenbaum, 1993)—the indiscriminate association of elements in a scene (whether part of an event or not) with other elements in the scene (see Altmann & Ekves, 2019, for an account of event representation which relies on relational binding *across time*). Relational binding is "blind" to which of these associations are arbitrary and which are systematic (as occurs when one element is predictive of another element). However, the systematic associations will likely also be encoded within *semantic memory*, that is, long-term experiential knowledge corresponding to concepts and *schema* (knowledge of situations and the typical events that may accompany each situation; Bartlett, 1932). This dual encoding of systematic associations—encoded both in semantic memory and in the relational binding of the participants in an event to their episodic context (*relational memory*)—will prove key to understanding how non-systematic, arbitrary associations impact on memory for abstract versus concrete concepts.

In tasks probing the episodic memory system, the episodic context is often *operationally defined* as some aspect of a percept or situation that is irrelevant to the central stimulus—for example, whether a test word is presented in red or green font or whether a line drawing is presented within a red or green frame (for discussion, see Migo et al., 2012). What factors influence the likelihood that one will encode and subsequently recall these irrelevant episodic details when prompted? Given that the role of context in comprehending abstract concepts is pervasive (e.g., Schwanenflugel & Shoben, 1983), we contend that one such factor is *abstractness*.

The most straightforward hypothesis is that episodic context generally (i.e., *any* type of episodic context), is more important for interpreting abstract (relative to concrete) concepts. Under this view, we should be more accurate at retrieving the arbitrary elements of the episodes that ground abstract concepts in particular contexts. That is, relational memory might be better for abstract than concrete concepts. Consider, for example, the difference between a typical instance of a chair, which is a chair regardless of the context in which it is experienced, and a typical instance of a decision. The context matters—whether advice was sought, dice were thrown, or whether the decision was to buy a house or a coffee. In these cases, the nature of the decision depends on the context in ways that a chair does not, hence the possibility that context matters more (and hence is more likely to be encoded) for decisions than for chairs. Or rather, that relational memory is engaged more for abstract that concrete concepts.

An alternative hypothesis is that the *type* of episodic context may influence the degree to which that context is encoded with an abstract concept. Elements that constitute experience of an abstract concept vary considerably across instantiations (i.e., they are less *situationally systematic*; Davis et al., 2020). Consider the differences between *decision* at the local café versus decision in the context of sentencing decisions in the justice system. Understanding such experiences may demand attention specifically to *systematic* elements of the context—for instance, understanding the meaning of *decision* in the courtroom requires tracking evidence, consequences, demeanor, and other characteristics related to the crime and alleged perpetrator. It requires semantic knowledge pertaining to *situations* and the likely participants and events that may accompany a situation. A decision in the justice system requires a broader set of schema than a decision in a café. van Kesteren et al. (2013) propose, based on neurobiological evidence, that the more schema knowledge is activated, the more relational memory is inhibited, in turn leading to *inhibition* of more arbitrary elements of the context, such as the color of the walls in the courtroom (for discussion, see Davis et al., 2020). Under this hypothesis, if processing abstract concepts entails activation of the sorts of systematic contexts typically necessary for comprehension (e.g., via activating schema or enhancing any systematic details that are co-present in the context), *memory for arbitrary elements of the context may be worse for abstract than concrete concepts.*

We opted to test these competing hypotheses by examining whether arbitrary contexts are differentially recognized when paired with abstract as compared to concrete concepts At stake is the role of relational memory when recognizing abstract and concrete concepts. A standard paradigm for assessing whether we encode arbitrary contents of a particular episode (e.g., the identity of a speaker) is the source memory task (see Davachi, 2006; Johnson et al., 1993; Yonelinas, 2001, 2002). In this task, participants are asked at a test phase to determine whether an item (e.g., a word) was previously presented in an exposure phase, and then are probed as to whether they can recognize some contextual detail that was present at encoding (e.g., the color of a frame that surrounded the word). In the studies below, context is operationally defined as an aspect of an episode (i.e., trial) that is irrelevant to the processing of the target stimulus embedded within that context, such as whether a target word is presented within a red or green frame, whether stimuli are presented in a male or female voice, or the quadrant of a screen in which the target words are presented. Important here is that arbitrary and irrelevant are not used interchangeably—a context may be arbitrary in its relation to word processing in the context of the experiment, yet typically relevant in the real-world processing of concepts. For instance, while the color of the surroundings is both arbitrary *and* irrelevant (the color of a wall is irrelevant when considering the meaning of *decision*), the spatial location of concepts in an experiment may be arbitrary despite tending to be relevant in real-world processing (whether *decision* is experienced in a casino or coffeehouse).

We expected that if relational binding (i.e., the binding of any relationship, systematic or arbitrary; Cohen & Eichenbaum, 1993) is stronger for abstract than concrete concepts, then even arbitrary contexts should be better encoded with abstract than concrete concepts.[2] On the other hand, if the lack of situational systematicity inherent to more abstract concepts indeed results in inhibition of arbitrary contexts, we would anticipate worse recognition of the arbitrary context in abstract concepts. Regardless of the direction of the effect, any effect of abstractness on source memory task performance would add to the evidence that semantic memory and episodic memory are integrated. Specifically, it would indicate that recognition of episodic context can be influenced by a semantic dimension (here, abstractness).

Experiment 1

In Experiment 1, we examined whether memory for an episodic context that is both irrelevant for processing word meaning *and* arbitrary (the color of a frame surrounding a word) is affected by whether the word that it is paired with is abstract or concrete. We used a source

[2] Because memory is generally better for concrete than it is for abstract words (e.g., Paivio et al., 1994), perhaps due to their imageability, we expected that overall memory for concrete concepts would be better, but that *when* abstract concepts are correctly recognized, memory for the contextual detail would be better.

memory task where after being exposed to a list of words presented individually in colored frames, participants were asked to judge whether a word had been present in the exposure phase, and if it had, to retrieve the color of the frame that surrounded it at encoding. As noted above, if relational binding is stronger for abstract than concrete concepts, then when abstract concepts are correctly recognized, the context should be better encoded. Alternatively, if the overall lack of situational systematicity inherent to more abstract concepts results in inhibition of arbitrary contexts, we should observe worse recognition of the arbitrary context in abstract concepts.

Less critically, because there is a well-established association between high confidence in having seen an item and greater likelihood of encoding the context in which that item was placed (e.g., Kirwan et al., 2008; Yu et al., 2012; see Rugg et al., 2012, for review), we sought to ensure that our procedure was working as it has in prior studies by asking participants to indicate their confidence in recognizing the word and frame. Here, we predicted that confidence in having seen the *word* will be associated with the likelihood of encoding the *context*.

Methods

Participants. We conducted a power analysis based on a pilot experiment (nearly identical in procedure to Experiment 1) of 40 participants. Based on an observed small-to-medium effect ($\eta_p^2 = .07$) and desired power = 0.90, 37 participants were required for this within-subjects design. Thus, we targeted 40–42 participants per experiment to account for possible attrition. In Experiment 1, 42 University of Connecticut (UConn) students (14 men, 28 women, mean age = 19.5 years) with normal or corrected-to-normal vision and hearing provided informed consent and received course credit for participating. Color-blind participants were ineligible for the study. There were no effects of demographic variables (age, gender) on any of our dependent measures. Two participants were excluded for non-compliance (i.e., pressing the same button on every trial), leaving *N* = 40. The study was approved by the UConn Institutional Review Board.

Stimuli. In the encoding phase, 100 (60 target, 40 non-target) abstract (e.g., *decision*) and 100 (60 target, 40 non-target) concrete (e.g., *chair*) noun concepts were used. Targets were non-synonyms. Non-targets were synonym words which functioned as positive responses for the synonym-judgment task described below. Stimuli were matched across all stimulus subsets on word length and word frequency (Brysbaert & New, 2009), and were sorted into abstract and concrete conditions based on Brysbaert et al.'s (2014) concreteness norms (Table 1). For each subject, half of the words were enclosed in red frames, and the other half in green, and this was balanced across concrete and abstract words, as well as between targets and non-targets. In the recognition phase, an additional 50 abstract and 50 concrete words—also matched on word length and frequency—which were not presented at encoding were added to the target and non-target items.

Table 1

Stimulus Characteristics

	Targets			Synonyms		
	n_{letter}	log^F	conc	n_{letter}	log^F	conc
Abstract	6.7	5.0	1.8	7.3	5.7	2.1
Concrete	7.0	5.1	4.9	6.2	5.7	4.8

Note. Mean values for word length (# of letters), log word frequency, and concreteness.

Procedure. Participants performed a two-phase source memory task. Stimuli were presented visually one at a time, in pseudorandomized order,[3] with an arbitrary frame context (either a red or a green frame). On each word, participants performed a synonym-judgment 1-back task. To ensure that they did not ignore the frames, the hand they used to make their response was determined by frame color (left hand for words in green frames and right for red). Stimuli were presented for 2000 ms with a 1000-ms interstimulus interval. Participants were told there would be a later memory test on the words, but not that memory for the contextual detail (i.e., frame color) would be tested.

In the recognition phase, participants performed two tasks for each word. First, they responded whether they had seen the word at encoding by selecting their degree of confidence in having seen it before (they could select high, medium, and low confidence for either “old” or “new”). Second, for old words, they indicated the color of the frame on initial encoding. The task was the same for new words, except that they were asked simply to select the color they thought the frame would have been had it been presented at encoding. Participants were given 6000 ms each for the old/new and the frame color judgment.

