Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

8-12-2022 1:00 PM

The Spacing Effect in Remote Information-Integration Category Learning

Anthony Cruz, *The University of Western Ontario*

Supervisor: Minda, John Paul, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Psychology

© Anthony Cruz 2022

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Cognitive Psychology Commons, and the Cognitive Science Commons

Recommended Citation

Cruz, Anthony, "The Spacing Effect in Remote Information-Integration Category Learning" (2022). *Electronic Thesis and Dissertation Repository*. 8718.
https://ir.lib.uwo.ca/etd/8718

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

The present study examined whether the temporal distribution of procedural category learning experiences would impact learning outcomes. Participants completed the remote category learning study on a smartphone in one of two learning conditions: Massed (control) or distributed. Consistent with expectations, distributed learners reached higher accuracy levels. This effect disappeared after accounting for reaction time differences, suggesting that it was driven by attentional mechanisms. Distribution may have made participants more likely to discover the optimal categorization strategy and more robust to sensory habituation. Counter to previous findings, participants favored distributed learning. These results suggest that adult category learning is facilitated by temporal spacing. Future work may further explore the effects of temporal and contextual distinctiveness of learning experiences on category learning outcomes.

Keywords: Category learning, spacing effect, COVIS theory, metacognition

Summary for Lay Audience

Throughout life, people learn to sort items into categories to help them make sense of the world. People rarely spend long periods of time studying new categories; instead, categories are usually learned in short experiences spaced out over time. For example, children don’t study the differences between cats and dogs, they slowly learn to distinguish between them through experience. The goal of this study was to see if spacing out learning experiences over time would improve a person’s ability to sort imaginary items into abstract categories. Participants learned to sort items on a smartphone either all at once (massed) or in short sessions spaced out over several days (distributed). Distributed learners were better at sorting the items. Massed learners became less sensitive to the differences between items and paid less attention over time. Distributed learners were more satisfied and keener to learn again. Both types of learners indicated a preference for distributed learning if trained again in the future. Future research should see if this learning method is effective for real-world categories such as skin lesions, mushrooms, or animal groups.

Table of Contents

List of Figures

List of Appendices

Introduction

Categorization is the act of sorting stimuli into discrete equivalence classes using a many-to-one stimulus-response mapping (Kéri, 2003). There is evidence that many categories, such as colors and phonemes, have evolutionary roots (Harnad, 2003), but many categories that humans use are learned. Experimental research has led to the development of several theoretical models of category learning. Prototype theories assume that while learning a category, humans develop a sense of its central tendencies. These central tendencies are stored in memory as prototypes and category judgments for newly encountered stimuli are made based on their similarity to these prototypes. There is wide support for these models of category learning (Posner & Keele, 1968; Smith & Minda, 1998), but prototype theory alone cannot explain category learning. Exemplar theories suppose that humans store in memory individual instances of a category. During learning, they compare newly encountered stimuli to their memories of previously encountered stimuli. These new stimuli are assigned to the category with which all pairwise similarities are the highest (Medin & Schaffer, 1978; Nosofsky, 1986). Both prototype and exemplar models rely on the assumption that humans compute similarity between newly encountered stimuli and some internal memory representation(s).

Decision bound theories assume that category learning is the process of learning to partition a stimulus space (Ashby & Townsend, 1986). One influential decision bound theory is COVIS (COmpetition between Verbal and Implicit Systems) theory, which assumes that there are two competing neural systems controlling category learning (Ashby et al., 1998). The verbal, or explicit, system tests verbalizable hypotheses related to the category space, adjusting until a successful rule has been discovered. For a task to be learned by this system, the category structure must be defined by some verbalizable rule; such structures are called rule-based. The implicit, or procedural, system learns stimulus-response pairings through feedback-based associative learning. This system is necessary for learning difficult-to-verbalize, or information-integration, category structures (Ashby & Valentin, 2017). Information-integration structures require simultaneous attention to at least two dimensions of variability. Participants tend to employ linear, deterministic decision boundaries. General linear classifiers can be used to compute the most likely strategy a participant is employing (Ashby & Gott, 1988).

Learning Schedule and the Spacing Effect

Natural categories differ markedly from artificial categories in the way that they are learned. Natural category learning typically occurs in short, temporally distinct learning experiences while artificial category learning typically involves one or few long training sessions. These differences in the arrangement of learning experiences are differences in learning schedules (Simon, 2008). The temporal distance between two adjacent learning experiences is a spacing gap. When experiences have a spacing gap of 0, the learning schedule is said to be massed. Otherwise, it is distributed (or, equivalently, spaced). In addition to being a better reflection of naturalistic learning processes, spacing tends to yield stronger learning outcomes than massing. This trend, called the spacing effect, was first documented by Herman Ebbinghaus and has since been replicated for a number of tasks (Simon, 2008).

In one procedural learning study using a 2x2 design, humans learned to type and experienced equal amounts of training in 1 or 2 sessions per day, each 1 or 2 hours long. It was found that shorter and less frequent training sessions resulted in more accurate keystrokes per minute while longer and more frequent training resulted in a higher percentage of uncorrected errors (Baddeley & Longman, 1978). Learners with 2 2-hour sessions per day (2x2h learners) learned least effectively while 1x1h learners learned most effectively. Spacing is also positively impactful for children's learning novel grammatical constructions (Ambridge et al., 2006) and for students learning to interpret electrocardiogram readings (Monteiro et al., 2017). Though spacing is clearly beneficial for these and other procedural learning tasks, it is unclear whether spacing will be more effective than massing for procedural category learning.

Educational psychologist Ernst Rothkopf wrote in 1977 that "spacing is the friend of recall, but the enemy of induction" (as cited in Kornell & Bjork, 2008), induction being the application of an existing categorization strategy to a novel stimulus. Kornell and Bjork (2008) tested this assertion by conducting two experiments in which participants observationally learned to categorize 72 paintings according to their artist (of which there were 12), later taking an induction test featuring 48 new paintings. Participants in the massed learning condition learned paintings one artist at a time and those in the distributed condition learned in an interleaved sequence such that paintings by the same artist were presented far apart in time. Participants in experiment 1a received both conditions and those in 1b experienced only one condition. Counter to expectations, distributed learning was more effective than massed learning across all test blocks in both experiments. They claimed this to be an example of the spacing effect and similar results have been produced by similar studies (e.g., Guzman-Munoz, 2017; Kornell et al., 2010; Wahlheim et al., 2011). However, participants in this study never truly took any breaks from learning. Both learning conditions had identical spacing gaps; only stimulus presentation order differed. This study did not make it clear whether the observed effect was truly a spacing effect or if it was a distinct interleaving effect.

A later study by Kang & Pashler (2012) directly addressed this limitation. They conducted a category induction task with a category structure similar to that of Kornell & Bjork (2008). This study had four between-participant conditions: massed, interleaved, temporal spaced, and simultaneous massed. The former two conditions are identical to those used by Kornell & Bjork (2008). Temporal spaced learners had long periods of nothingness between painting presentations. Simultaneous massed learners viewed 4 paintings by the same artist at one time. Interleaving and simultaneous massed presentation led to stronger performances than the other groups, suggesting that the benefit of interleaving arises from the juxtaposition of different categories, which emphasizes inter-category differences, rather than temporal spacing (Kang & Pashler, 2012).

Following Kornell and Bjork's publication on the interleaving effect, Vlach, Sandhofer, and Kornell (2008) produced research that explored the potential for a spacing effect in children’s category induction. In a mixed 2x2 design, participants completed either a memory task or a category induction task and underwent both spaced and massed training. Spaced training involved a 30-second inter-stimulus interval during which participants were given a toy to play with. Spacing yielded better learning in both tasks and the magnitude of this effect was equal for both tasks. In a later study, Vlach, Ankowski, and Sandhofer (2012) compared immediate and 15-minute delay category induction test performance on a sample of 2 year-olds in either simultaneous, massed, or spaced learning conditions. Performance on the immediate induction test was highest for the simultaneous presentation condition but performance on the delayed induction test was highest for the spaced condition. Coupled with a later finding that spaced learners perform worse on retrieval tests conducted during learning (Vlach et al., 2021), this suggests that the benefits of temporal spacing only emerge after some delay.

Given the temporal distinctiveness of natural category learning experiences, it is possible that the spacing effect is at play during natural category learning. The studies by Vlach and colleagues demonstrate that this is a possibility in children, but it is unclear if these results would replicate in adults. The spacing effect does not facilitate word or grammar acquisition in adults as it does in children (Smith & Scarf, 2017) and the magnitude of the spacing effect, in general, may decrease as adults age (Simone et al., 2013). Additionally, all of the aforementioned studies of spacing in category induction have used spacing gaps on the order of seconds. These small spacing gaps were likely used for convenience, but this limits the generalizability of these results, as the magnitude of the spacing effect is dependent upon the size of the spacing gap (Cepeda et al., 2009) and spacing gaps of this magnitude do not reflect learning in naturalistic settings, wherein several hours may pass between adjacent learning experiences. The impact of larger spacing gaps in category induction has not been explored.

Prior results also ignore the fact that, in nature, temporal spacing often leads to contextual variety. Procedural category learning is negatively impacted by minute contextual changes. Crossley, Ashby, and Maddox (2014) showed this using a three-stage learning-unlearning-relearning study. They found that task-irrelevant changes in background color could impact participants' relearning of a procedural category structure. It was harmful for background colors during learning and unlearning to match while it was beneficial for background colors during learning and relearning to match. Social context may also be of importance; participants in one study learned more effectively in the presence of another human than alone with a computer (Stephens et al., 2010). Changing the motor responses associated with category judgments generally has a negative impact on performance (Hughes & Thomas, 2021). Whitehead, Zamary, and Marsh (2021) examined transfer of category knowledge from ideal contexts, in which stimulus perception is unobstructed, to impoverished contexts, in which some stimulus features are missing or obstructed. They did not directly compare performance between contexts (as this was not their concern in the study), but participants in the impoverished context appeared to perform worse.

Attentional Mechanisms

The attention attenuation hypothesis holds that the spacing effect occurs as a result of diminished attention over the course of massed study (Kornell et al., 2010). Attention may diminish because participants begin to feel that they are getting diminishing returns from continued practice. In two experiments employing a category structure similar to that used by Kornell & Bjork (2008), Wahlheim, Dunlosky, and Jacoby (2011) found strong support for the attention attenuation hypothesis. In their first experiment, they found that performance decreased across learning blocks when stimuli were presented in a massed, rather than an interleaved, fashion. In a second experiment, participants were given control over the duration of stimulus presentations. It was found that participants experiencing massed presentations spent less time studying than those experiencing interleaved presentations. Moreover, when study time was included as a covariate in a comparison of categorization performance between massed and interleaved presentations, the main effect of presentation style disappeared. Although this study analysed the interleaving effect, rather than the spacing effect, it is reasonable to expect that this same effect would hold for massed and distributed learning schedules; distributed learning may be superior to massed learning because participants gradually pay less attention over the course of long training sessions.

Deficient processing theory holds that the depth of stimulus processing decreases over the course of learning due to sensory habituation, a decreased response to frequent stimulus repetitions (Hintzman, 1974). Sensory habituation is closely linked to repetition suppression (Nordt et al., 2016), the phenomenon in which neurons show dampened responses to repeated presentations of information to which they are sensitive (Barron et al., 2016). In populations of neurons, dampened responses could indicate representational overlap or similarity. Sensory habituation may reasonably be expected to occur with repeated presentations of nonidentical, highly similar stimuli, such as those often used in category learning tasks. Kenney (2009) argued against this possibility, finding that the discriminability of stimuli does not impact the magnitude of the spacing effect as deficient processing theory would predict. Like previous studies, however, this study conflates the spacing effect with the interleaving effect. Distributed learning may be superior to massed learning because repeated stimulus presentations are processed less deeply during massed learning.

Memory-Based Mechanisms

Large spacing gaps between adjacent learning experiences may cause forgetting. The forgetting-as-abstraction hypothesis holds that forgetting may aid abstraction because irrelevant information is forgotten sooner than relevant information and cause the retrieval of learnt experiences becomes more effortful, creating a desirable difficulty in the learning process (Vlach, 2014). This hypothesis builds upon study-phase retrieval theory, holding that memory is ameliorated by the recall of past learning events (Thios & D’Agostino, 1976). The encoding variability hypothesis also builds upon study-phase retrieval theory, supposing that learning is more effective when more distinct retrieval cues are stored in memory. Though one might expect contextual variety to negatively impact category induction (e.g., Crossley et al., 2014), the effect of natural contextual variety on category learning outcomes has not been explored. Contextual variety may support memory by acting as a retrieval cue. Additionally, contextual variability causes context and stimulus features to become less correlated (Melton, 1970), making it easier to ignore decision-irrelevant information during subsequent category learning experiences. The forgetting of irrelevant information, more effortful retrieval, and contextual retrieval cues may be partially responsible for the spacing effect.

