The effect of encoding conditions on learning in the prototype distortion task

Jessica C. Lee¹ · Evan J. Livesey¹

Published online: 17 November 2016
© Psychonomic Society, Inc. 2016

Abstract The prototype distortion task demonstrates that it is possible to learn about a category of physically similar stimuli through mere observation. However, there have been few attempts to test whether different encoding conditions affect learning in this task. This study compared prototypicality gradients produced under incidental learning conditions in which participants performed a visual search task, with those produced under intentional learning conditions in which participants were required to memorize the stimuli. Experiment 1 showed that similar prototypicality gradients could be obtained for category endorsement and familiarity ratings, but also found (weaker) prototypicality gradients in the absence of exposure. In Experiments 2 and 3, memorization was found to strengthen prototypicality gradients in familiarity ratings in comparison to visual search, but there were no group differences in participants' ability to discriminate between novel and presented exemplars. Although the Search groups in Experiments 2 and 3 produced prototypicality gradients, they were no different in magnitude to those produced in the absence of stimulus exposure in Experiment 1, suggesting that incidental learning during visual search was not conducive to producing prototypicality gradients. This study suggests that learning in the prototype distortion task is not implicit in the sense of resulting automatically from exposure, is affected by the nature of encoding, and should be considered in light of potential learning-at-test effects.

Electronic supplementary material The online version of this article (doi:10.3758/s13420-016-0252-z) contains supplementary material, which is available to authorized users.

✉ Jessica C. Lee
jessica.c.lee@sydney.edu.au

¹ School of Psychology, University of Sydney, Sydney, NSW 2006, Australia

Keywords Recognition · Prototype distortion · Visual search · Category learning · Implicit

A key debate within the category learning literature pertains to whether some kinds of category structures can be learned implicitly and whether they recruit independent neural structures to explicit learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005, 2011; Poldrack & Foerde, 2008; Smith & Grossman, 2008). One paradigm that is central to the debate is the A/Not-A prototype distortion task. In this task, participants are exposed to a set of exemplars that are generated by distorting a prototype, creating a category of stimuli that are physically similar to each other and whose average is the prototype. Learning about the category (A) is evident if participants can subsequently distinguish between novel exemplars and foils (make accurate A/Not-A judgements), or if the prototype is given the highest category endorsement on test (e.g., Posner & Keele, 1968). Alternatively, learning may be evident if, when judging novel exemplars of varying distortion levels, category endorsements, or recognition judgements vary as a function of the similarity of the exemplars to the category prototype. This pattern of generalization, referred to as a prototypicality gradient, is present when the highest category endorsement or recognition is given for low distortions of the prototype and lowest endorsement or recognition is given for high distortions of the prototype. These effects (collectively referred to as prototype effects) can all be taken as evidence that participants have learnt something about the similarity structure of the category. Learning in the prototype distortion task has been labeled as implicit because of seemingly intact categorization performance in amnesic patients, coupled with the incidental nature of the exposure conditions, each of which will now be reviewed.

Intact category learning in memory-impaired patients

Prototype effects are often regarded as implicit because of several striking studies that have reported intact categorization in memory-impaired patients (Bozoki, Grossman, & Smith, 2006; Kéri, Kálmán, Kelemen, Benedek, & Janka, 2001; Kéri et al., 1999; Knowlton & Squire, 1993; Reed, Squire, Patalano, Smith, & Jonides, 1999). The earliest study demonstrating this was by Knowlton and Squire (1993), who measured categorization and recognition performance in amnesic patients and healthy controls after being exposed to a series of dot patterns centered around a prototype (for a description of these stimuli, see Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968). Amnesic and control participants were asked to point to the dot closest to the middle of each stimulus and were then told to use any knowledge they had acquired in a subsequent categorization and recognition test. While control participants were able to discriminate between category exemplars that they had previously seen, and category exemplars that were novel on test, amnesic patients were impaired at discriminating between new and old exemplars and yet still displayed categorization performance equivalent to the normal controls (see also Knowlton, Mangels, & Squire, 1996, for a similar finding in probabilistic category learning; Knowlton, Ramus, & Squire, 1992, for a similar finding in artificial grammar learning; and Squire & Knowlton, 1995, for a related case study). A similar result was found by Kéri et al. (2001; see also Kéri et al., 1999) using the same dot pattern stimuli, with impaired explicit recognition but equivalent A/Not-A categorization performance in patients with mild to moderate Alzheimer's disease relative to a control group. These studies (see also Bozoki et al., 2006; Reed et al., 1999, for similar results with different stimuli) seem to provide evidence for a dissociation between categorization and recognition memory, suggesting the existence of a separate system that can learn about categories in the absence of declarative knowledge.