Data analysis. Data were analyzed using R (R Core Team, 2013). All responses of less than 150 ms were removed (3.8% of responses)—because a decision and response could not be made at that speed, these responses were assumed to be in error, or an attempted response to the previous trial after that trial had timed out. Memory for items (i.e., words) and their contexts (i.e., frame color) was first analyzed using descriptive statistics, calculating accuracy, hit rate, miss rate, correct rejections, false alarms, and d' (calculated as z(Hit) – z(FA)) for all words, and

[3] Each participant saw the same order, which was randomized and then edited such that synonym pairs were adjacent.

accuracy was also assessed by level of confidence. Context (i.e., frame) memory accuracy was calculated only for target hits, and was assessed across confidence levels. Context memory accuracy was analyzed as a function of word type (abstract or concrete)[4] and confidence in having seen the word at encoding (low, medium, or high). Logistic mixed effects models (*lme4* package; Bates et al., 2017) were used to analyze the data, with subject and word as random intercepts,[5] and word type (abstract or concrete), level of confidence (low, medium, high), and their interaction as treatment-coded fixed effects. Thus, the models were of the following form:

$$accuracy \sim wordType + confidence + (type{:}confidence) + (1|subject) + (1|word)$$

For each effect, we report model estimates, *z*-values, and *p*-values. Each predictor was entered in a successive model, and statistical significance was assessed by comparing the models using likelihood ratio tests.[6] For brevity and readability, full model details are reported in tables, while only the statistical significance of the model comparisons is reported in text. For all analyses, *p*-values < .05 were considered statistically significant.

Results

Item recognition. Before reporting on our measure of primary interest (context memory), we first assess overall recognition memory (as well as hit, miss, correct rejection, and false alarm rates) for concrete and abstract words to provide a baseline measure of recognition

---

[4] While we designed our study and conducted our power analysis in accordance with a 2 (word type) × 3 (confidence) design, as we note above, there is no clear-cut dichotomy between abstract and concrete concepts (e.g., Vigliocco et al., 2009). Thus, at the suggestion of one reviewer, we also implemented identical models using concreteness as a continuous predictor. These models produced nearly identical conclusions to the binary models, presumably because we selected our concepts to fall at the extreme ends of the concreteness continuum.

[5] We initially attempted to implement models with random slopes for word type over subject, confidence over subject, and confidence over word, but these models did not converge. Thus, we opted for intercepts-only models.

[6] Our hypotheses strictly concerned accuracy on the source memory task, but for completeness, we have included analogous models of response time (using linear mixed effects models) in the Supplemental Material available online. Briefly, these models largely showed the same patterns in the word recognition tasks—faster RTs for concrete words and more confident responses among *all* words, and no effect of concreteness in targets only—and divergent effects in the context recognition tasks. That is, participants were generally no faster to make context recognition judgments for abstract vs. concrete words.

memory for concrete and abstract words (Table 2). Hit rates were higher and false alarms lower (an effect known as the *mirror effect*; Glanzer & Adams, 1985) in concrete than abstract words, an effect which has previously been observed for concreteness (Glanzer & Adams, 1990). For overall accuracy, there were main effects of both word type and confidence. Concrete words were better recognized than abstract ($\chi^2(1) = 10.36$, $p = .001$), and accuracy increased with greater confidence ($\chi^2(2) = 593.35$, $p < .001$). Their interaction was non-significant ($\chi^2(2) = 3.43$, $p = .18$). d' analysis showed that when considering response sensitivity, accuracy remained better for concrete concepts, $t(39) = -5.37$, $p < .001$. Among targets only (i.e., non-synonym words presented at encoding), there was no main effect of word type on recognition memory ($\chi^2(1) = 0.29$, $p = .59$), but a main effect of confidence level ($\chi^2(2) = 681.14$, $p < .001$), with recognition memory accuracy increasing as confidence level goes up. The interaction was non-significant ($\chi^2(2) = 4.04$, $p = .13$). Figure 1a shows means and 95% CIs for word and context memory (collapsing across confidence levels), and the detailed model results are shown in Table 3.

Table 2

Mean Item Recognition Accuracy

Word type	Acc	Hit	Miss	CR	FA	d'
Abstract	.73	.77	.23	.66	.34	1.21
Concrete	.78	.81	.19	.73	.27	1.57

Note. Acc = overall item recognition accuracy; CR = correct rejection; FA = false alarm.

Table 3

Summary of Models Predicting Accuracy for All Words, Targets, and Frame Recognition

	All Words			Targets Only			Frame Recognition		
	Est	*z*	*p*	Est	*z*	*p*	Est	*z*	*p*
Word Type: Concrete	0.30	3.25	**.001**	0.07	0.54	.587	0.19	2.30	**.02**
Confidence: Medium	0.46	7.18	< **.001**	1.19	10.64	< **.001**	0.12	0.98	.33
Confidence: High	1.54	22.62	< **.001**	2.91	23.01	< **.001**	0.17	1.45	.15
Word Type × Confidence: Medium	-0.01	-0.13	.90	-0.36	-1.78	.07	0.06	0.23	.82
Word Type × Confidence: High	0.18	1.47	.14	-0.06	-0.26	.79	0.17	0.74	.46

Note. For word type, abstract is the reference level, and for confidence, low is the reference.

Context (frame color) memory. To test our primary question—whether context memory is better for abstract concepts—we included only trials for which the target word had been correctly recognized. There was a main effect of word type, where the frame color was less likely to be remembered for abstract words ($\chi^2(1) = 5.16$, $p = .02$; Figure 1b), but no main effect of confidence level in having seen the word at encoding. The interaction of word type and confidence was non-significant. Detailed model results are shown in Table 3.

Because of the baseline advantage for concrete words in item recognition (evident in both accuracy and d') it is necessary to examine whether this advantage could have biased the context memory models. That is, the strength with which the word was encoded (which is well-represented by d'), not concreteness, may have driven context memory performance. Accordingly, we also constructed models with d' as a predictor to determine whether, after accounting for encoding strength, memory for frame color is still inferior for abstract words. A likelihood ratio test comparing the model with both d' and word type versus the model with only d' was significant, $\chi^2(1) = 5.27$, $p = .02$, suggesting that the effect of word type, where frame recognition was worse in abstract than it was in concrete concepts, was significant even after accounting for the baseline advantage for recognizing concrete words.

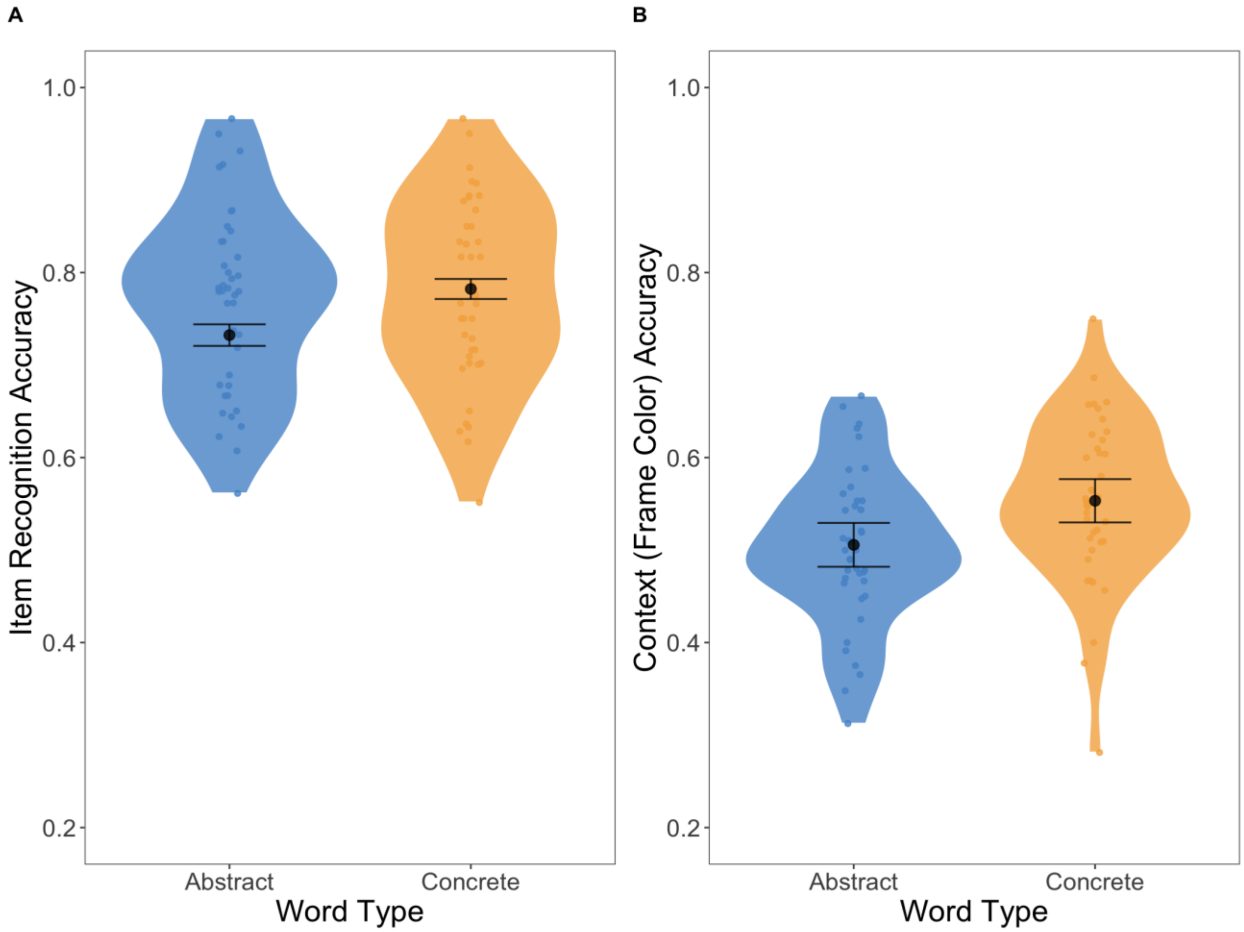

Figure 1. Effects of concreteness on (a) overall item recognition accuracy and (b) context (i.e., frame color) memory. Solid black point reflects the condition mean. Error bars are estimated 95% confidence intervals around the means. Individual points within each density violin are individual subjects.