Categorical Perception

Learned categories confer many benefits, one being cognitive economy, the reduction of variability among stimuli to levels relevant for some purpose at hand (Rosch, 1975). Category judgments are guided by perceptual representations of stimuli, which can be optimized for categorization behavior by minimizing between-category similarity and maximizing within-category similarity (Hughes & Thomas, 2021). In a seminal study by Liberman, Harris, Hoffman, and Griffith, participants listened to plosives along a */b-d-g/* spectrum and classified them as b, d, or g. Then, participants completed an ABX discrimination procedure in which two stimuli (A and B) from different parts of the morph space were presented and participants needed to judge whether a third stimulus, X, was in the same category as A or the same category as B. There were clear locations along the phoneme spectrum at which participants' probability of making a phoneme classification jumped sharply. Around these locations, called category boundaries, discriminability was the highest (Liberman et al., 1957).

This phenomenon, categorical perception, has been defined many ways. Repp (1984) argued that, for stimuli varying along a continuum, categorical perception may be empirically observed when the probability of making a given category judgment changes dramatically at some point along the continuum (the category boundary). Goldstone (1994) later defined categorical perception as perception of the stimulus space such that relevant cues are emphasized (acquired distinctiveness) and irrelevant cues are de-emphasized (acquired equivalence). For stimuli defined along continuously varying dimensions, this would involve enhanced sensitivity to relevant directions of variation and less of sensitivity to irrelevant directions of variation. Harnad (2003) defined categorical perception more broadly as the compression of within-category differences and/or expansion of between-category differences. Categorical perception reduces the unnecessary variability among and enhances the relevant features of to-be-categorized stimuli.

Categorical perception can be induced for novel stimulus sets using artificial category learning procedures. Goldstone (1994) demonstrated this in a series of experiments. In the first two experiments, stimuli were 16 squares arranged in a 4x4 grid varying in size and brightness. Experiment 1 was used to adjust the stimuli such that adjacent stimuli at all points in the grid were equally discriminable. In experiment 2, there were three experimental groups in which participants categorized stimuli according to brightness, size, or both brightness and size. Experimental groups completed a feedback-based category learning procedure followed by a perceptual discrimination task; controls only completed the perceptual discrimination task. Discriminability in categorizers was compared against controls to seek evidence of acquired distinctiveness and acquired equivalence. Strong evidence of acquired distinctiveness was found for all dimensions in each experimental group. Acquired equivalence was found for size but not brightness. Experiments 3 and 4 were identical to 1 and 2 except that saturation was used instead of size. Acquired distinctiveness was again found for all stimulus dimensions, with effects appearing larger around category boundaries. No evidence was found for acquired equivalence, however, with irrelevant stimulus dimensions sometimes becoming enhanced. Despite inconsistent evidence for acquired equivalence, within-category compression can be elicited by learning (Livingston et al., 1998). Folstein, Palmeri, and Gauthier (2013) also found evidence for acquired distinctiveness, finding that stimulus pairs varying in the direction relevant to categorization were more discriminable and showed less repetition suppression than those varying in the irrelevant direction. In summary, categorical perception can be elicited by artificial category learning, with strong evidence supporting sensitization to relevant information and occasional evidence of desensitization to irrelevant information.

Although categorical perception can arise for artificially learned categories, there are certain stimuli that humans seem predisposed to perceive categorically. Older infants only stay attuned to phonemes used in their ambient language(s) (Kuhl et al., 2006; Werker & Tees, 1984), but infants below 6 months in age can discriminate amongst phonemes from many languages (Perszyk & Waxman, 2016), suggesting that humans have a natural propensity to categorically perceive speech. Humans likely have a similar predisposition for color perception, as nonhuman primates and pre-linguistic infants show signs of categorical perception of color (Ozturk et al., 2013; Zhang et al., 2021). Humans have a special predisposition to categorically perceive faces (Etcoff & Magee, 1992; Kanwisher et al., 1997; Kotsoni et al., 2001). Levin and Beale (2000) demonstrated that categorical perception could be induced for stimuli on an artificial face morph spectrum, but that that it was elicited more strongly for upright rather than inverted faces, reflecting humans' natural propensity for categorically perceiving faces. It is not clear how temporal spacing impacts the development of categorical perception. However, naturalistic learning experiences, defined by temporal distinctiveness, may be more effective than artificial category learning at inducing categorical perception.

Metacognition

As humans choose when and how they learn, a learning strategy cannot be effective if it is not desirable. It is important to assess learners' beliefs about and attitudes toward different learning strategies. These beliefs and attitudes, metacognitive beliefs, arise from learners' ability to monitor and evaluate their own learning processes (Mitsea & Drigas, 2019). Metacognitive beliefs guide learners toward learning strategies that they feel are optimal (Metcalfe, 2009; Morehead et al., 2017). Flavell (1979) argued that metacognitive belief consists of three primary subcomponents: Knowledge of person, knowledge of task, and knowledge of strategy. A learner's knowledge of person consists of their perceptions of self and others engaging in a learning process, including intraindividual differences, interindividual differences, and cognitive universals. A learner's knowledge of task includes task demands, or what the learner will need to do, as well as the available information during the task. Knowledge of strategy, most relevant to the current study, concerns the effectiveness of different strategies for a given undertaking. Metacognitive beliefs, especially knowledge of strategy, impact a learner's choice of learning schedule.

In their study of how training schedule impacts keyboarding skill acquisition, Baddeley and Longman (1978) followed participant's final training sessions with a 3-question multiple choice survey asking how satisfactory they found their training schedule, which schedule they would prefer if they were trained again, and how keen they would be to undergo training again on the same schedule. Despite having the best rate of learning, 1x1h learners indicated lower satisfaction and lower keenness. 2x2h learners indicated the highest levels of satisfaction and keenness despite showing the poorest performance. This misalignment between participants' performance and their metacognitive beliefs, metacognitive incongruence, has been observed in a number of other studies. In their seminal paper on the interleaving effect, Kornell and Bjork (2008) conducted a within-subjects comparison of interleaved and massed study in a category induction task. The majority of participants performed better on the interleaved presentation schedule but tended to believe that they did better on the massed presentation schedule. Prior to this, Simon and Bjork (2001) identified a similar bias in a study of motor learning; participants learning on a massed schedule performed worse and greatly overestimated their performance on a delayed test. These metacognitive incongruences do not necessarily improve with experience (Tauber & Dunlosky, 2015). A learner's knowledge of strategy often fails to reflect the effectiveness of optimal learning strategies, with suboptimal strategies often being evaluated as more effective or favorable.

This pattern of results may arise from the challenges associated with implementing optimal study techniques or from participants' pre-existing beliefs. Challenges associated with implementing optimal study techniques, termed desirable difficulties, enhance long-term learning but inspire a sense of challenge that may deter participants from using them (Bjork et al., 2013). The accessibility bias inherent in repetition tasks may inspire a sense of processing fluency, making learners feel that they are learning most effectively when they engage in long training sessions (Doyle & Hourihan, 2016; Wahlheim et al., 2012). Spaced learning is challenging in that participants must, during each learning session, readjust to the learning process. Pre-existing beliefs about massed and distributed learning also play a role. Vlach, Bredemann, and Kraft (2019) found that preschool-aged children believe massed and distributed learning to be equal in effectiveness, but that 6-10 year-olds showed a bias toward massed learning. They argue that this bias may be learned from the adults around them; teachers and parents most likely encourage massed learning, as it makes them appear more studious. Wang and Xing (2019) found in a study comparing interleaved and massed learning that participants' metacognitive illusion disappeared when they were given a description of why their performance was stronger in the interleaved study condition. Since learners choose how and when they learn, learning strategies should be implemented in a way that minimizes the potential for metacognitive incongruence.

The Current Study

The goal of the current study was to test the effectiveness of an ecologically valid category learning paradigm and to determine the role of learning schedule in that paradigm's outcomes. Participants were assigned to either a massed or a distributed learning schedule. They completed the study remotely on their smartphones, allowing for ease of use and, for distributed learners, contextual variety. An information-integration category structure was used to mimic the difficult-to-verbalize nature of many natural categories. Several studies have found support for a spacing effect in category induction (Vlach et al., 2012, 2021, 2008), but these studies use observational learning, 30-second spacing gaps, and child participants. In the present study, I sought evidence for the spacing effect in feedback-based category induction in an adult sample with more substantial spacing gaps. Additionally, rather than an inter-stimulus interval, spacing gap here is defined as the time between adjacent study sessions. The primary outcomes of interest were task performance (accuracy), categorical perception, and metacognition. I hypothesized that participants' accuracy in the information-integration category learning task would increase over the course of training but that distributed learners would reach higher overall accuracy levels compared with those who underwent massed learning. This finding would reflect a spacing effect.

Participants' perceptions of the stimulus space were captured before and after learning using a similarity judgment task. It was hypothesized that both groups would show within-category compression, between-category expansion, acquired distinctiveness, and acquired equivalence after learning, as these outcomes have been found in previous artificial category learning paradigms (Folstein et al., 2013; Goldstone, 1994; Livingston et al., 1998). The impact of temporal spacing on the development of categorical perception has not yet been explored. Building upon the forgetting-as-abstraction hypothesis, I expected the magnitude of categorical perception effects to be larger in the distributed learners, as the forgetting between learning experiences in this group may have assisted in de-emphasizing category-irrelevant information. The final outcome of interest was metacognition. Participants' attitudes towards their learning paradigms were assessed using an adaptation of Baddeley and Longman's (1978) 3-question post-study survey. It was hypothesized that participants would be biased toward the suboptimal massed learning schedule, reflecting the metacognitive incongruence found in past studies (Baddeley & Longman, 1978; Kornell & Bjork, 2008; Simon & Bjork, 2001).

Methods

Participants

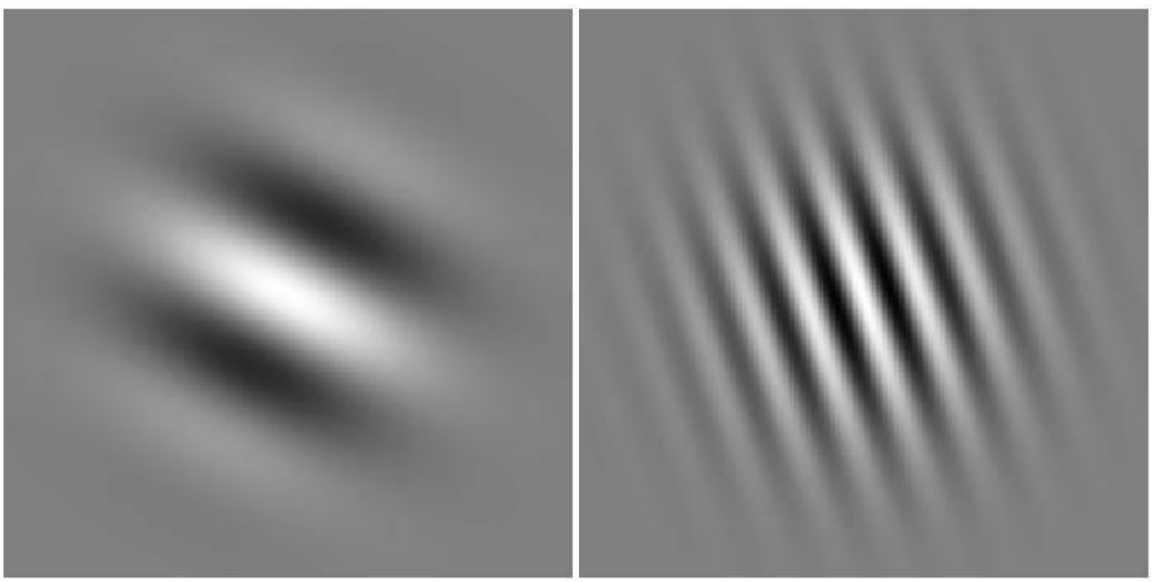
Participants in this study were recruited from Prolific between the 13th of April and the 16th of May 2022. A total of 104 participants consented to participate in the final version of this study. After consenting, participants were randomly assigned to either the massed or distributed learning condition. Participants were paid a total of 11.97 GBP for their participation. Participants spoke fluent English, owned a smartphone with internet access, and had normal or corrected-to-normal vision. All experimental procedures and materials were approved by the Western Research Ethics Board (see Appendices). After excluding participants who did not complete the entire experiment or who completed different components out of order, a final sample size of 96 was obtained. Participants had a median age of 23 (IQR: [21, 26]). Age did not differ significantly between learning conditions.

Stimuli

Stimuli were generated using the GRT package (Matsuki, 2017) in R version 4.1.1 (R Core Team, 2021). Stimuli were grayscale Gabor patches with varying spatial frequencies (f) and orientation angles (θ). For category learning, these parameters were sampled from two multivariate Gaussian distributions with equal covariance matrices (such that the Pearson correlation between f and θ was $r = .78$ in both groups) to generate 128 unique stimuli (See Figure 1 for examples). There were 64 stimuli in each category, arbitrarily labelled as A and B. The mean f for each group was $(\mu_A, \mu_B) = (11, 17)$. The mean θ (in degrees relative to vertical) was $(\mu_A, \mu_B) = (81, 64)$. We chose these distributions such that a deterministic information-integration category boundary could be drawn. This type of category structure is difficult to verbalize and is resilient to taxes on working memory, stress, and sleep deprivation (Hughes & Thomas, 2021). The deterministic nature of the boundary makes it possible (albeit unlikely) for participants to derive a general, 100% accurate categorization strategy. The linear category boundary is given by the equation $\theta = 30.89f + 2.54$. This line intersects the x-axis (f) at an angle of $68.54°$. The maximum possible accuracy attainable using single-dimensional rule-based strategies was 73.44% using an f-based strategy or 69.53% using a θ-based strategy.