Discussion

Context memory was worse for abstract concepts, and this was true even after controlling for a baseline advantage in recognizing concrete words. Thus, the results of Experiment 1 ran counter to the simple hypothesis that relational memory is better for abstract than concrete concepts. Instead, context memory was *worse* for abstract than for concrete words. Why did this difference emerge? This relational memory advantage for concrete concepts could be because, as suggested by the alternative hypothesis that we raised, when processing abstract concepts, highly arbitrary information (e.g., frame color) is inhibited in favor of more systematic information. Of course, in our experiment, there was no systematic information present in the context of the target word to be encoded. However, Davis et al. (2020) propose, based on neurobiological evidence (e.g. van Kesteren et al., 2013), that when recognizing the sparsely distributed patterns of information in the environment that serve as cues to the activation of abstract concepts, there is greater reliance on top-down schema-based information than when recognizing information congruent with concrete concepts. This produces, for abstract concepts, greater inhibition of the mechanisms that bind arbitrary elements within the episode to one another (this inhibition resulting from the complementarity observed by van Kesteren et al., 2013, between the brain regions associated with schema and relational binding; i.e., medial prefrontal cortex and hippocampus, respectively). To give an example, a word like “decision” will activate more schema-based information during its comprehension than “chair,” resulting in greater inhibition of arbitrary information co-present in its context.

Another possibility is that relational binding is generally better for abstract concepts, but the specific contextual detail used in this task happened to promote better binding for concrete than abstract concepts to a color frame context. That is, concrete concepts may be more amenable to a mnemonic strategy wherein a color adjective (i.e., “red” or “green”) could readily be bound to concrete objects (e.g., “table”), making context memory better for concrete words. If true, by changing the to-be-remembered context to one that is not more readily bound with concrete than abstract concepts, we should observe a relational memory advantage for abstract concepts.

Experiment 2

In Experiment 2, we utilized a variant of the source memory paradigm, where instead of the frame, the context to be encoded was a male or female voice—the idea being that unlike color adjectives, speaker voice is not (at least not in any obvious way) more easily bound to concrete than abstract concepts. In fact, person-related social properties—which could arguably include voice—may be more important for abstract than concrete concepts (Barsalou & Wiemer-Hastings, 2005). Concepts were presented auditorily, and memory was assessed on visually presented words (e.g., Wilding & Rugg, 1996). If the simple hypothesis—that contextual detail generally is encoded to a greater extent in abstract concepts—is correct, source memory (i.e., was it spoken by a male or female voice?) should be better for abstract concepts.

Methods

Participants. Forty-two UConn undergraduates (7 men, 35 women, mean age = 18.9 years) with normal or corrected-to-normal vision who had not participated in Experiment 1 provided informed consent and were given course credit for their participation. As in Experiment 1, there were no effects of demographic variables (age, gender) on any of our dependent measures. One participant was excluded for non-compliance (again, pressing the same button throughout the experiment), leaving $N = 41$.

Stimuli. The words were the same as those used in Experiment 1, but rather than being presented visually they were instead recorded by a male and a female speaker, with half the words presented by the male speaker and half by the female speaker. As with frame color, this list was held constant across participants. There were no differences in the length of the sound files between the two speakers, and all files were normalized to a peak amplitude.

Procedure. In the encoding phase, the procedure was the same as in Experiment 1. The voice of the speaker determined the hand participants used to make their judgments. In the memory phase, the first judgment—whether the word was in the initial set (old) or not (new)—was the same. For the second judgment, participants were asked to indicate whether the person who said the word in the initial set was "Jane" or "Sid." The test phase was conducted with visually presented words, as in Experiment 1.

Data analysis. Data were analyzed in the same way as in Experiment 1.

Results

Item recognition. Accuracy and hit, miss, correct rejection, and false alarm rates across all words are shown in Table 3. Among all words, there was a significant main effect of both word type, with concrete words showing better recognition ($\chi^2(1) = 6.77$, $p = .009$), and confidence level, with both medium and high showing greater accuracy than low confidence ($\chi^2(2) = 610.85$, $p < .001$). The word type × confidence interaction was non-significant ($\chi^2(2) = 4.26$, $p = .12$). d' analysis revealed that after considering response sensitivity, accuracy was better for concrete concepts, $t(40) = -3.49$, $p = .001$. Among targets, there was a main effect of confidence ($\chi^2(2) = 961.49$, $p < .001$), but not of word type ($\chi^2(2) = 0.39$, $p = .53$). The interaction was significant ($\chi^2(2) = 9.18$, $p = .01$) at high confidence, suggesting that at greater memory strength, item recognition was worse for abstract words. Means and 95% CIs for the main effects of word type on word memory are visualized in Figure 2a, and detailed model results are shown in Table 5.

Table 4

Mean Word Recognition Accuracy

Word type	Acc	Hit	Miss	CR	FA	d'
Abstract	.70	.72	.28	.64	.36	1.04
Concrete	.73	.77	.23	.67	.33	1.28

Note. Acc = overall item recognition accuracy; CR = correct rejection; FA = false alarm.

Table 5

Summary of Models Predicting Accuracy for All Words, Targets, and Voice Recognition

	All Words			Targets Only			Voice Recognition		
	Est	*z*	*p*	Est	*z*	*p*	Est	*z*	*p*
Word Type: Concrete	0.23	2.62	**.008**	0.08	0.63	.53	0.23	2.42	**.016**
Confidence: Medium	0.33	5.27	< **.001**	1.51	13.17	< **.001**	-0.02	-0.13	.89
Confidence: High	1.40	22.12	< **.001**	3.29	26.57	< **.001**	0.39	3.21	**.001**
Word Type × Confidence: Medium	0.11	1.03	.30	0.32	1.62	.11	0.19	0.74	.46
Word Type × Confidence: High	0.23	2.07	**.04**	0.61	3.03	**.002**	0.34	1.47	.14

Note. For word type, abstract is the reference level, and for confidence, low is the reference.

Context (voice source) memory. To test our primary question—whether context (here, voice source) memory is better for abstract concepts—we again included only trials for which the word had between correctly recognized. There was a main effect of word type, with source memory for the voice context worse for abstract words ($\chi^2(1) = 5.70$, $p = .017$), as well as a main effect of confidence ($\chi^2(2) = 25.22$, $p < .001$). The interaction of word type and confidence was non-significant ($\chi^2(2) = 2.49$, $p = .29$). Thus, here, like in Experiment 1, participants were less likely to recognize the context correctly for abstract as compared to concrete words. Means and 95% CIs are shown in Figure 2b, and the detailed model results are shown in Table 5.

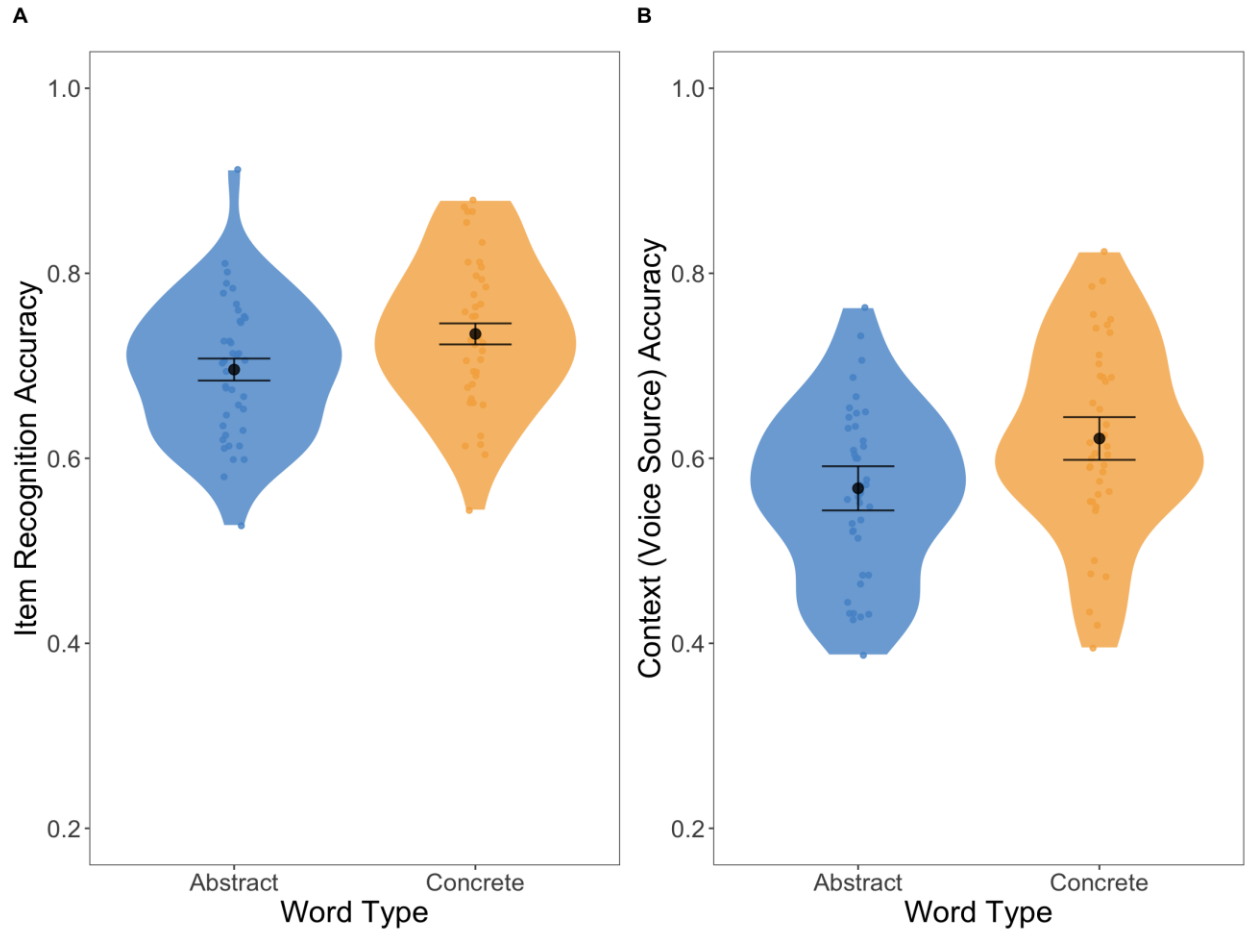

Figure 2. Effects of concreteness on (a) overall item recognition accuracy and (b) source (i.e., voice) memory. Solid black point reflects the condition mean. Error bars are estimated 95% confidence intervals around the means. Individual points within each density violin are individual subjects.

As in Experiment 1, there was a baseline advantage for concrete words in item recognition (again, evident in both accuracy and d’, Table 3). Thus, to test whether the strength with which the word was encoded, not concreteness, drove context memory performance, like in Experiment 1, we constructed models with d' as a predictor. A likelihood ratio test comparing the model with both d' and word type versus the model with only d' was significant, $\chi^2(1) = 5.75$, $p = .016$, suggesting that the effect of word type, where source memory was worse for abstract than it was for concrete concepts, was significant even after accounting for the baseline advantage for recognizing concrete words.

Discussion

Like in Experiment 1, context memory was worse for abstract concepts. This was the case even when the to-be-remembered context was, in principle, no more likely to be bound with concrete as compared to abstract concepts. Thus, the two arbitrary episodic details (color and voice) that we have examined thus far appear to be better remembered in the context of concrete as compared to abstract concepts. This is consistent with the hypothesis that when processing abstract concepts, arbitrary information is inhibited in favor of more systematic information. Note that if our interpretation of the results of Experiments 1 and 2 is correct, that is, if a semantic dimension (abstractness) does indeed affect recognition of episodic context, this would support an integrated view of semantic and episodic memories.

However, both Experiments 1 and 2 showed a baseline memory advantage for concrete words, and thus they may have been more strongly encoded. Although we did adjust for this advantage in our statistical analysis, avoiding this potential confound altogether would be more convincing. Accordingly, we conducted a third experiment where we controlled for this baseline concreteness advantage in encoding strength.

Experiment 3

In Experiment 3, we simplified the test phase by probing *only* recognition memory: half of the words were presented in the same frame color as they were at encoding (i.e., frame color *retained*), while half of the words were presented in a different frame color (i.e., frame color *changed*). The idea here is that we can control for strength of encoding by comparing the *relative* advantage conferred by keeping the context constant from exposure to test between abstract and concrete concepts—i.e., while the memory trace left by abstract concepts may be weaker overall, the benefit of maintaining the same frame color between exposure and test may be larger for abstract than concrete concepts. On the other hand, if recognition memory accuracy for abstract concepts is *worse* when the frame color at encoding is retained at test, it would suggest—in line with the alternative hypothesis—that arbitrary episodic detail may be inhibited in abstract concepts.

Methods

Participants. Forty UConn undergraduates (10 men, 30 women, mean age = 19.2 years) with normal or corrected-to-normal vision who had not participated in Experiment 1 or 2 provided written informed consent and received course credit. As in Experiment 1, individuals with color-blindness were ineligible, and again, there were no effects of demographic variables (age, gender) on any of our dependent measures. Four subjects were removed for non-compliance, leaving $N = 36$.

Stimuli. The stimuli were the same as those in Experiments 1 and 2, and frame color assignment was counterbalanced across participants.

Procedure. The encoding procedure was the same as in Experiment 1. At test, participants were asked to identify as many old words as possible, ignoring the color of the frame. Words were presented in the red and green frames. Half of the words retained the frame color from encoding, and half changed color.

Data analysis. Item recognition data were analyzed in the same way as in Experiments 1 and 2. However, frame retention (retained vs. changed) was used as a second fixed effect in the mixed logit model (thus replacing *confidence* in the model presented in Experiment 1), and we assessed the word type × frame retention interaction as the critical test of our hypothesis.

Results and discussion

Accuracy and hit, miss, correct rejection, and false alarm rates across all words are shown in Table 6. In overall old/new item recognition memory, there was a main effect of word type (est = 0.36, z = 3.85, p < .001; model: $\chi^2(1) = 14.36$, p < .001), where memory was better for concrete words. Among targets only, however, there was no concreteness advantage (est = 0.07, z = 0.66, p = .51; model: $\chi^2(1) = 0.43$, p = .51), but there was a significant main effect of frame retention (est = -.15, z = -2.03, p = .042; model: $\chi^2(1) = 4.06$, p = .044), where accuracy was surprisingly *worse* when the context was retained than when it was changed.

Turning to our question of primary interest, there was an interaction between word type and frame retention (est = .34, z = 2.23, p = .026; model: $\chi^2(1) = 4.92$, p = .027; Figure 3), providing additional evidence that concreteness can influence memory for episodic contexts. Importantly, accuracy was *worse* when the frame color was retained in abstract concepts, again operating counter to the hypothesis that episodic context *in general* is a critical part of processing highly abstract concepts. Rather, the results are consistent with the idea that when it is arbitrary, episodic context may in fact be *inhibited* in abstract concepts.

Table 6

Mean Item Recognition Accuracy

Word type	Acc	Hit	Miss	CR	FA	d'
Abstract	.77	.77	.23	.77	.23	1.59
Concrete	.82	.82	.18	.84	.16	2.06

Note. Acc = overall item recognition accuracy; CR = correct rejection; FA = false alarm.