Figure 1

Sample Category Learning Stimuli



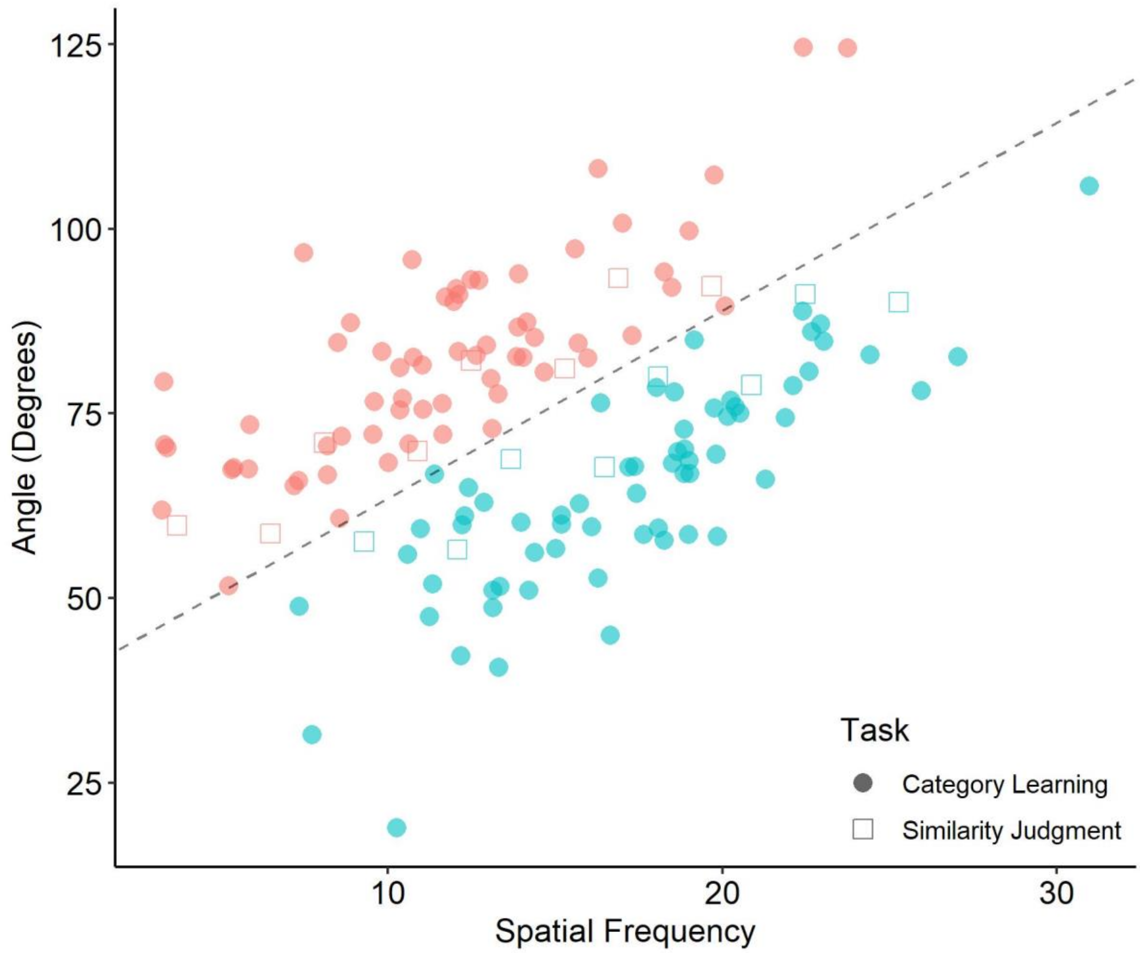
Note. On the left is an exemplar from category A with $f = 3.24$ and $\theta = 61.87°$. On the right is an exemplar from category B with $f = 10.23$ and $\theta = 18.89°$. These category A and B exemplars represent, respectively, the minimum spatial frequency and orientation angle in the stimulus set.

16 unique stimuli were generated for the similarity judgment task. A 4x4 grid was generated by creating all possible ordered pairs (x, y) such that $x \in \{\pm 18, \pm 6\}$ and $y \in \{\pm 4.5, \pm 1.5\}$. These values were chosen such that the grid would be centered at the origin. Adjacent pairs at a given y value were separated

by a distance of 12 while adjacent pairs at a given x value were separated by a distance of 3. The grid was represented as a 16x2 matrix and then rotated $68.54°$ relative to the x-axis using a rotation matrix. The $68.54°$ angle of rotation set the former horizontal axis of the grid parallel to the category boundary. This rotated grid was then recentered at the point $(f, \theta) = (14.48, 74.94)$, a point which lies on the category boundary. This spatial frequency is the mean of the most optimal f-based boundary ($f = 14.48$) and the f value associated with the intersection of the category boundary and the most optimal θ-based boundary ($\theta = 74.03$). The chosen orientation angle was that orientation angle on the category boundary associated with the chosen spatial frequency. Figure 2 shows the entire stimulus space for the study. Images used in both tasks were 352 x 352 pixels in size. The task was programmed such that the size of the stimuli would scale according to the size of each participant's smartphone screen. Testing was primarily done on a Google Pixel 6 and with Google Chrome's developer tools, which allow webpages to be viewed in a variety of screen dimensions.

Figure 2

Stimulus Space



Note. The dashed line indicates the optimal decision boundary. Category A and B exemplars are colored red and blue, respectively.

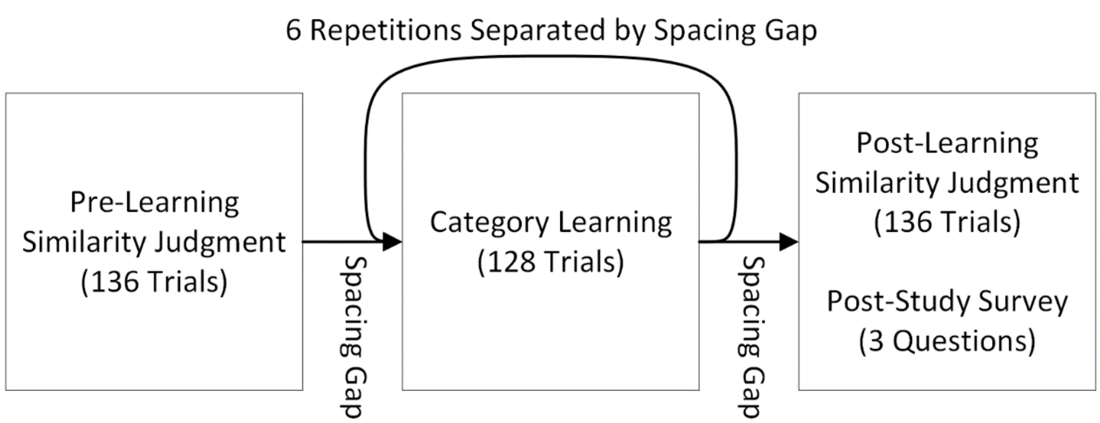
Procedure

Participants were recruited via Prolific. They first completed a Qualtrics form indicating their consent to participate. Then, they were then randomly assigned to a study condition and invited, via Prolific, to complete study components corresponding to their assigned condition. The experiment consisted of two similarity judgment tasks, six blocks of category learning tasks, and a 3-

question multiple-choice survey. Massed learners (controls) completed each study component in one session. Distributed learners completed one study component per session, resulting in eight sessions. They were instructed to leave 6-18 hours between each study component. The first and last study components were expected to be 15 minutes in duration and the remaining components were expected to be 10 minutes in duration, resulting in a total expected time commitment identical to that of the massed learners. See Figure 3 for a flowchart depicting each participant's progression through the study. Each experimental task was programmed in jsPsych 6.0.0 (de Leeuw, 2015) and hosted on Pavlovia.org.

Figure 3

Study Progression Flowchart



Note. Massed learners had a spacing gap of 0; no breaks were given between adjacent tasks. Distributed learners were instructed to take breaks (spacing gaps) of 6-18 hours between adjacent tasks. In both conditions, there was no break between the post-learning similarity judgment task and post-study survey.

Pre-Learning Similarity Judgment

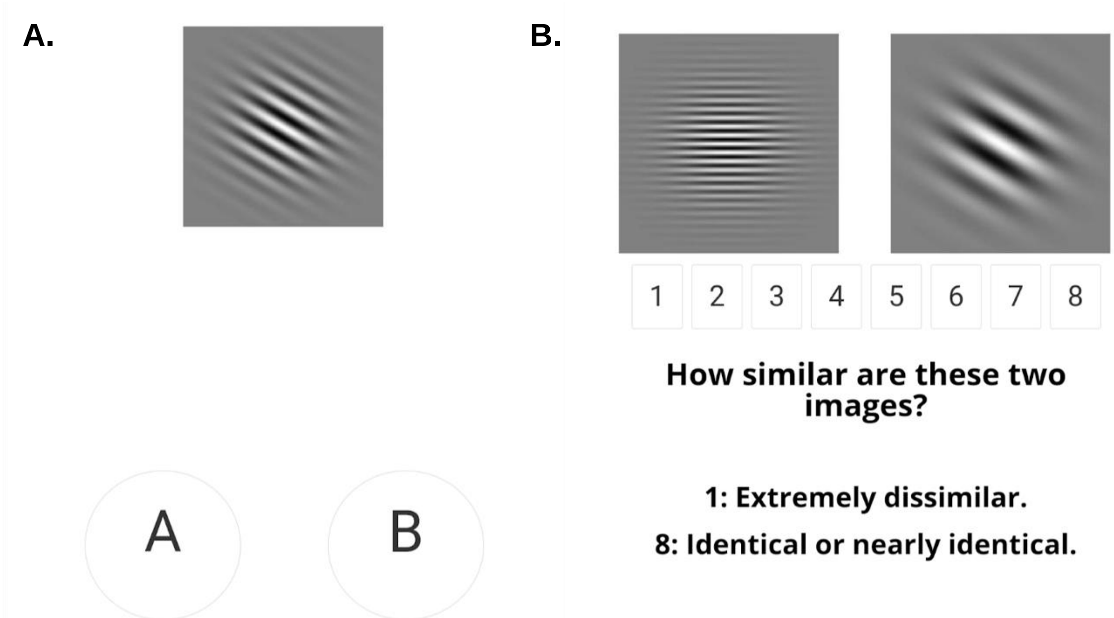
Participants first completed a similarity judgment task. Participants saw two stimuli appear side-by-side (in a randomized order) in the center of their phone screen and were asked to evaluate how similar they appeared to be. Images were programmed to each take 45% of the width of their container; this made the images scale to each participant's screen size. Participants used a 1-8 Likert scale to provide their responses and a 10-second time limit was imposed for each trial. There were constant on-screen instructions stating that 8 should refer to pairs that are "identical or nearly identical" and 1 to pairs that are "extremely dissimilar." Participants were shown all 136 possible pairs of similarity judgment stimuli. Of these pairs, 16 were identical, 64 were between-category, and 56 were within-category. In the distributed condition, this task alone was the first study component.

Category Learning

Participants were asked to complete 6 blocks of category learning with 128 trials per block. In the distributed condition, one block constituted one study component. Participants were instructed to sort stimuli into category A or category B by pressing the corresponding A and B buttons at the button of the screen. Each trial of learning began with 500ms of fixation. The stimulus would then appear until the participant made a judgment or until 10s had passed. This was followed immediately by 700ms of corrective feedback. If participants failed to provide a response in time, they were asked to respond more quickly next time. Figure 4 shows the interfaces for the category learning and similarity judgment tasks.

Figure 4

Experimental Task Interfaces



Note. On the left (A) is a screenshot of one category learning trial. On the right (B) is a screenshot of one similarity judgment trial. Whitespace above and below the task interfaces has been cropped for simplicity. Screenshots were captured on a Google Pixel 6; interfaces may have scaled differently on different screens.

Post-Learning Similarity Judgment

The procedure for this task is identical to that used in the pre-learning similarity judgment task. In the distributed condition, this task (alongside the post-learning survey) comprised the eighth and final study component.

Post-Learning Survey

After completing the post-learning similarity judgment task, participants were given the following description of the study conditions:

> In this study, you experienced one of two possible training schedules.
> **Schedule A** involved back-to-back completion of the category learning sessions.
> **Schedule B** involved completion of one session at a time over the course of several days.
> Please answer 3 questions about your experience and press "Submit."

The questions asked participants about how satisfactory they found their learning schedule (satisfaction), which schedule they would choose if they were to participate again (preference), and how keen they would be to undergo training again in the same schedule (keenness). Preference was a dichotomous choice between Schedule A and Schedule B. Satisfaction and keenness were assessed on a 5-point Likert scale. These questions are adaptations of those used by Baddeley and Longman (1978).

Data Analysis

All analyses were conducted in R version 4.1.1 (R Core Team, 2021).

Category Learning

For each block of category learning, participants' accuracy (proportion of learning trials that were answered correctly) and mean reaction time (RT) were recorded. Null trials (trials that were skipped) were ignored in both of these computations. A 6x2 (block x condition) mixed factorial analysis of variance (ANOVA) was conducted to determine whether accuracy changed significantly across the various blocks of learning, whether accuracy differed between the two conditions, and whether these two factors interacted. A similar analysis was conducted with RT as the dependent variable. We recorded for each block of learning whether a participant had surpassed the maximum reasonable accuracy attainable with a random guessing strategy (60.13%) and whether they had surpassed the maximum possible accuracy attainable under a unidimensional rule-based learning strategy (73.44%). The maximum reasonable accuracy attainable with a random guessing strategy was determined by taking the 99th percentile of 10,000 Monte Carlo simulations. Fisher's exact test was used to see if these classifications differed according to learning condition during the first and last blocks of learning.

A general linear classifier was fit for a single dimensional θ-based model, single dimensional f-based model, and an information-integration model for each participant at each block of learning. Akaike's information criterion (AIC) was computed for each model fit to determine which of the three models best explained participants' categorization strategy during each block of learning. For the first and final blocks of learning, Fisher's exact test was conducted with categorization strategy and learning condition as the grouping variables.

Similarity Judgment

Similarity judgments were converted into dissimilarities. We did this conversion using the following mapping: $D_{i,j}^2 = S_{i,i} + S_{j,j} - 2S_{i,j}$, where $S_{i,j}$ and $D_{i,j}$ respectively denote the similarity judgment for and dissimilarity between stimuli *i* and *j*. This mapping has the property that the dissimilarity between any stimulus and itself is zero (Buja et al., 2008). This mapping relies on the assumption that identical stimulus pairs will receive maximal similarity judgments. Any trials for which $D_{i,j}^2 < 0$ were considered invalid and excluded from analysis.

Multidimensional Scaling (MDS) was used to visualize and qualitatively assess the perceptual space. An MDS solution was computed using average pairwise dissimilarities before and after learning for each learning condition, resulting in 4 unique MDS solutions. It was expected that some categorical clustering would emerge after learning for both learning conditions. Consistent with prior research on categorical perception, it was expected that this clustering would emerge due to an within-category compression and expansion in the categorization-relevant direction of variability (Folstein et al., 2013; Goldstone, 1994; Livingston et al., 1998). The magnitude of these effects was expected to be greater in the distributed learners.

A representational dissimilarity matrix (RDM) was constructed for each participant at each time point. These RDMs were tested against a categorical model RDM and a physical distance model RDM computed using the L_1 metric (Manhattan distance). The L_1 metric was chosen due to the separable nature of the stimulus features (Soto & Wasserman, 2010). We conducted a Spearman correlation between each participant's vectorized RDMs and each vectorized model RDM. We compared fits for both of these models between both groups at both time points in a 2 x 2 (time point x condition) mixed factorial ANOVA. Based on the results of this ANOVA, model fits were compared against 0 and against each other. It was expected that L_1 model fits would decrease after learning and that categorical model fits would increase after learning. It was also expected that distributed learners would have stronger categorical model fits and weaker L_1 model fits after learning.

Similarity judgments were classified as either identical, within-category, or between-category. As their dissimilarities were set to 0, identical pairs were excluded from analysis. A 2x2x2 (learning condition x time point x pair type) ANOVA was conducted with dissimilarity as the dependent variable. It was expected that there would be no main effect of learning condition or time point, but that between-category pairs would be more dissimilar than within-category pairs. Additionally, an interaction between pair type and time point was expected such that between-category pairs became more dissimilar after learning and within-category pairs became less dissimilar.

A subset of 24 adjacent stimulus pairs were classified according to direction of variation and location within the stimulus space. Direction of variation was either relevant, varying in the direction perpendicular to the category boundary, or irrelevant, varying in parallel with the category boundary. Location was either outer, not crossing the center of the stimulus set, or inner, crossing the center of the stimulus set. Relevant-inner pairs crossed the true category boundary. This classification system is similar to that used by Folstein, Palmeri, and Gauthier (2013). Change in dissimilarity was computed for each pair. A 2x2x2 (learning condition x direction of variation x location) mixed factorial ANOVA was conducted with change in dissimilarity as the dependent variable. It was expected that change would be larger for pairs varying in the relevant direction of variation but that there would be no effect of learning condition or location.

Post-Learning Survey

Question 1 (satisfaction) was analyzed using an independent samples t-test with condition as the grouping variable. Responses were coded on a Likert scale with scores ranging from 0 to 4 (inclusive). This test was used to determine whether satisfaction differed between the two groups. The same analysis was conducted for question 3 (keenness). Question 2 (preference) was first analyzed used a Chi-Square goodness-of-fit test with preference as the grouping variable. Then, it analyzed using a Chi-Square test of independence with learning condition and preference as the grouping variables. This test was used to determine if participants preferred one schedule over another and whether this preference is affected by their assigned learning condition. It was expected that satisfaction and keenness would be higher in the massed condition and that all learners would tend to prefer massed learning.

Results

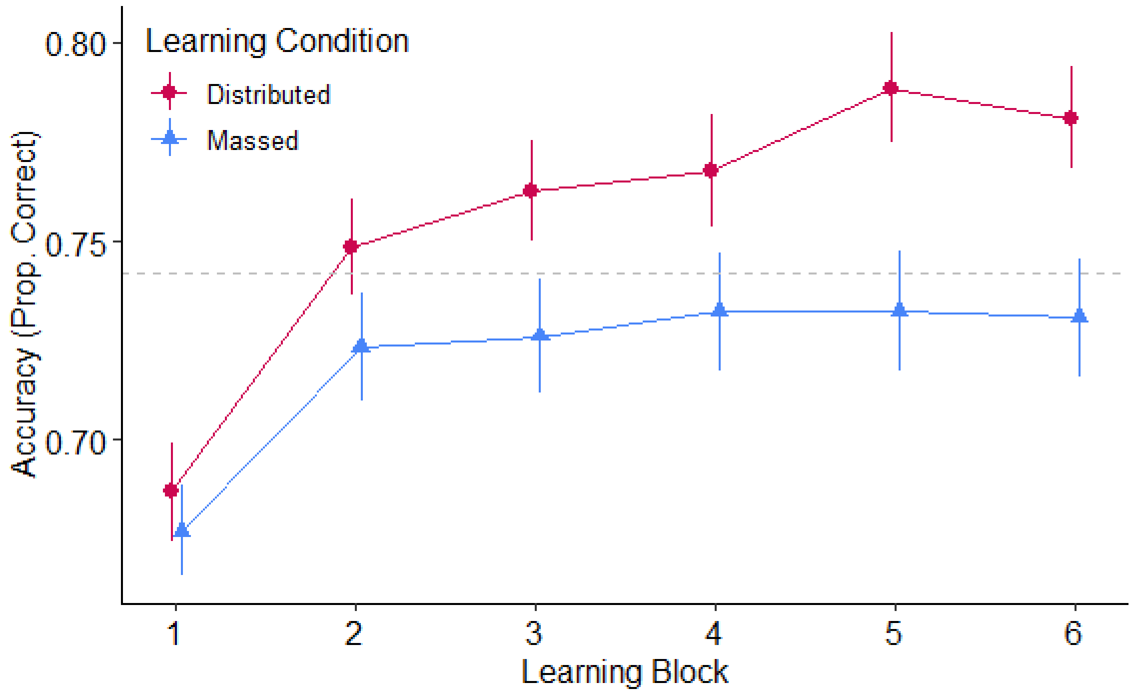
Participants who did not complete all category learning blocks or who completed blocks out of order were removed from all analyses. The final sample size was $N = 96$ ($n_{massed} = 49$, $n_{distributed} = 47$). All statistical analyses were conducted using a significance threshold of $\alpha = .05$ unless otherwise stated. There was some variability in distributed learners' spacing gaps. A one-way within-participants ANOVA was conducted to determine if distributed learners' spacing gaps varied over the course of training. Spacing gaps did not change significantly over the course of training, $F(4, 184) = 1.272, p = .282$. The mean spacing gap for distributed learners was 13.40 hours ($SE = 0.04, 95\%\ CI$: $[0.13, 27.78]$).

Category Learning

A 6x2 (block x learning condition) mixed factorial ANOVA was conducted with accuracy as the dependent variable (see Figure 5). The main effect of block on accuracy was significant, $F(5, 470) = 25.255, p < .001, \eta_p^2 = .212$. Accuracy increased from block to block. The main effect of learning condition on accuracy was significant, $F(1, 94) = 4.857, p = 0.03, \eta_p^2 = .049$. Participant accuracy was higher in the distributed condition ($M = .76, SE = .01, 95\%$ CI: $[.56, .91]$) than in the massed condition ($M = .72, SE = .01, 95\%$ CI: $[.48, .89]$). The interaction between condition and block was not significant, $F(5, 470) = 2.162, p = .057, \eta_p^2 = .022$. There was no effect of learning condition on accuracy during the first block of learning, $t(93.093) = .581, p = .563, d = .119$.

Figure 5

Accuracy over Time

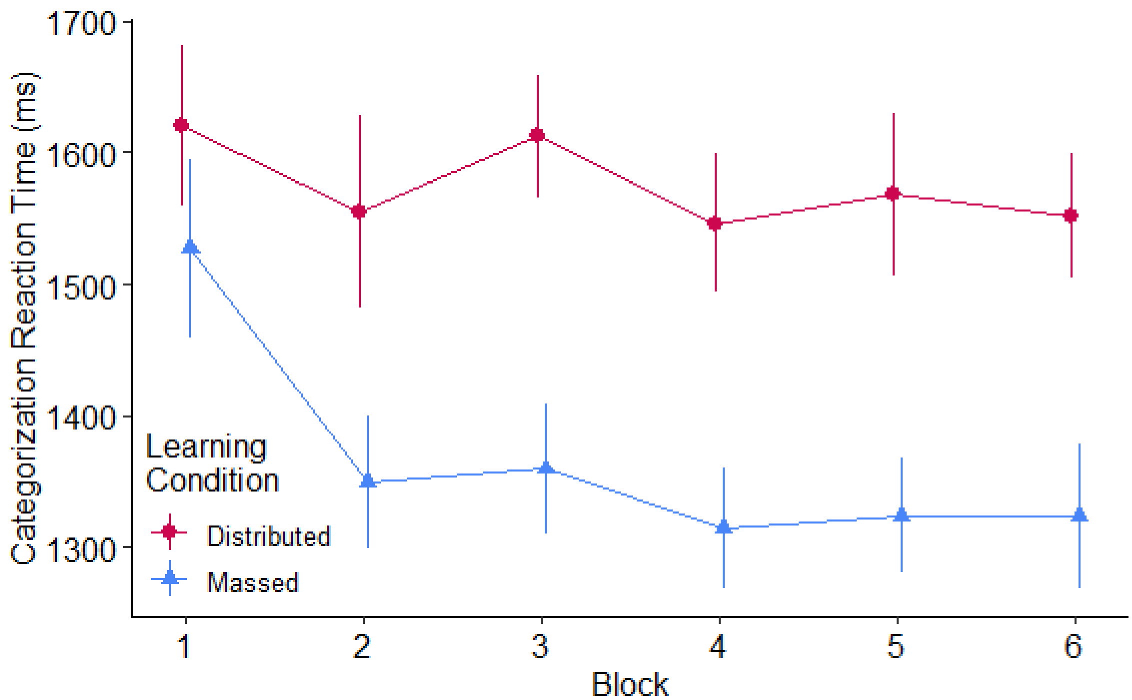


Note. Vertical bars represent standard errors. The dashed line represents the maximum accuracy attainable using a single-dimensional rule-based categorization strategy (73.44%).

A 6x2 (block x learning condition) mixed factorial ANOVA was conducted with mean reaction time (in milliseconds) as the dependent variable (see Figure 6). The main effect of block on reaction time was significant, $F(5,470) = 5.523, p < .001, \eta_p^2 = .055$. Reaction times decreased from block to block. The main effect of learning condition on reaction time was significant, $F(1,94) = 10.255, p = .002, \eta_p^2 = .098$. Reaction time was higher in the distributed condition ($M = 1576, SE = 23, 95\%$ CI: $[961, 2411]$) than in the massed condition ($M = 1366, SE = 22, 95\%$ CI: $[834, 2132]$). The interaction between condition and block was not significant, $F(5,470) = 1.635, p = .149, \eta_p^2 = .017$. There was no effect of learning condition on reaction time during the first block of learning, $t(93.298) = 1.033, p = 0.304, d = 0.21$.