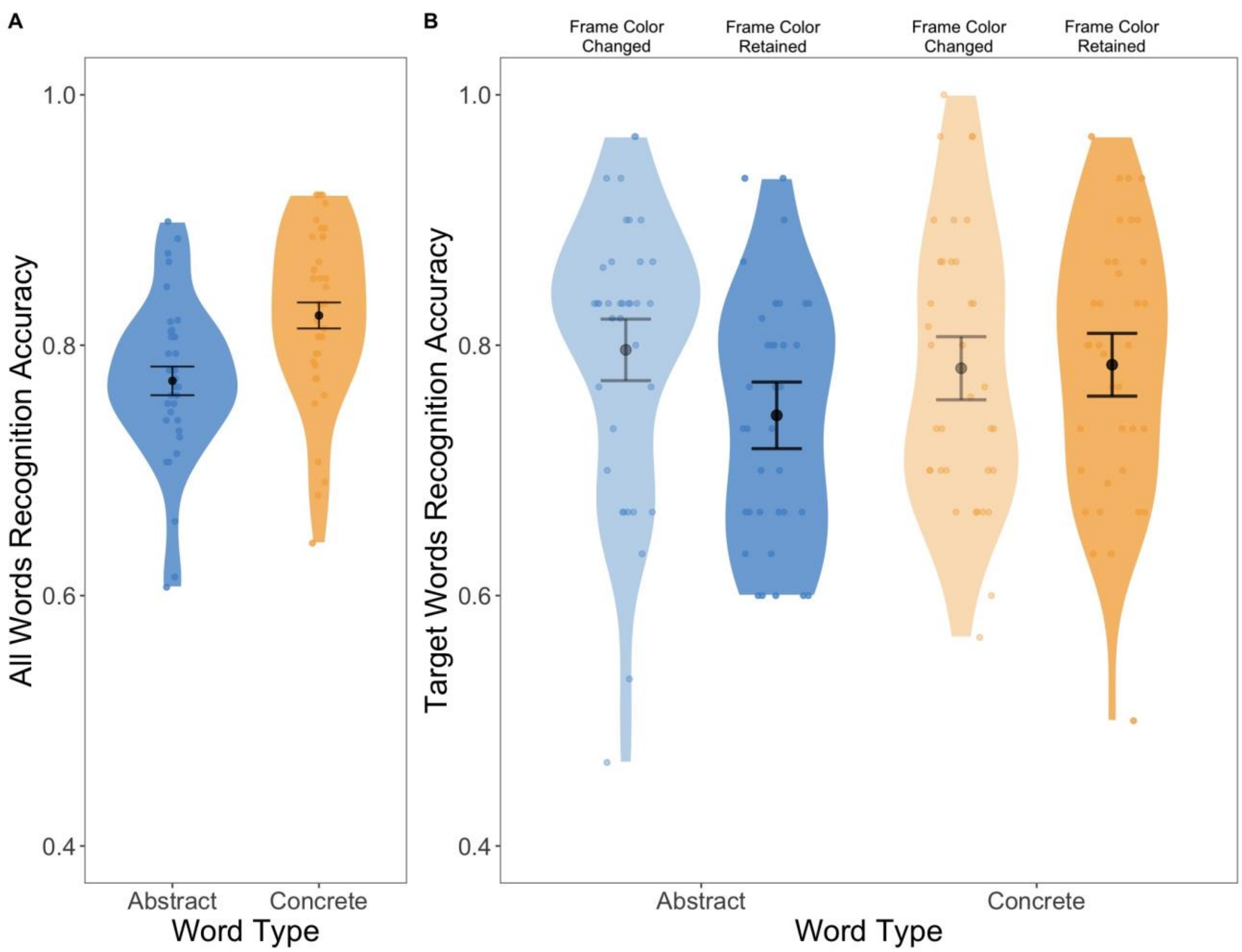

Figure 3. Plots showing (a) the main effect of concreteness on item recognition memory for all words and (b) the interaction between word type and frame retention on target recognition memory accuracy. Solid black point reflects the condition mean. Error bars are estimated 95% confidence intervals around the means. Individual points within each density violin are individual subjects.

Experiment 4

In Experiment 4, we tested whether the apparent inhibition we observed in Experiment 3 for retained-context abstract words would extend to voice—a type of context which is arguably a person-related social property, a class that may be particularly important for abstract concepts (e.g., Barsalou & Wiemer-Hastings, 2005). Specifically, we tested whether word recognition would be *hindered* in abstract concepts (relative to concrete concepts) when the same speaker from the encoding phase also presented the word at recognition.

Methods

Participants. Thirty-nine UConn undergraduates (12 men, 27 women, mean age = 19.3) with normal or corrected-to-normal vision who had not participated in Experiments 1, 2, or 3 provided written informed consent and received course credit. There were again no effects of demographic variables (age, gender) on any of our dependent measures. Two participants were excluded due to non-compliance, leaving $N = 37$.

Stimuli. The stimuli were the same as those in Experiments 1–3, and voice source assignment was counterbalanced across participants.

Procedure. The encoding procedure was the same as in Experiment 2, while the recognition procedure was borrowed from Experiment 3. At test, participants were asked to identify as many old words as possible, irrespective of the identity of the speaker. Words were presented by the male and female voices. Half of the words retained the voice source from encoding, and half changed to the other voice used at encoding.

Data analysis. Data analysis was identical to that in Experiment 3.

Results and discussion

Accuracy and hit, miss, correct rejection, and false alarm rates across all words are shown in Table 7. As in Experiment 3, in overall old/new item recognition memory, there was a main effect of word type (est = .25, $z = 2.61$, $p = .009$; model: $\chi^2(1) = 6.69$, $p = .010$; Figure 4a), where memory was better for concrete words. Also like in Experiment 3, among targets only there was no concreteness advantage (est = .06, $z = 0.45$, $p = .65$; model: $\chi^2(1) = 0.20$, $p = .66$). Unlike in Experiment 3, however, the effect of (voice) retention was non-significant (est = -.11, $z = -1.59$, *p* = .11; model: $\chi^2(1) = 2.51$, *p* = .11). Turning to our question of primary interest, there was no type × retention interaction (est = .11, *z* = .82, *p* = .41; model: $\chi^2(1) = 0.67$, *p* = .41; see Figure 4).

Table 7

Mean Item Recognition Accuracy

Word type	Acc	Hit	Miss	CR	FA	d'
Abstract	.70	.66	.34	.79	.21	1.32
Concrete	.75	.71	.29	.83	.17	1.66

Note. Acc = overall item recognition accuracy; CR = correct rejection; FA = false alarm.

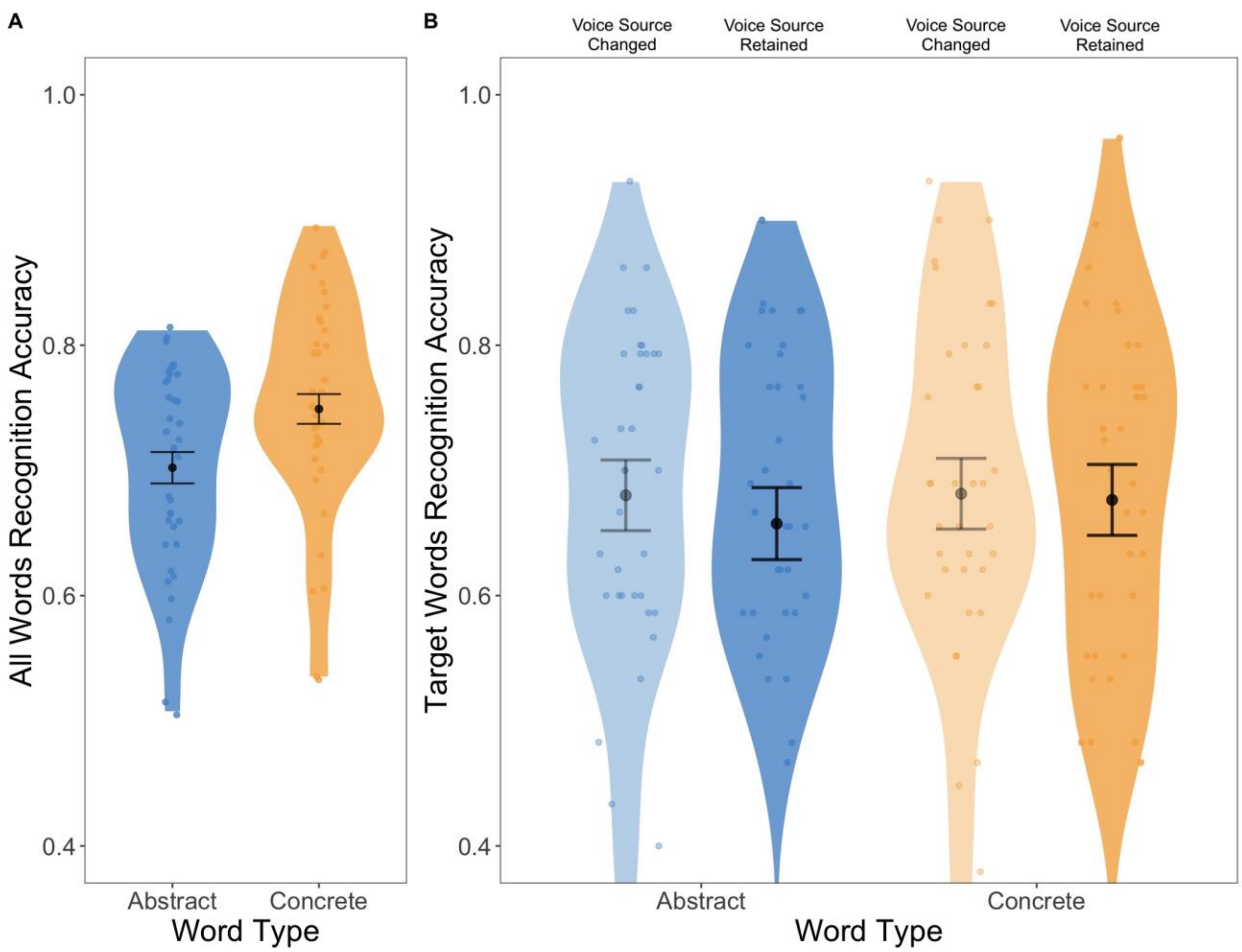

Figure 4. Plots showing (a) the main effect of concreteness on item recognition memory for all words and (b) the interaction between word type and voice retention on target recognition memory accuracy. Solid black point reflects the condition mean. Error bars are estimated 95% confidence intervals around the means. Individual points within each density violin are individual subjects.

Although the pattern observed in Experiment 4 was the same as that observed in Experiment 3 numerically (a disadvantage for abstract concepts, but no difference for concrete concepts, when the context at encoding was retained at test), in Experiment 4, this difference was not reliable. Given the similar pattern, one possibility is that we simply failed to detect the effect of voice in Experiment 4 due to greater variability in the data compared to Experiment 3. Another possibility, however, is that because abstract concepts tend to be associated with social-communicative contexts, speaker identity is, in general, more relevant to the real-world processing of abstract concepts than frame color is, and thus less likely to be inhibited. That is, while voice context was arbitrarily related to the meanings of the words in the experimental context, it may not be inhibited to the degree that frame color is because it *tends to be* more generally relevant for the meanings of abstract concepts. (In the General Discussion, we not only provide preliminary evidence in support of the second possibility but also provide the underlying theoretical rationale.) To test the hypothesis that more relevant (yet still arbitrary in the context of the experiment) episodic context might facilitate context sensitivity in abstract concepts, we conducted Experiment 5.

Experiment 5

In Experiment 5, we take as our starting point that location is more situationally relevant for the recognition of abstract concepts (that is, for the recognition of the patterns of information that cue the concept) than for the recognition of concrete concepts; a river is a river no matter where it is located, but a decision at a casino is different in non-arbitrary ways from a decision in a court of law. From this starting point, we surmise that abstract concepts are more reliant on situationally determined location than are concrete concepts (below, we provide evidence that this is true, as well as evidence that situationally determined location is more relevant for understanding abstract concepts than is voice or color), and that they may, therefore, more strongly engage the hippocampal-based mechanisms that encode location (Epstein & Kanwisher, 1998). If location is a more constitutive component of abstract than concrete concepts, we might expect that location is more strongly encoded with the other cues to a given abstract concept, making it more resistant to the inhibition that can occur due to activation of the concept (as mediated by schema knowledge—see above).

We presented the same words from Experiments 1–4 in different quadrants of the display. The location of each word was either changed or retained at the recognition phase. As in Experiments 3 and 4, we were interested in whether retaining the context—this time, spatial location—would confer a recognition benefit for abstract concepts. We anticipated better performance overall during the recognition phase when the location of the word was retained. And we anticipated that this better performance would favor abstract over concrete words.

Methods

Participants. Forty-one UConn undergraduates (16 men, 25 women, mean age = 19.4) with normal or corrected-to-normal vision who had not participated in Experiments 1–4 provided written informed consent and received course credit. No effects of demographic variables (age, gender) on any of our dependent measures were observed. One participant was omitted due to an experimenter error (the data output file was configured incorrectly), leaving $N = 40$.

Stimuli. The stimuli were the same as those in Experiments 1–4, and location retention (i.e., which stimuli had their location retained between the encoding and recognition phases) was counterbalanced across participants.

Procedure. The encoding procedure was similar to that used in Experiments 1–4, except that the words could appear in one of four quadrants on the screen. In the recognition phase, half of the words retained their location from encoding, and half changed. The same number of words changed to each of the other three quadrants (i.e., when the location changed, it was equally likely that the word would appear in each of the other three quadrants).

Data analysis. Data analysis was identical to that in Experiments 3 and 4.

Results and discussion

Accuracy and hit, miss, correct rejection, and false alarm rates across all words are shown in Table 8. In overall old/new item recognition memory, there was once again a main effect of word type (est = 0.29, $z = 2.74$, $p = .006$; model: $\chi^2(1) = 7.38$, $p = .007$; Figure 5a), where memory was better for concrete words. As in Experiment 4, when examining only target words there was no concreteness advantage (est = -.001, $z = -0.01$, $p = .99$), but there was a main effect of location retention (est = 0.28, $z = 2.22$, $p = .03$; model: $\chi^2(1) = 4.83$, $p = .03$), such that recognition memory was better when the spatial context was retained (i.e., the word appeared in the same location on the screen as it had at exposure).

Turning to our question of primary interest, we also observed the predicted interaction between word type and location retention (est = -0.50, $z = -2.00$, $p = .04$; model: $\chi^2(1) = 3.96$, $p = .05$; Figure 5b). Thus, like in Experiments 1–3, we again find evidence that concreteness can influence recognition of episodic context—that is, the episodic memory system is recruited differently depending on semantic content. More importantly for our current question was the direction of the effect—we observed a *facilitatory* effect of location retention on recognition of abstract concepts. While recognition memory was better overall when the spatial context from exposure was retained at recognition, the benefit was greater for abstract concepts.

Table 8

Mean Item Recognition Accuracy

Word type	Acc	Hit	Miss	CR	FA	d'
Abstract	.67	.62	.38	.76	.24	1.10
Concrete	.72	.67	.33	.83	.17	1.63

Note. Acc = overall item recognition accuracy; CR = correct rejection; FA = false alarm.

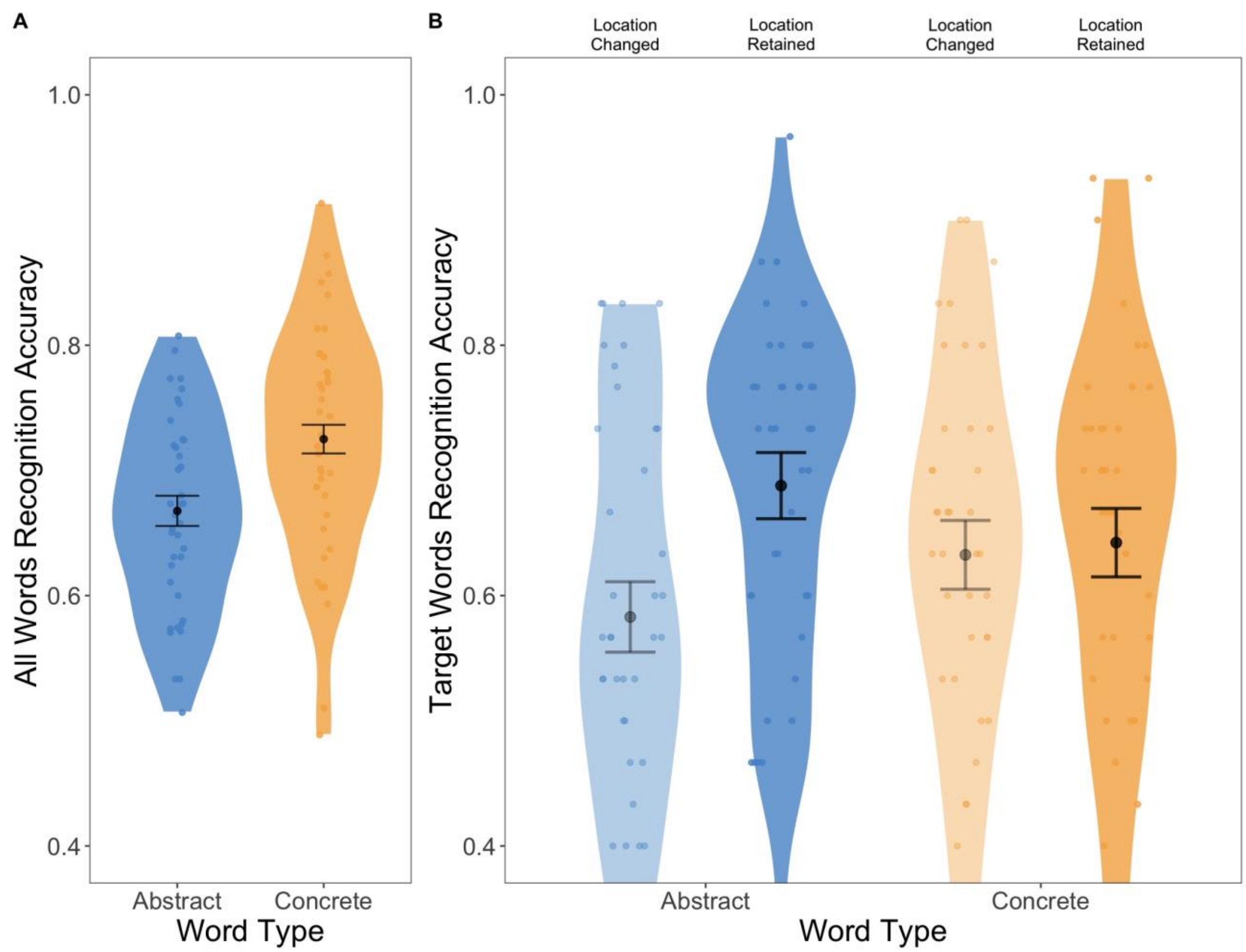

Figure 5. Plots showing (a) the main effect of concreteness on item recognition memory for all words and (b) the interaction between word type and location retention on target recognition memory accuracy. Solid black point reflects the condition mean. Error bars are estimated 95% confidence intervals around the means. Individual points within each density violin are individual subjects.

This context reinstatement benefit for abstract concepts is consistent with our conjecture that location tends to be a relevant cue to interpreting abstract concepts in the real world, and that this real-world importance means that location is likely to be encoded with abstract concepts (despite that in the experiment it is manipulated as an arbitrary context).

As a test of the validity of this intuition—i.e., that spatial location tends to be more relevant to understanding abstract concepts—we collected a set of ratings from 60 additional UConn undergraduate students on how important each type of context is in affecting the meaning of each concept tested across the five experiments. For color, participants (2 men, 18 women, mean age = 18.5 years) were asked to indicate how important the color of the surrounding context was in affecting the meaning of each concept. For voice (5 men, 15 women, mean age = 19.2 years), they were asked to rate how important the voice of a speaker talking about each concept was in affecting its meaning. And finally, for location, participants (5 men, 15 women, mean age = 19.2 years) assessed the degree to which the location something appears in might influence its meaning. (See Appendix for full instructions.) Participants indeed rated location as most relevant to abstract concepts followed by voice and then color (Figure 6).[7] Because location was rated as relatively important (and more important than voice or color for understanding the meaning of concepts, these ratings are consistent with our conjecture that context reinstatement aids recognition memory for abstract concepts, but *only* when that context is relevant to interpreting the meaning of abstract concepts in the real world.

---

[7] While we also collected ratings for concrete concepts and include them here for completeness, to the extent that concrete concepts are indeed more situationally systematic than abstract concepts, the relative importance of these features should be less predictive of whether they are episodically bound to the concept.