Figure 6

Reaction Times over Time

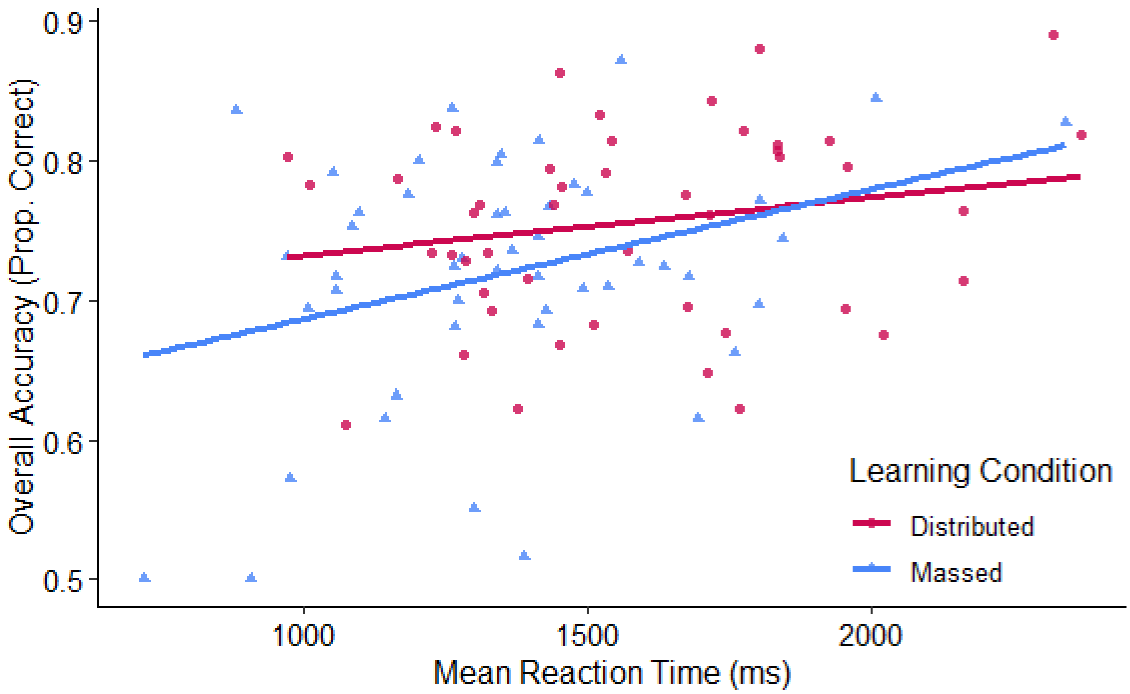


Note. Vertical bars represent standard errors.

Within each condition, an exploratory Pearson correlation between overall accuracy and mean reaction time was computed. There was a significant positive correlation in the massed condition, $r(47) = .328, p = .021$. There was no significant correlation in the distributed condition, $r(45) = .199, p = .179$. After controlling for the correlation between accuracy and reaction time, the main effect of condition disappeared, $t(92.484) = 1.271, p = .207$.

Figure 7

Accuracy as a Function of Reaction Time

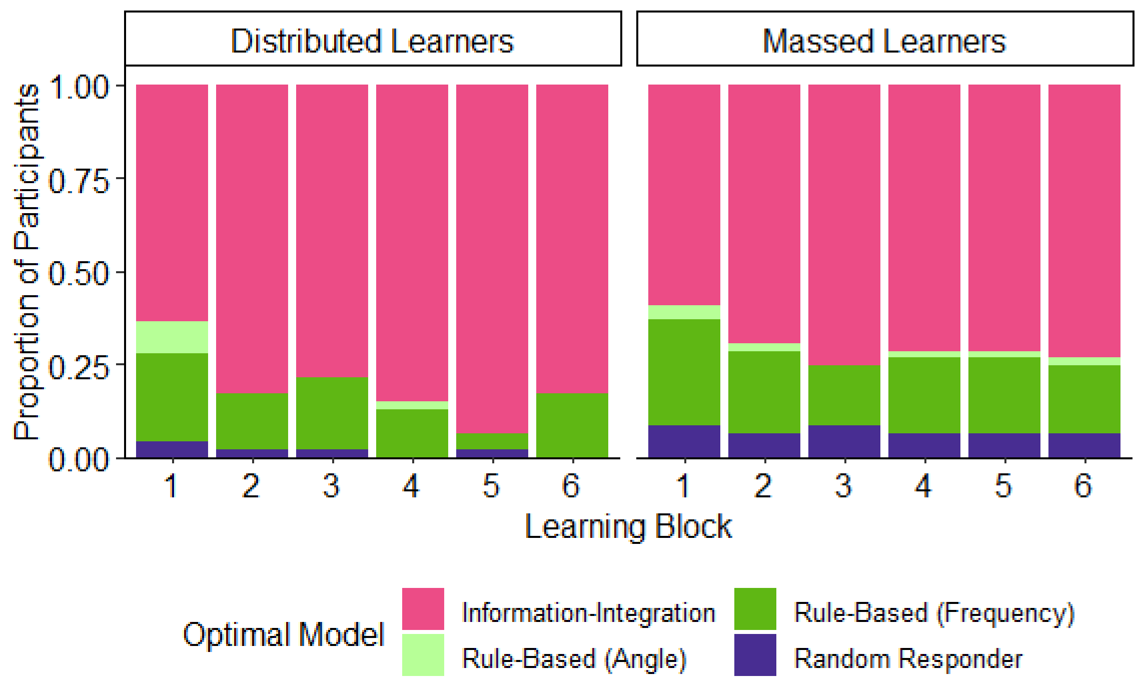


Fisher's exact test was used to determine if there was a relationship between learning condition and strategy use in the first and final blocks of learning. There was no significant association between learning condition and above-chance performance during the first learning block, $p = .606$. 81.3% of participants performed above chance levels during the first learning block. There was no significant association between learning condition and above-chance performance during the final learning block, $p = .269$. 91.7% of participants performed above chance levels during the final learning block. There was a marginally significant association between learning condition and above-criterion performance during the first learning block, $p = .06$. Above-criterion performance was more likely for distributed learners (34%) than for massed learners (16.3%).

There was a significant association between learning condition and above-criterion performance during the final learning block, $p = .042$. Above-criterion performance was more likely for distributed learners (68.1%) than for massed learners (47%). There was no significant association between learning condition and strategy use during the first learning block, $p = .675$. There was no significant association between learning condition and strategy use during the final learning block, $p = .296$. Figure 8 shows participants' strategies over the course of learning.

Figure 8

Categorization Strategies across Learning



Data were subset to only include participants who were best fit by an information-integration model (the optimal model) during the final block of learning. This subset corresponds to the proportion of participants shown in pink in Figure 8. A 6x2 (block x learning condition) mixed factorial ANOVA was conducted with accuracy as the dependent variable. The main effect of block on accuracy was significant, $F(5, 365) = 35.098, p < .001, \eta_p^2 = .325$. Participants' accuracy increased over the course of learning. The main effect of learning condition on accuracy was not significant, $F(1, 73) = 1.744, p = .191, \eta_p^2 = .023$. The interaction between block and learning condition was not significant, $F(5, 365) = 1.582, p = .164, \eta_p^2 = .021$.

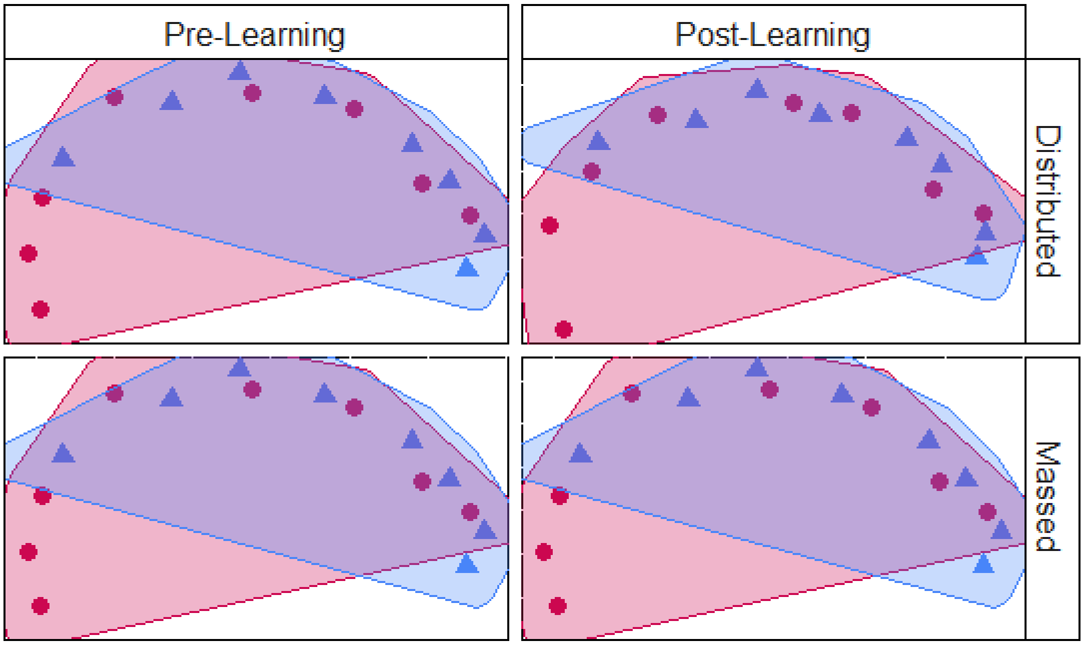
A 6x2 (block x learning condition) mixed factorial ANOVA was conducted with Akaike's Information Criterion (AIC) as the dependent variable. The main effect of block on AIC was significant, $F(5, 365) = 36.018, p < .001, \eta_p^2 = .33$. Participants' AIC decreased over the course of learning. The main effect of learning condition on AIC was not significant, $F(1, 73) = .001, p = .973, \eta_p^2 < .001$. The interaction between block and learning condition was not significant, $F(5, 365) = .82, p = .536, \eta_p^2 = .011$.

Similarity Judgment

Figure 9 shows the multidimensional scaling solutions for each time point and learning condition. It appears that there are no substantial qualitative differences among the four plots.

Figure 9

Multidimensional Scaling Solutions



Note. Scales are in arbitrary units and are fixed across all four plots.

Each participant's dissimilarity data at each time point were correlated with an L_1 physical distance model and a categorical model. The resulting Spearman correlation coefficients were normalized using Fisher's Z-Transformation prior to analysis.

A 2x2 (learning condition x time point) mixed factorial ANOVA was conducted with the L_1 RSA model fit (Z-Transformed Spearman's r) as the dependent variable. The main effect of learning condition was not significant, $F(1, 94) = 0.537, p = .465, \eta_p^2 = .006$. The main effect of time point was not significant, $F(1, 94) < .001, p = .983, \eta_p^2 < .001$. The interaction between condition and time point was significant, $F(1, 94) = 6.382, p = .013, \eta_p^2 = .064$. Bonferroni-corrected post-hoc testing revealed no significant pairwise differences, all $p's > .0125$.

A 2x2 (learning condition x time point) mixed factorial ANOVA was conducted with the categorical RSA model fit (Z-Transformed Spearman's r) as the dependent variable. The main effect of learning condition was not significant, $F(1, 94) = .157, p = .692, \eta_p^2 = .002$. The main effect of time point was not significant, $F(1, 94) = .054, p = .817, \eta_p^2 = .001$. The interaction between condition and time point was not significant, $F(1, 94) = 2.904, p = .092, \eta_p^2 = .03$.

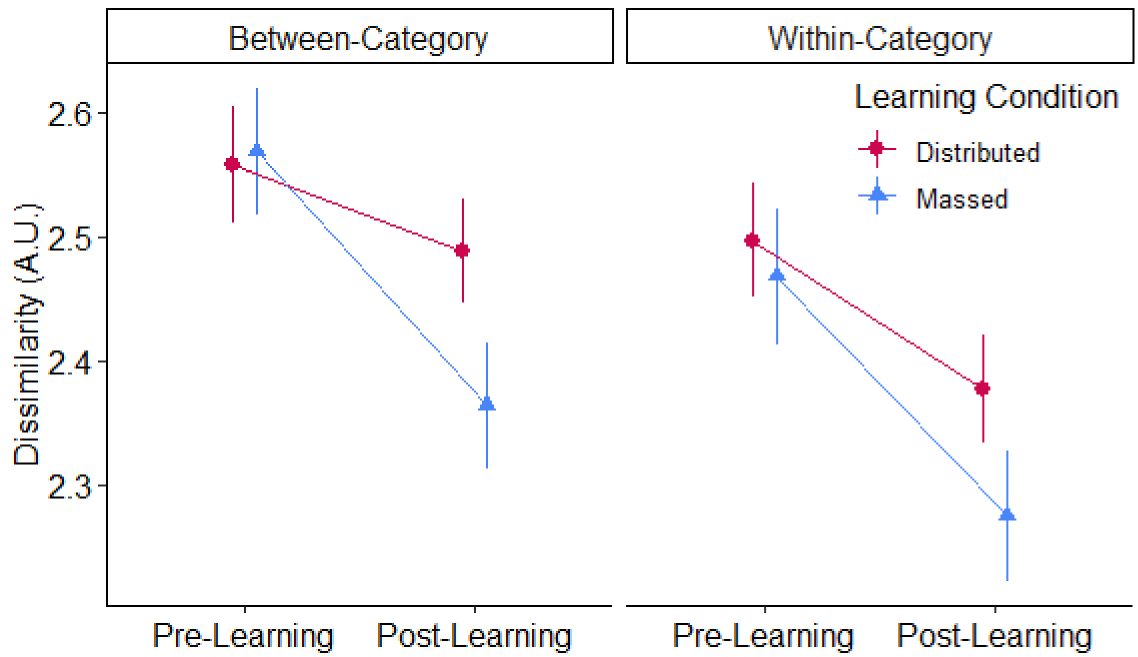
Both model fits were compared to 0 using a one-sample t-test. L_1 model fits were significantly greater than 0, $t(191) = 68.268, p < .001, d = 4.927$. Categorical model fits were significantly greater than 0, $t(191) = 58.984, p < .001, d = 4.257$. These model fits were compared to each other using a Welch two-sample t-test. L_1 model fits ($M = .671, SE = .0005, 95\%$ CI: $[.481, .817]$) were significantly stronger than categorical model fits ($M = .221, SE = .0003, 95\%$ CI: $[.124, .304]$), $t(228.11) = 47.52, p < .001, d = 4.85$.

A 2x2x2 (learning condition x pair type x time point) mixed factorial ANOVA was conducted with dissimilarity as the dependent variable. Since the dissimilarities for all identical stimulus pairs were set to 0, identical pairs were excluded from this analysis. Arithmetic means within subjects, pair types, and time points were taken and used for this analysis. Trials that resulted in a negative squared dissimilarity (see Methods) were considered invalid and excluded from these arithmetic means. After means were taken, a Box-Cox transformation with λ = 1.706 was used to normalize the data before analysis.

Summary data were computed using the transformed dataset. An inverse transformation was performed on the summary data before reporting. The main effect of learning condition of dissimilarity was not significant, $F(1, 94) = .82, p = .367, \eta_p^2 = .009$. The main effect of pair type on dissimilarity was significant, $F(1, 282) = 24.387, p < .001, \eta_p^2 = .08$. Dissimilarity was higher for between-category pairs ($M = 2.51, SE = .002, 95\%$ CI: $[1.707, 3.151]$) than for within-category pairs ($M = 2.422, SE = .002, 95\%$ CI: $[1.617, 3.036]$). The main effect of time point on dissimilarity was significant, $F(1, 282) = 68.155, p < .001, \eta_p^2 = .195$. Dissimilarity was higher before learning ($M = 2.539, SE = .002, 95\%$ CI: $[1.71, 3.184]$) than after learning ($M = 2.391, SE = .002, 95\%$ CI: $[1.656, 3.015]$). The interaction between condition and time point was significant, $F(1, 282) = 7.872, p = .005, \eta_p^2 = .027$. Four post-hoc t-tests with Bonferroni correction were performed. Within the massed learning condition, dissimilarity was significantly higher before learning ($M = 2.537, SE = .004, 95\%$ CI: $[1.645, 3.225]$) than after learning ($M = 2.339, SE = .004, 95\%$ CI: $[1.571, 3.012]$), $p < .001$. Distributed learners' dissimilarity did not change significantly after learning, $p = .037$. Pre-learning, dissimilarity did not differ significantly between learning conditions, $p = .855$. Post-learning, dissimilarity was marginally higher for distributed learners ($M = 2.446, SE = .003, 95\%$ CI: $[1.807, 3.003]$) than for massed learners ($M = 2.339, SE = .004, 95\%$ CI: $[1.571, 3.012]$), $p = .019$. All other interactions were not significant, $p's > .05$. These results are visualized in Figure 10.

Figure 10

Between- and Within-Category Dissimilarity before and after Learning



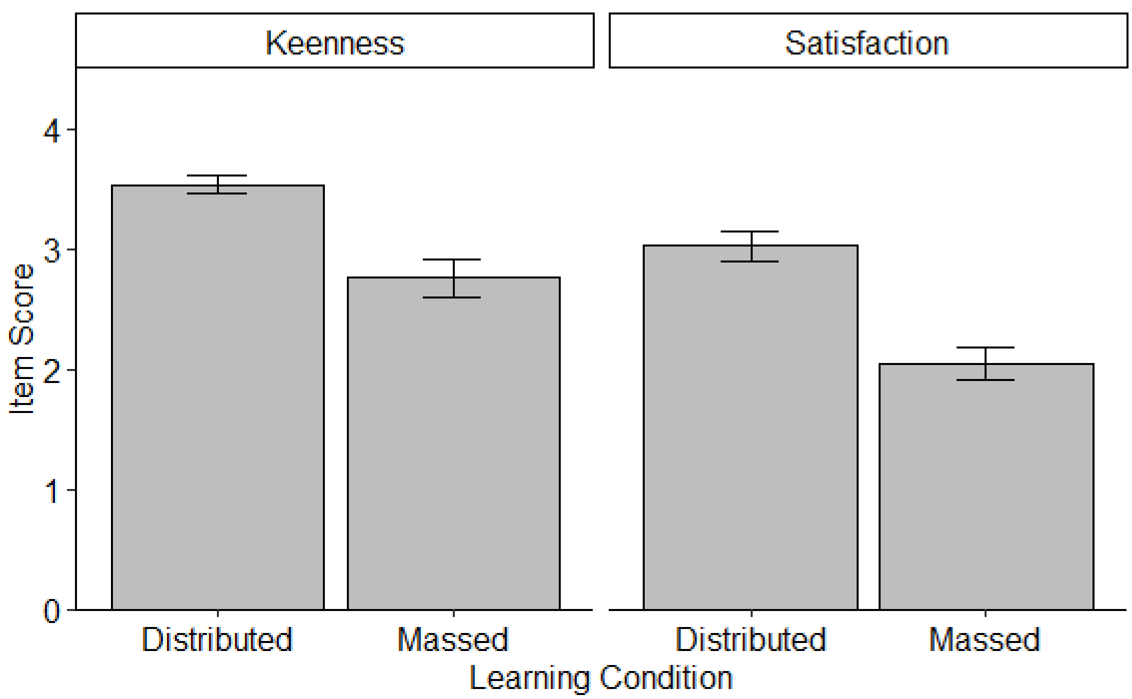
Similarity data were filtered such that only similarity judgments for adjacent pairs would remain in the next analysis. Change in dissimilarity from pre-learning to post-learning was calculated and averaged within each participant and level of learning condition, direction of variation, and location. One participant's data were excluded because they had one datum that varied by more than 3 standard deviations from the mean. A 2x2x2 (learning condition x direction of variation x location) mixed factorial ANOVA was conducted with change in dissimilarity as the dependent variable (i.b. Folstein et al., 2013). The main effect of learning condition was not significant, $F(1, 93) = .803, p = .372, \eta_p^2 = .009$. The main effect of direction of variation was not significant, $F(1, 279) = .155, p = .694, \eta_p^2 = .001$. The main effect of location was not significant, $F(1, 279) = .15, p = .699, \eta_p^2 = .001$. All interaction effects were not significant, p's $> .05$.

Metacognition

A Welch two sample t-test was used to compare satisfaction and keenness between learning conditions. Distributed participants showed a higher level of satisfaction ($M = 3, SD = .9$, 95% CI: $[1.2, 4]$) than did the massed participants ($M = 2, SD = .9$, 95% CI: $[0.2, 3.8]$), $t(93.904) = 5.462, p < .001, d = 1.113$. Distributed participants showed a higher level of keenness ($M = 3.5, SD = .6$, 95% CI: $[3, 4]$) than did the massed participants ($M = 2.8, SD = 1.1$, 95% CI: $[1, 4]$), $t(72.001) = 4.504, p < .001, d = .908$. See Figure 11.

Figure 11

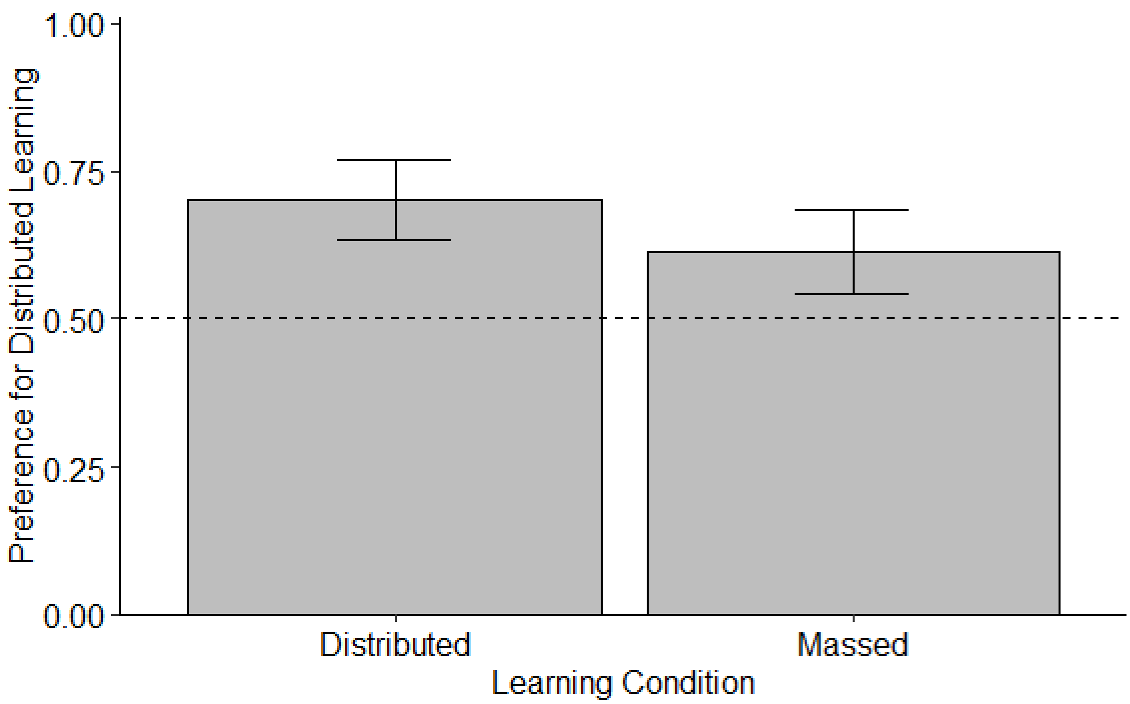
Keenness and Satisfaction between Learning Conditions



A Chi-Square goodness of fit test was used to determine if participants tended to prefer distributed or massed training in survey item 2. Participants tended to prefer distributed training, $X^2(1, 96) = 9.375, p = .002$. A Chi-Square test of independence was used to determine if participant preferences differed by learning condition. Participant preferences did not differ according to learning condition, $X^2(1, 96) = .507, p = .476$. See Figure 12.

Figure 12

Proportion of Participants Preferring Distributed Learning



Exploratory analyses were conducted to determine whether participants' accuracy could predict survey responses. There was no significant correlation between accuracy and satisfaction, $r(94) = .187$. There was no significant correlation between accuracy and keenness, $r(94) = .268$. A logistic regression was conducted to determine whether preference for distributed training could be predicted from a participant's overall accuracy. This regression was not significant, $p = .956$.

Discussion

The large effect of learning block on accuracy indicates that participants were able to learn the task regardless of condition. Mean accuracy reached well above chance levels within the first block of learning. There was a small main effect of learning condition on accuracy. This supports my hypothesis that the spacing effect would be present in this study, as distributed learners showed higher overall accuracy levels. As both types of learners showed similar performance levels during the first block of learning, this difference emerged over the course of training due to the experimental manipulation. There was a small main effect of block on reaction time such that reaction times tended to decrease over the course of learning. Learning condition had a medium effect on reaction times such that massed learners had lower reaction times than distributed learners, but both types of learners had similar reaction times during the first block of learning, suggesting that this effect was also due to the experimental manipulation. This echoes the results of previous spacing effect studies in which massed learners spend less time on repetitions (Carpenter, 2020).

Overall accuracy and mean reaction times had a medium positive correlation for the massed learners. This correlation was not significant among distributed learners and when controlling for this correlation, the main effect of learning condition on accuracy disappeared. This finding supports the attention attenuation hypothesis, which holds that the spacing effect is born from decreased attention over time in massed learners. Studies by Vlach et al. (2008) and Kornell et al. (2010) claim to have found results inconsistent with the attention attenuation hypothesis, as spacing effects were present even though participants in both massed and spaced learning conditions experienced stimuli for equal durations of time. However, participants in these studies did not have control over their stimulus presentation times and it should not be assumed that they paid full attention for the duration of stimulus presentations. Participants in the current study could end a repetition at their own volition, making this a stronger assessment of attention attenuation.

During the first and last blocks of learning, distributed learners were more likely to exceed the maximum accuracy attainable under a suboptimal categorization strategy. When data were subset to only include learners who used an information-integration strategy during the final block of learning, there was no main effect of learning condition on accuracy or on Akaike's Information Criterion (model fits). Together, these results suggest that temporal spacing played a role in pushing participants away from suboptimal learning strategies.

Multidimensional scaling did not reveal any substantial qualitative differences in participants' perceptions of the stimulus space in relation to learning condition or time point. In the model-based representational similarity analysis, model fits did not change substantially after learning and did not differ between learning conditions. Participants' similarity judgment data were well fit by both the categorical and the physical distance model, but the physical distance model had significantly stronger model fits. There was no evidence of between-category expansion, acquired distinctiveness, or acquired equivalence. Together, these results fail to support my hypotheses. This task did not produce categorical perception in participants.

There was a large decrease in dissimilarity for both within- and between-category pairs after learning. Although within-category compression is often a sign of categorical perception, here it seems more likely to indicate sensory habituation, a general loss of sensitivity to variations among stimuli after long series of repetitions. The small interaction between learning condition and time point revealed that dissimilarity decreased for massed learners but not for distributed learners. There was no main effect of learning condition on dissimilarity before learning. This suggests that massed learners were driving the main effect of time point on dissimilarity. Temporal spacing may have protected distributed learners from sensory habituation. These findings support deficient processing accounts of the spacing effect. Vlach et al. (2021) argued against this explanation for the spacing effect. They found that, despite performing better on a delayed induction test, spaced learners focused less visual attention on category-relevant stimulus features. However, this finding does not necessarily refute the deficient processing theory; massed learners may have spent more time focusing on task-relevant features to account for their deficient processing.