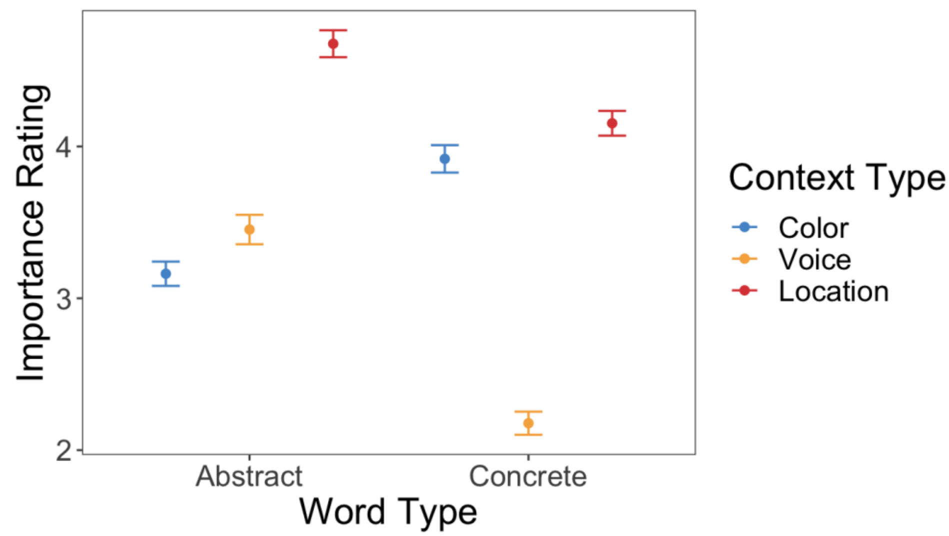

Figure 6. Mean (± 95% confidence intervals) ratings of the importance of each type of context in affecting the meaning of each concept. Higher ratings reflect greater importance.

General Discussion

We tested whether arbitrary episodic contexts are better encoded with abstract concepts, or with concrete concepts. In Experiments 1 and 2, there was a concreteness advantage for recognizing episodic contexts. In Experiment 3, episodic context preservation conferred a *disadvantage* for recognizing abstract concepts, suggesting the presence of a mechanism whereby arbitrary associations are *inhibited* in the episodic experience(s) of the situations that activate abstract concepts (later, we discuss possible mechanisms). In Experiment 4, we observed a null effect: there was no benefit or disadvantage of context preservation for recognizing abstract concepts when that context was a speaker's voice. Finally, in Experiment 5, we varied the location in which abstract concepts were presented at encoding. The motivation here was to use a context that might be more relevant to regular processing of abstract concepts: because abstract concepts are particularly dependent on situational location for determining their meaning (as demonstrated in our rating study above; for additional discussion, see Davis et al., 2020), even arbitrarily associated location might be better encoded with abstract concepts. Here, we found a benefit of location retention for abstract concepts at test. To summarize, we observed that the way the episodic memory system is recruited during conceptual processing can be modulated by semantic content, and in particular, its recruitment differs as a function of concreteness (notwithstanding Experiment 4, which we return to below).

Across several literatures it is agreed that context is critical for understanding abstract concepts. However, there are differences across frameworks in terms of the *type* of context specified as being particularly important to processing abstract concepts, ranging from semantically constraining context (e.g., “The evidence was presented in court and the judge made her *decision*) in context-availability theory (Schwanenflugel & Shoben, 1983), to thematic associations in the qualitatively different representations framework (e.g., *decision*, *judge*, *gavel*; Crutch & Warrington, 2005), to meaningful situational and internal factors in grounded cognition (Barsalou, 1999; Barsalou & Wiemer-Hastings, 2005). The present study examined whether there is a basic mechanism that might unify these approaches—namely, sensitivity to episodic information, and consequently, better relational memory for abstract concepts. However, the results suggest an alternative, more complicated picture: memory for episodic context tends to be *worse for* abstract concepts, unless that context is one which is typically more informative to abstract concept meaning. In the following, we seek to unpack these findings by exploring potential relations between concreteness and the episodic memory system, neurocognitive considerations for abstract concept representation, and potentially promising avenues for further exploring the neurocognitive dynamics underpinning representation of abstract concepts.

Concreteness, Context, and Episodic Memory

Concreteness is a powerful organizing factor in semantic memory (e.g., De Deyne, 2017; Hollis & Westbury, 2016), and concreteness effects are near ubiquitous in recognition memory studies (e.g., Nelson & Schreiber, 1992; Paivio et al., 1994; Wattenmaker & Shoben, 1987). The present results suggest that such effects can extend beyond stronger memory for concrete concepts to include better associative, *relational memory* for arbitrary contexts for more concrete concepts. This is not the case, however, when the arbitrary context is location-based, perhaps because, as we suggested above, abstract concept meanings might be more sensitive to situational location in the real world. One important consideration here is the way in which we might expect context to be differentially recruited for processing relatively concrete and relatively abstract concepts, as this has implications for the relation between context sensitivity and concreteness.

In a review of the pervasiveness of context effects in cognition and perception, Yeh and Barsalou (2006) present two theses for how context affects concept processing: (1) contexts and concepts mutually activate each other, such that when processing a context, associated concepts are activated, and vice versa (e.g., *coffee* activates *café*, and *vice versa*); and (2) when processing a concept in a particular context, properties of the concept which are relevant to that context become active (e.g., thinking about *decision* when determining an appropriate drink late at night would activate different properties than thinking about *decision* when determining appropriate dress for a virtual meeting). These two theses have different implications for the relation between context sensitivity and concreteness.

The first thesis resonates strongly with context availability theory, and likely suggests a concrete word advantage: concrete concepts activate contexts more strongly because they have stronger implicit ties to specific contexts (e.g., *coffee–café*), and denser networks of contextual associations (e.g., *coffee*: café, milk, mug, sugar, latte, etc.; Schwanenflugel & Shoben, 1983; for discussion, see Kousta et al., 2011). Thus, a mechanism similar to that which underpins context availability effects may have facilitated building implicit, direct associations to the context (such as the surrounding color) for concrete concepts in the present study. Indeed, it is possible that source memory effects (and presumably, hippocampal processing) are more closely related to the first thesis, as they deal with implicit, proximal connections between stimulus and context (for review, see Eichenbaum, 2013). Relatedly, an fMRI study of recognition memory for abstract and concrete concepts showed a relation between hippocampal activation and a behavioral concreteness advantage (Fliessbach et al., 2006). In this same study, abstract concepts showed greater left inferior frontal gyrus activation at encoding, perhaps reflecting a more effortful search for potentially relevant contexts and associations, and convergent with the extant neuroimaging literature on abstract concept processing (e.g., Binder et al., 2005; Hoffman et al., 2010; for review, see Wang et al., 2010). Under this explanation, arbitrary episodic context is more strongly associated with concrete concepts because concrete concepts are generally easier to contextualize, owing to their dense networks of contextual associations. However, it does not explain why some arbitrary episodic contexts are *inhibited* in processing abstract concepts (Experiment 3), while others *facilitate* abstract concept processing (Experiment 5).

Yeh and Barsalou’s (2006) second thesis may be more pertinent to abstract concept processing: when processing *decision* in the context of your choice of beverage at 9pm in the local café, the activated properties will be different from when processing decision in the context of a judge determining the appropriate sentence for a felon convicted of battery. That is, schema knowledge—semantic knowledge of situations and the events and elements of which they are typically composed—can vary considerably across instantiations of abstract concepts. *Decision* has a number of possible interpretations, and its precise meaning—and thus, the properties activated—depends on the situation and (systematically) associated schema-based knowledge (for related work on semantic diversity, see e.g., Hoffman et al., 2013, 2015).

Research on the neural dynamics underpinning schema processing (e.g., van Kesteren et al., 2013) suggests that activating such systematic associations may in fact inhibit the formation of associations with arbitrary elements of an episode. This dynamic is rooted in the interplay between neural systems in medial frontal and medial temporal lobe, where medial frontal activation when processing systematic associations may dampen activation of medial temporal lobe (i.e., hippocampal structures), thereby inhibiting the formation of arbitrary bindings. If abstract concepts do indeed implicitly activate systematic, schema-based contextual information, this could explain why arbitrary episodic context tends to be inhibited for abstract concepts, but that when it comes from a class that tends to be informative, such as spatial location, context facilitated abstract concept recognition. Exploring the interplay between systematic and arbitrary contextual information—and the associated neural dynamics—is a crucial direction for future work, which we return to below.

This explanation requires that our intuition that location is more important than color or speaker voice for understanding the meaning of abstract concepts is correct. We provided evidence for this intuition by showing that, in a separate rating study, participants rated location as most relevant to interpreting the meaning of abstract concepts in the real world, followed by voice and then color. We also confirmed our intuition that for abstract concepts, voice is slightly more important than color, which might explain the null result observed in Experiment 4. This also seems intuitively reasonable given that social-communicative contexts are strongly associated with more abstract concepts (Barsalou & Wiemer-Hastings, 2005). Moreover, voice is a cue to identity, and when communicating with different people, having a nuanced understanding of *their* understanding of e.g. justice as distinct from *someone else's* understanding of justice, is critical. Hence the informativeness of voice. Given the hypothesized role of systematic, schema-based knowledge in understanding abstract concepts, the finding that location is most informative for abstract concepts is unsurprising—after all, the activation of schema-based knowledge depends on spatial qualities. Whatever the nuanced differences between one speaker's concept and another's, both likely depend on situationally determined location for their meaning (speakers are merely conduits for information that is a proxy for the direct experience of the spatiotemporally distributed cues that signal an instance of a concept).