Participants in the distributed learning condition were more satisfied with their learning and were keener to undergo distributed training again; these effects were both large. Participants in both learning conditions also tended to indicate a preference for distributed learning if trained again in the future. These results reflect a metacognitive congruence; survey responses seemed favorable to the optimal learning condition. These findings fail to support my hypothesis that participants would respond more favorably toward massed learning. Exploratory analyses revealed that post-study survey results could not be predicted by participants' overall accuracy levels. Some mechanism other than performance must be underlying these evaluations.

Processing fluency and prior beliefs have been proposed as influences of metacognitive evaluations. Processing fluency is unlikely to play a role here. The massed group most likely had higher processing fluency, so metacognitive incongruence would have been found if processing fluency was driving these evaluations. The role of prior beliefs is unclear. If prior beliefs about spaced learning drove metacognitive incongruence in previous studies, then, if people's general sentiment toward spaced learning has not changed, metacognitive incongruence should have been found in the current study. In recent years, however, educational activities have largely shifted to online, asynchronous formats due to the COVID-19 pandemic. This shift may have changed people's attitudes towards distributed learning. However, convenience may serve as an alternate explanation for these metacognitive evaluations. Participants in the current study completed their learning sessions remotely from a smartphone; there was no inconvenience due to travel. Massed learners needed to dedicate substantial periods of time to the task, however, while distributed learners had nearly negligible training times. This difference may have made massed learners feel more inconvenienced than distributed learners. The opposite may be true of in-person studies, however; distributed training for in-person studies could result in repeated journeys to and from a physical location and a still-substantial amount of time being dedicated to the task. In Baddeley and Longman's (1978) study, for instance, the 1x1h learners travelled to the training center twice as often as 2x2h learners. Since total study time was kept constant between groups, 1x1h learners had to endure this inconvenience for a longer span of time. In studies with shorter spacing gaps, spacing serves to lengthen the duration of the study, again making spacing inconvenient. The current results suggest that prior beliefs, alone, are not likely to be driving the metacognitive incongruence often found in studies of learning schedule.

Practical Implications

Artificially learned categories play a vital role in daily life. Radiologists use a visual search procedure, similar to categorization, to classify their findings (Waite et al., 2019). Cardiologists and pulmonologists make diagnostic judgments using electrocardiograms and chest x-rays, respectively (Rourke et al., 2016). Compared with traditional rule-based learning, category learning (or perceptual training) is a more effective paradigm for training participants to detect melanomas (Xu et al., 2016). Optimizing artificial category learning procedures could improve the degree to which these artificial categories are learned. The ecologically valid spaced learning paradigm presented here could be adapted to help students learn categories more effectively than with traditional methods.

Limitations and Future Directions

Participants in this study were recruited from Prolific, a platform that allows the easy recruitment of a diverse, global sample for psychological research. Participants recruited from this platform may be more adept at performing psychological studies than the general population. In the current study, participants' median number of previous study approvals, an index of psychological study experience, was 126 (IQR: [82, 188]). Experience did not differ significantly between learning conditions and the spacing effect was observed in a (presumably) less experienced SONA pilot sample, making it unlikely that the group differences seen here are explained by differences in prior study experience. Still, it should be noted that participants in the current study tended to have substantial experience participating in other psychological studies and this may have impacted the results. Participants in this study were also diverse. Results may have differed had the study been conducted on a more culturally homogenous sample (e.g., Unsworth et al., 2005), such as that which might be expected from an in-person replication. A study directly comparing the effects of spacing among different cultural groups may be of interest.

In his paper finding evidence that acquired distinctiveness and equivalence could be evoked be artificial category learning, Goldstone (1994) preceded his category learning studies with perceptual discrimination studies. The results of these perceptual discrimination studies were used to choose stimulus parameters such that adjacent stimuli were equally discriminable across all points in the stimulus space. There stimulus space was a 4x4 grid, similar to the similarity judgment stimuli used in the current study. Such consideration was not given to the design of stimuli in the current study. Future iterations of this study should tune the stimulus space in a similar manner, as discriminability may not be equal at all points in the current study's stimulus space.

Substantial consideration should also be given to what one considers a single learning experience. Studies of the spacing effect tend to treat each stimulus presentation as its own learning experience (e.g., Vlach et al., 2008). In the present study, one block of learning (128 stimulus presentations) was treated as one learning experience, instead. This is sensible because the goal of category learning is to learn to partition the stimulus space, not to commit exemplars to memory. The approach used in this study is also more analogous to classroom learning or studying, in which students review all the information in each lesson and then take a break before returning to study again. It would be impractical to ask students to apply a large spacing gap between the individual items that comprise a single lesson. This type of spacing may make category induction more difficult (see Kornell et al., 2010). The discriminative contrast between adjacent stimuli facilitates induction (Kang & Pashler, 2012; Kornell & Bjork, 2008) and temporal spacing makes this contrast more difficult to achieve. A future replication of the current study could vary inter-stimulus interval between-participants to determine if inter-stimulus intervals interact with the between-block spacing gaps.

The present study provided trial-by-trial feedback but did not inform participants about their performance on entire blocks. Block-level feedback may have helped participants identify the optimal categorization strategy more easily. Distributed learners had ample time to independently reflect on their performance between training sessions, but massed learners did not. Block-level feedback for all learners may have reduced the spacing effect if it was, in part, driven by distributed learners' ability to reflect on block-level performance. This is only speculation, however, as self-reflection in this type of task has not yet been examined. Future research should address whether additional feedback or reflection on performance play a role in the spacing effect.

An information-integration category structure was used in this study as a person's ability to learn these structures is robust to taxes on working memory, sleep deprivation, and stress (Hughes & Thomas, 2021), all of which might reasonably be expected when asking participants to complete this task remotely on their smartphones. These structures also do not require selective attention (Ashby & Valentin, 2017). Since learning categories that do not require selective attention may be a developmental default for children (Sloutsky & Sophia Deng, 2019), the category learning task used here may well represent the way that humans naturally acquire many categories. Still, it would be valuable to replicate this study using different category structures. Information-integration structures are not robust to changes in context (e.g., Crossley et al., 2014). According to COVIS, humans learn rule-based category learning tasks using the declarative system, which relies on working memory and is robust to contextual changes. A rule-based task may therefore strengthen the spacing effect as it would minimize the negative impact of contextual variety on distributed learning and weaken massed learning due to longer periods of maintaining working memory. An in-person replication of this study would also be valuable for this reason; the fixed context may help further facilitate information-integration learning for distributed learners.

Several aspects of the present study were uncontrollable and unknowable due to the study's remote nature. There was most likely variety in participants' phone sizes, viewing distance, and viewing angle during learning. These factors could have impacted the detail with which participants experienced the stimulus presentations. Participants' internet connection strength most likely impacted image loading times, as well, and reaction time data do not account for this. There is value to the ecological validity afforded by this study's design and there is no reason to believe that these factors differed substantially enough between learning conditions to alter the results, but these factors still provide some uncontrollable sources of variability. These considerations should be kept in mind when interpreting the results.

In the present study, participants in the distributed condition were instructed to complete sessions 6-18 hours apart but were ultimately able to choose their spacing gaps. This enriched the data acquired from this task, as spacing gap could be treated as a continuous variable and adherence to the experimental manipulation could be assessed. This also adds ecological validity to the task, as humans typically guide their own learning in this way. Inter-block spacing gaps did not seem to be related to task performance in the distributed group and did not seem to vary significantly throughout learning. However, there was a very large range of variability in spacing gaps, with some participants completing learning sessions within an hour of each other. As demonstrated by Cepeda, Coburn, Rohrer, Wixted, Mozer, and Pashler (2009), all spacing gaps do not yield the same benefit. Procedural category learning may be optimized by a different spacing gap than the one assessed in this study. Future research may employ tighter controls on participants' spacing gaps and employ between-participants designs to determine which spacing gaps may be most optimal. A wide pool of literature explores expanding and compressing spacing gaps; spacing gaps which, with each successive learning experience, become larger or smaller. Expanding learning schedules may produce stronger learning (Carpenter et al., 2012). A replication of the present study with added expanding and compressing spacing gap conditions may be of interest for future work.

It should be noted that other explanations for the spacing effect have been proposed (Hintzman, 1974). According to the consolidation hypothesis, massed learning causes the consolidation of later learning experiences to interfere with the consolidation of previous learning experiences. Some experiments have supported this retroactive interference proposal, but it is unlikely to have played a role in this study, as each block of learning consisted of identical to-be-learned material. The rehearsal hypothesis asserts that participants voluntarily retrieve and reprocess learning experiences during spacing gaps. It is also unlikely that rehearsal played a role in this spacing effect, as the artificial stimuli used in the present study were uninteresting and highly confusable. It is difficult to imagine participants voluntarily reprocessing this task in enough detail to positively impact future learning. Still, these explanations for the spacing effect have been proposed and should be considered.

Deficient processing and attention attenuation are distinct but may reasonably be expected to interact. Habituation is intensified when stimulus repetitions are more frequent (Thompson, 2009). When participants spend less time studying an individual stimulus, the frequency with which they encounter new stimuli increases, enhancing sensory habituation. When participants experience habituation, they will likely feel diminishing returns on their time spent studying, causing them to spend less time. Thus, these two attentional mechanisms may work together in a positive feedback loop during long learning sessions. Future research should explore this possibility.

With decreased attention (indexed by lower reaction times) being one possible mechanism for the spacing effect observed here, perhaps prompting participants to take their time when they provide a fast response would minimize the spacing effect by encouraging, but not forcing, participants to take more time when evaluating stimulus presentations. If attention attenuation and deficient processing do interact in a positive feedback loop, then this manipulation may reduce sensory habituation, as well. Conversely, it may be worthwhile to explore the effect of further decreasing attention. If attention attenuation and deficient processing interact in a positive feedback loop, then asking participants to provide faster responses may amplify the sensory habituation seen in the current study. Future research should address the relationship between attention attenuation and deficient processing.

While contextual variety is assumed to be present between and within distributed learners, the current study did not directly assess participants' physical context at each block of learning. Variations in mental state, such as substance use or sleep, was also not addressed. Future iterations of this study may use some form of ecological momentary assessment to record participants' physical context during each learning block as well as such variables as sleep and substance use to provide richer information on how long periods of spacing can impact category learning.

The similarity judgment task required 136 trials to account for each possible pairwise comparison. Since one goal of the current study design was the minimize the time that distributed learners spent on individual task sessions, this may have been suboptimal. An alternative assessment of pairwise similarity could be obtained from a multi-arrangement procedure, a task in which participants move images in a two-dimensional space such that similar images are close and dissimilar images are distant. The results of multi-arrangement procedures are strongly correlated with those of traditional (dis)similarity judgment tasks, but they require less time to complete and allow participants to see the entire stimulus space at once (Kriegeskorte & Mur, 2012). It is unclear if this task would be sensitive to sensory habituation, as was the case with the similarity judgment task in the current study, but this method merits exploration. An alternative approach to this problem would be to find the just-noticeable difference (JND) for participants before and after learning. This could be done for each stimulus dimension individually or, in the case of an information-integration task such as this one, for variations parallel or perpendicular to the category boundary. This procedure could yield more fine-grained results than a discrimination or similarity judgment task while still being sensitive to sensory habituation. The exploration of other assessments of participants’ perceptual space should be a highlight for future work.

Whether convenience plays a role in these metacognitive evaluations could be tested by performing an in-lab replication of the present study. Distributed learners in such a replication would be more inconvenienced than massed learners, as they would need to physically travel to a fixed location multiple times per day for several days. If convenience drives metacognitive evaluations, metacognitive incongruence would be found in an in-person replication. Such a replication could also begin with a short survey on participants’ attitudes towards spaced learning, providing insight into participants’ prior beliefs.

Satisfaction, preference, and keenness were not related to task performance. The current study did not ask participants to explicitly evaluate their learning, so it is unclear whether they were able to accurately assess their performance. Past studies have found that massed learners often over-estimate their performance; such depth would have been valuable in the current study. In addition to the attitudinal measures included in the current study, future research might ask participants to evaluate their learning.

Conclusion

The present study found support for a spacing effect in adult information-integration category learning. The spacing effect seen here appears to have been driven by attentional mechanisms, with massed participants paying less attention and losing sensitivity to variations in the stimulus space over the course of learning. Spacing seems to have protected against inattention and sensory habituation. Spacing may have facilitated the discovery of information-integration learning strategies, as the spacing effect was not present in the subset of participants who had discovered these strategies. Counter to expectations, metacognitive congruence was found, with survey responses tending to favor the optimal learning condition. These findings open opportunities for research into how COVIS theory's predictions interact with the spacing effect as well as how participant experiences impact their metacognitive judgments.

References

Ambridge, B., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development*, *21*(2), 174–193. https://doi.org/10.1016/j.cogdev.2005.09.003

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295x.105.3.442

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(1), 33–53. https://doi.org/10.1037//0278-7393.14.1.33

Ashby, F. G., & Townsend, J. T. (1986). Varieties of Perceptual Independence. *Psychological Review*, *93*(2), 154–179. https://www.ncbi.nlm.nih.gov/pubmed/3714926

Ashby, F. G., & Valentin, V. V. (2017). Chapter 7 - Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (second, pp. 157–188). Elsevier. https://doi.org/10.1016/B978-0-08-101107-2.00007-5

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*(8), 627–635. https://doi.org/10.1080/00140137808931764

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: A means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1705). https://doi.org/10.1098/rstb.2015.0355

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics: A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, *17*(2), 444–472. https://doi.org/10.1198/106186008X318440

Carpenter, S. K. (2020). Distributed practice or spacing effect. In *Oxford Research Encyclopedia of Education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.859

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

Crossley, M. J., Ashby, F. G., & Maddox, W. T. (2014). Context-dependent savings in procedural category learning. *Brain and Cognition*, *92C*, 1–10. https://doi.org/10.1016/j.bandc.2014.09.008

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Doyle, M. E., & Hourihan, K. L. (2016). Metacognitive monitoring during category learning: How success affects future behaviour. *Memory*, *24*(9), 1197–1207. https://doi.org/10.1080/09658211.2015.1086805

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240. https://doi.org/10.1016/0010-0277(92)90002-y

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *The American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology. General*, *123*(2), 178–200. https://doi.org/10.1037//0096-3445.123.2.178

Guzman-Munoz, F. J. (2017). The advantage of mixing examples in inductive learning: a comparison of three hypotheses. *Educational Psychology Review*, *37*(4), 421–437. https://doi.org/10.1080/01443410.2015.1127331

Harnad, S. (2003). Categorical Perception. In *Encyclopedia of Cognitive Science*. Encyclopedia of Cognitive Science (01/12/03). Nature Publishing Group: Macmillan. https://eprints.soton.ac.uk/257719/

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. *Theories in Cognitive Psychology: The Loyola Symposium.*, *386*. https://psycnet.apa.org/fulltext/1975-00291-001.pdf

Hughes, G. I., & Thomas, A. K. (2021). Visual category learning: Navigating the intersection of rules and similarity. *Psychonomic Bulletin & Review*, *28*(3), 711–731. https://doi.org/10.3758/s13423-020-01838-0

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311. https://www.ncbi.nlm.nih.gov/pubmed/9151747

Kenney, A. (2009). *The spacing effect in inductive learning*. http://akenney.fastmail.fm.user.fm/works/9.61Paper.pdf

Kéri, S. (2003). The cognitive neuroscience of category learning. *Brain Research. Brain Research Reviews*, *43*(1), 85–109. https://doi.org/10.1016/s0165-0173(03)00204-2

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*(2), 498–503. https://doi.org/10.1037/a0017807

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception*, *30*(9), 1115–1125. https://doi.org/10.1068/p3155

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*, 245. https://doi.org/10.3389/fpsyg.2012.00245

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, *62*(2), 386–401. https://doi.org/10.3758/bf03205558

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*(3), 732–753. https://doi.org/10.1037//0278-7393.24.3.732

Matsuki, K. (2017). *grt: General Recognition Theory* (R package version 0.2.1). https://CRAN.R-project.org/package=grt

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606. https://doi.org/10.1016/s0022-5371(70)80107-4

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

Mitsea, E., & Drigas, A. (2019). A journey into the metacognitive learning strategies. *International Journal of Online and Biomedical Engineering*, *15*(14), 4–20. https://doi.org/10.3991/ijoe.v15i14.11379

Monteiro, S., Melvin, L., Manolakos, J., Patel, A., & Norman, G. (2017). Evaluating the effect of instruction and practice schedule on the acquisition of ECG interpretation skills. *Perspectives on Medical Education*, *6*(4), 237–245. https://doi.org/10.1007/s40037-017-0365-x

Morehead, K., Dunlosky, J., & Foster, N. L. (2017). Do people use category-learning judgments to regulate their learning of natural categories? *Memory & Cognition*, *45*(8), 1253–1269. https://doi.org/10.3758/s13421-017-0729-9

Nordt, M., Hoehl, S., & Weigelt, S. (2016). The use of repetition suppression paradigms in developmental cognitive neuroscience. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *80*, 61–75. https://doi.org/10.1016/j.cortex.2016.04.002

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39

Ozturk, O., Shayan, S., Liszkowski, U., & Majid, A. (2013). Language is not necessary for color categories. *Developmental Science*, *16*(1), 111–115. https://doi.org/10.1111/desc.12008

Perszyk, D. R., & Waxman, S. R. (2016). Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition*, *153*, 175–181. https://doi.org/10.1016/j.cognition.2016.05.004

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363. https://doi.org/10.1037/h0025953

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (4.1.1). R Foundation for Statistical Computing. https://www.R-project.org/

Repp, B. H. (1984). Categorical Perception: Issues, Methods, Findings. In N. J. Lass (Ed.), *Speech and Language* (Vol. 10, pp. 243–335). Elsevier. https://doi.org/10.1016/B978-0-12-608610-2.50012-1

Rosch, E. (1975). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. https://doi.org/10.1037/e666602011-017

Rourke, L., Cruikshank, L. C., Shapke, L., & Singhal, A. (2016). A neural marker of medical visual expertise: implications for training. *Advances in Health Sciences Education: Theory and Practice*, *21*(5), 953–966. https://doi.org/10.1007/s10459-016-9712-7

Simon, D. A. (2008). Scheduling and Learning. In M. Guadagnoli, A. S. Benjamin, J. S. de Belle, B. Etnyre, & T. A. Polk (Eds.), *Human Learning: Biology, Brain, and Neuroscience* (Vol. 139, pp. 61–72). Elsevier.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(4), 907–912. https://www.ncbi.nlm.nih.gov/pubmed/11486923

Simone, P. M., Bell, M. C., & Cepeda, N. J. (2013). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *The Journals of*

Gerontology. Series B, Psychological Sciences and Social Sciences, *68*(5), 674–680. https://doi.org/10.1093/geronb/gbs096

Sloutsky, V. M., & Sophia Deng, W. (2019). Categories, concepts, and conceptual development. *Language, Cognition and Neuroscience*, *34*(10), 1284–1297. https://doi.org/10.1080/23273798.2017.1391398

Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology*, *8*, 962. https://doi.org/10.3389/fpsyg.2017.00962

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Soto, F. A., & Wasserman, E. A. (2010). Integrality/separability of stimulus dimensions and multidimensional generalization in pigeons. *Journal of Experimental Psychology. Animal Behavior Processes*, *36*(2), 194–205. https://doi.org/10.1037/a0016560

Stephens, R., Perfors, A., & Navarro, D. (2010). Social context effects on the impact of category labels. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 1411–1416. https://digital.library.adelaide.edu.au/dspace/handle/2440/63957

Tauber, S. K., & Dunlosky, J. (2015). Monitoring of learning at the category level when learning a natural concept: will task experience improve its resolution? *Acta Psychologica*, *155*, 8–18. https://doi.org/10.1016/j.actpsy.2014.11.011

Thios, S. J., & D’Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Thompson, R. F. (2009). Habituation: a history. *Neurobiology of Learning and Memory*, *92*(2), 127–134. https://doi.org/10.1016/j.nlm.2008.07.011

Unsworth, S. J., Sears, C. R., & Pexman, P. M. (2005). Cultural influences on categorization processes. *Journal of Cross-Cultural Psychology*, *36*(6), 662–688. https://doi.org/10.1177/0022022105280509

Vlach, H. A. (2014). The spacing effect in children’s generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(1), 246–254. https://doi.org/10.1037/a0025260

Vlach, H. A., Bredemann, C. A., & Kraft, C. (2019). To mass or space? Young children do not possess adults’ incorrect biases about spaced learning. *Journal of Experimental Child Psychology*, *183*, 115–133. https://doi.org/10.1016/j.jecp.2019.02.003

Vlach, H. A., Kaul, M., Hosch, A., & Lazaroff, E. (2021). Attending less and forgetting more: Dynamics of simultaneous, massed, and spaced presentations in science concept learning. *Journal of Applied Research in Memory and Cognition; Washigton*. https://doi.org/10.1016/j.jarmac.2021.10.007

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750–763. https://doi.org/10.3758/s13421-010-0063-y

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory & Cognition*, *40*(5), 703–716. https://doi.org/10.3758/s13421-011-0180-2

Waite, S., Grigorian, A., Alexander, R. G., Macknik, S. L., Carrasco, M., Heeger, D. J., & Martinez-Conde, S. (2019). Analysis of perceptual expertise in radiology - Current knowledge and a new perspective. *Frontiers in Human Neuroscience*, *13*, 213. https://doi.org/10.3389/fnhum.2019.00213

Wang, J., & Xing, Q. (2019). Metacognitive illusion in category learning: Contributions of processing fluency and beliefs. *Advances in Cognitive*

Psychology / University of Finance and Management in Warsaw, *15*(2), 100. https://doi.org/10.5709/acp-0260-3

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Whitehead, P. S., Zamary, A., & Marsh, E. J. (2021). Transfer of category learning to impoverished contexts. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-02031-7

Xu, B., Rourke, L., Robinson, J. K., & Tanaka, J. W. (2016). Training melanoma detection in photographs Using the perceptual expertise training approach. *Applied Cognitive Psychology*, *30*(5), 750–756. https://doi.org/10.1002/acp.3250

Zhang, Q., Lei, L., & Gong, T. (2021). Categorical perception as a combination of nature and nurture. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/5r19w7z2

Appendix A

Initial Research Ethics Board Approval Letter

Western Research

Date: 2 February 2022

To: Dr. John Paul Minda

Project ID: 119955

Study Title: Effect of Learning Schedule on Remote Category Learning

Short Title: Remote Category Learning

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 04/Mar/2022

Date Approval Issued: 02/Feb/2022 16:10

REB Approval Expiry Date: 02/Feb/2023

Dear Dr. John Paul Minda

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. **All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study.**

Documents Approved:

Document Name	Document Type	Document Date	Document Version
stimulus_distribution	Other Data Collection Instruments	15/Dec/2021	
category_learning_instructions	Other Data Collection Instruments	15/Dec/2021	
similarity_judgment_instructions	Other Data Collection Instruments	15/Dec/2021	
unique_id_screen	Other Data Collection Instruments	15/Dec/2021	
SONA Letter - CLEAN	Recruitment Materials	25/Jan/2022	
Qualtrics Consent_and_Email	Online Survey	26/Jan/2022	
Qualtrics Demographics	Online Survey	26/Jan/2022	
LOI_C - CLEAN	Implied Consent/Assent	26/Jan/2022	

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Pre-Consent Screening	Screening Form/Questionnaire	15/Dec/2021	

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Appendix B

Amendment 1 Research Ethics Board Approval Letter

Western Research

Date: 15 March 2022

To: Dr. John Paul Minda

Project ID: 119955

Study Title: Effect of Learning Schedule on Remote Category Learning

Application Type: NMREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: 01/Apr/2022

Date Approval Issued: 15/Mar/2022 16:21

REB Approval Expiry Date: 02/Feb/2023

Dear Dr. John Paul Minda,

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Qualtrics Demographics (Revised) - CLEAN	Online Survey	23/Feb/2022	
Qualtrics Consent_and_Email (Revised) - CLEAN	Online Survey	23/Feb/2022	
LOI_C - CLEAN (9 Mar 2022)	Implied Consent/Assent	09/Mar/2022	

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Massed Pavlovia - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 1 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 2 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 3 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 4 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 5 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 6 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 7 - CLEAN	Other Materials	23/Feb/2022	
Distributed Session 8 - CLEAN	Other Materials	23/Feb/2022	

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Ms. Katelyn Harris, Ms. Zoë Levi, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix C

Amendment 2 Research Ethics Board Approval Letter

Western Research

Date: 12 April 2022

To: Dr. John Paul Minda

Project ID: 119955

Study Title: Effect of Learning Schedule on Remote Category Learning

Application Type: NMREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: 06/May/2022

Date Approval Issued: 12/Apr/2022 16:20

REB Approval Expiry Date: 02/Feb/2023

Dear Dr. John Paul Minda,

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Code - Prolific Massed Session	Online Survey	05/Apr/2022	
Code - Prolific Session 1	Online Survey	05/Apr/2022	
Code - Prolific Session 2	Online Survey	05/Apr/2022	
Code - Prolific Session 3	Online Survey	05/Apr/2022	
Code - Prolific Session 4	Online Survey	05/Apr/2022	
Code - Prolific Session 5	Online Survey	05/Apr/2022	
Code - Prolific Session 6	Online Survey	05/Apr/2022	
Code - Prolific Session 7	Online Survey	05/Apr/2022	
Code - Prolific Session 8	Online Survey	05/Apr/2022	
LOI_C (5_4_2022)	Implied Consent/Assent	05/Apr/2022	
Qualtrics - LOIC	Online Survey	05/Apr/2022	
Similarity Judgment Instructions	Online Survey	08/Apr/2022	
Similarity Judgment Task	Online Survey	08/Apr/2022	
Category Learning Instructions	Online Survey	08/Apr/2022	
Category Learning Task	Online Survey	08/Apr/2022	
Recruitment Information - Prolific - April 8 Revision CLEAN	Recruitment Materials	08/Apr/2022	2
3-Question Survey	Online Survey	08/Apr/2022	

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Recruitment Information - Prolific - April 8 Revision TRACKED	Other Materials	08/Apr/2022	

REB members involved in the research project do not participate in the review, discussion or decision.