Thus, our favored interpretation of the present findings is that abstract concepts activate systematic, schema-based contextual information, and when processing *decision*, the activation of such systematic information may in fact *inhibit* formation of arbitrary associations (van Kesteren et al., 2013; for discussion, see Davis et al., 2020). However, when contextual associations are *relevant* for recognizing (or understanding the meaning of) an abstract concept, those associations are better remembered. This would explain why our arbitrary episodic contexts were not well remembered for abstract concepts (Experiments 1 and 2) and why context retention may have in some cases even *inhibited* word recognition (Experiment 3). It would also explain why, when using a context to which abstract concepts are more sensitive in the real world (i.e., spatial location), retention in fact facilitated word recognition (Experiment 5).

Limitations and Next Steps

The synonym judgment task used at encoding may have worked to a disadvantage: as abstract concepts tend to have more diverse meanings, synonym judgments may be more difficult for abstract concepts, as it must be determined whether *any particular sense* of the word is a synonym to the target (Hoffman et al., 2013). Thus, an abstract concept like *decision* when paired with *judgment* might leave fewer resources available to process immediately available relational information (i.e., in the present study, the frame color or the voice) because we must search for a context in which *decision* and *judgment* are in fact synonyms (a recent computational model makes this prediction; Popov & Reder, 2020). Support for this account comes from the fact that memory for concrete synonyms tended to be particularly strong (see Supplementary Material for analysis of our non-target, synonym trials). However, it is worth noting that the same synonym judgment task in Experiment 5 resulted in a context-retention *advantage* for abstract concepts, and that a resource-limited account would not predict the context reinstatement *disadvantage* shown in Experiment 3 (and Experiment 4, though this effect was not statistically reliable).

It is also noteworthy that context reinstatement, for the most part, did not improve item recognition. This may be because we only used two contexts in Experiments 3 and 4—context-retention advantages may not be observed when the context is shared across too many items (Park et al., 2006). And indeed, a main effect for context retention did emerge in Experiment 5, where the concepts could occur in 1 of 4 locations on a screen. Nevertheless, with just two contexts (in Experiments 1–4), reinstatement still *impaired* item recognition for abstract words, implying that a context-retention *disadvantage* can be detected with only two contexts.

While further research is necessary to better understand the interaction between abstract concepts and arbitrary episodic contexts, it is unlikely that such work will be fruitful without nuanced consideration of what it *means* for a concept to be relatively concrete or abstract. One contributing factor to perceived abstractness may be the degree to which a concept is experienced as spread over space and time (Davis et al., 2020). That is, while a relatively concrete concept like *coffee* is experienced rapidly in a spatially circumscribed space, a relatively abstract concept like *justice* takes longer to unfold, and may comprise multiple spatially distinct events (e.g., bank is robbed, suspect apprehended, court appearance, sentencing). Attention to such events in comprehension necessitates apprehension of the systematic relations among events, perhaps at the expense of more arbitrary ones. Further exploring the spatiotemporal properties of concepts may shed further light on the present findings.

The present set of experiments also demand further work on the neurobiological mechanisms underpinning such effects. A key motivation for this set of experiments was the notion that relational binding—the process of binding contextual detail (e.g., a colored frame, or a spatial location) to a target stimulus (e.g., a picture, or a word) when encoding episodes in memory—might be the mechanism by which abstract concepts are sensitive to contextual information. Importantly, relational binding is subserved by the hippocampal system (Cohen & Eichenbaum, 1993). While Experiments 1–3 largely suggest that the hippocampal system might be inhibited when processing abstract concepts with arbitrary contexts, thus inhibiting relational memory (in line with Davis et al., 2020), Experiment 5 leaves open the possibility that abstract concepts do indeed engage hippocampal mechanisms when spatial location is invoked, perhaps because location typically situationally relevant when processing abstract concepts in the real world.

Conclusions

This work suggests that arbitrary episodic detail is better bound with concrete than abstract concepts. The encoding context facilitated recognition of abstract concepts only in a location-based context, perhaps because that episodic detail is more relevant to constraining abstract concept meaning in the real world. Abstract concepts rely on situational context for interpretation, and given that activation of situational information is known to inhibit formation of arbitrary associations (van Kesteren et al., 2013; for discussion, see also Davis et al., 2020), formation of arbitrary associations may often be *inhibited* in abstract concepts on account of implicit activation of such systematic, schema-based contexts. More broadly, the way in which the episodic memory system is recruited appears to differ as a function of concreteness, suggesting that engagement of the episodic memory system is modulated by semantic content. The episodic and semantic memory systems are not modular—this and an accumulation of work in recent years instead suggest an interactive, integrated memory system.

Acknowledgments

We would like to thank Emma Dzialo, Conor Hylton, and Isabella Caban for help with data collection. We would also like to thank Ken McRae for early discussions on experimental design, and members of the Yee and Altmann Labs—in particular, Yanina Prystauka, Zachary Ekves, Gitte Joergensen, Hannah Morrow, and Kyra Krass—for helpful discussions on this project over the years.

Open Practices Statement

None of the experiments were preregistered, but the data—including Supplemental Material—are openly available in an OSF repository (https://osf.io/jcmpr/?view_only=84a652fb36b74dd6b6b049e909ebd579).

References

Asmuth, J., & Gentner, D. (2017). Relational categories are more mutable than entity categories. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2007–2025.

Badre, D., & Wagner, A. D. (2002). Semantic retrieval, mnemonic control, and prefrontal cortex. *Behavioral and Cognitive Neuroscience Reviews*, *1*(3), 206–218.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(04), 577–660.

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170144.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). Cambridge, UK: Cambridge University Press.

Bartlett, F. C. (1932). Remembering: An experimental and social study. Cambridge, UK: Cambridge University.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *3294*, 1–45.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625–641.

Cohen, N. J., and Eichenbaum, H. (1993). *Memory, Amnesia and the Hippocampal System*. Cambridge, MA: MIT Press.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(3), 615–627.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700.

Davis, C. P., Altmann, G. T. M., & Yee, E. (2020). Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology*.

De Deyne, S. (2017). Mapping the lexicon using large-scale empirical semantic networks. Talk presented at the Annual Meeting of the Psychonomic Society, Vancouver, BC.

Eichenbaum, H. (2013). Memory on time. *Trends in Cognitive Sciences*, *17*(2), 81–88.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 5–16.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730.

Hoffman, P., Binney, R. J., & Ralph, M. A. L. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex*, *63*, 250–266.

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin & Review*, *23*(6), 1744–1756.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Kirwan, C. B., Wixted, J. T., & Squire, L. R. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *Journal of Neuroscience*, *28*(42), 10541–10548.

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. Consciousness and Cognition, 21(3), 1435–1455.

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, *31*(2), 237–260.

O’Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1196–1204.

Park, H., Arndt, J., & Reder, L. M. (2006). A contextual interference account of distinctiveness effects in recognition. *Memory & Cognition*, *34*(4), 743-751.

Popov, V., & Reder, L. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46.

R Core Team (2013). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Sarah, S. Y., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia*, *50*(13), 3070–3079.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), The psychology of word meanings (pp. 223–250). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82–102.

Smith, M. L., & Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, *19*(6), 781–793.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences*, *95*(26), 15855–15860.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychologia*, *41*(3), 280–292.

Tulving, E. (1983). Elements of episodic memory. Oxford, UK: Clarendon.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25.

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359.

Wattenmaker, W. D., & Shoben, E. J. (1987). Context and the recallability of concrete and abstract sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 140–150.

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, *119*(3), 889–905.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *The American Journal of Psychology*, *119*(3), 349–384.

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *356*(1413), 1363–1374.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Yu, S. S., Johnson, J. D., & Rugg, M. D. (2012). Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgments. *Hippocampus*, *22*(6), 1429–1437.

Appendix

Instructions for Color Importance Rating Task

You will be asked to think about the relationship between things and the color of their surroundings. The meanings of some things, like an apple, may depend little on the color of the surroundings. An apple in a blue setting may have the same meaning as an apple in a green setting. But other things, like dancing, might have a slightly different meaning depending on the color of the surrounding context.

Please tell us how much the meaning of each of the following things depends on the color of its surroundings. That is, how important is the color of its surroundings to its meaning? There are no right answers, so simply go with your first instinct.

Instructions for the Voice Importance Rating Task

You will be asked to think about the relationship between things and the voice of the speaker mentioning them. The meanings of some things, like an apple, may depend little on who is talking about them. The word apple when spoken in a woman's voice may have largely the same meaning as when spoken in a man's voice. Other things, like dancing might have a slightly different meaning depending on whether the voice is male or female.

Please tell us how much the meaning of each of the following things depends on the voice of the person speaking about it. That is, if someone talks to you about it, how important is their voice to its meaning? There are no right answers, so simply go with your first instinct.

Instructions for the Voice Importance Rating Task

You will be asked to think about the relationship between things and the locations that they appear in. The meanings of some things, like an apple, may depend little on the surrounding location. An apple in Canada may have the same meaning as an apple in Chile, and an apple on *the kitchen table may have the same meaning as an apple on the counter. Other things, like dancing, might have a slightly different meaning depending on the location.*

Please tell us how much the meaning of each of the following things depends on its location. That is, how important is its location to its meaning? There are no right answers, so simply go with your first instinct.