However, there are some issues with these studies that limit their interpretation. The original demonstration by Knowlton and Squire (1993), while compelling, has been criticized on the basis of using stimuli that generate a “false” prototype effect, where a prototypicality gradient can result even in the absence of exposure (Nosofsky, Denton, Zaki, Murphy-Knudsen, & Unverzagt, 2012; Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004). These studies suggest that participants are relying on working memory and learning about the category on test in an explicit fashion (Palmeri & Flanery, 1999). This learning-at-test effect may be especially pronounced if the test phase contains a large proportion of low-distortion exemplars and foils that make the contrast between categorical and

noncategorical stimuli obvious (Nosofsky et al., 2012), or if the test phase contains a large number of low-distortion exemplars and prototype presentations which could highlight the prototypical features of the category (Zaki & Nosofsky, 2004).

Another issue concerns the small sample sizes that these studies (e.g., Knowlton & Squire, 1993; Reed et al., 1999) typically use. Since the dissociation requires a null effect (that performance on categorization is equivalent between amnesic and control groups), a small sample size means that there will be less power to detect a possible group difference if one exists. However, a large difference in recognition accuracy that naturally results from comparing memory-impaired participants with healthy participants might still be detected. Zaki (2004) confirmed this hypothesis in a meta-analysis on 12 studies that compared memory-impaired participants against controls on performance in various category-learning tasks, a subset of which were prototype distortion tasks. Contrary to the suggestions of the studies mentioned above, it was concluded that there was indeed an overall impairment in categorization performance in amnesic patients, consistent with the idea that prototype effects are dependent on declarative memory and underpinned by a single memory system.

Assuming that the intact prototype effects in amnesic patients are in fact due to exposure, Nosofsky and colleagues (Nosofsky & Zaki, 1998; Zaki, Nosofsky, Jessup, & Unverzagt, 2003) proposed that the dissociations discussed above can still be explained using a single system of memory. Dissociation can be predicted even if a common ability (e.g., sensitivity in discriminating between distinct exemplars in memory) underlies both categorization and recognition tasks if we assume that this ability is deficient in amnesic patients, and that this affects recognition judgements more than categorization judgements. In other words, there is a parameter difference between tasks responsible for the observed dissociations whereby changes in memory sensitivity between groups results in a large impairment in recognition and a small impairment in categorization (see Berry, Henson, & Shanks, 2006, for a related single-system view). This idea has been supported by the finding that after a delay of 1 week between exposure and test, participants' ability to discriminate between new and old exemplars was impaired, while categorization accuracy was unaffected (Nosofsky & Zaki, 1998). This dissociation in healthy participants supports the idea that even when assuming a single memory system, changes on one measure do not necessarily entail changes on the other. In summary, while the studies with amnesic populations are striking, the methodological issues concerning the stimuli and the ability of single-system theories to explain dissociations undermine the conclusion that learning in the prototype distortion task is implicit.

Emergence under incidental learning conditions

A less-cited reason that prototype effects can be considered implicit is that experiments demonstrating this effect have tended to expose participants to category exemplars under conditions where they are not encouraged to deliberately encode the stimuli. Most studies require participants to view the stimuli passively, such as thinking about their appearance (Bozoki et al., 2006), or perform a task that incidentally exposes them to the stimuli, such as pointing to the dot closest to the center of each stimulus (Knowlton & Squire, 1993). It is often assumed that because these conditions do not explicitly mention the existence of a category, any learning that occurs must be incidental. However, these passive viewing situations do not preclude the possibility that participants are intentionally encoding the stimuli in some way during this initial phase, assuming (correctly) that they will later be useful. Classifying these exposure conditions as incidental would be more convincing if it could be ensured that participants' explicit cognitive functions were more fully engaged by performing a more difficult task during the exposure phase. A prototypicality gradient under these conditions would suggest that participants can learn about categories of stimuli that are physically similar to one another in an automatic or incidental fashion, satisfying one of the defining characteristics of implicit learning (Cleeremans & Jiménez, 1998; Frensch, 1998; Stadler & Frensch, 1994). One way in which to achieve incidental exposure is suggested by another implicit learning paradigm, namely contextual cueing in visual search.

In a contextual cueing task (Chun & Jiang, 1998), participants search through a configuration of distractors (usually rotated letter *L*s) for a target (usually a rotated letter *T*). Unbeknown to participants, certain configurations appear multiple times throughout the experiment. Participants show reliable reductions in reaction time in response to repeatedly presented (old) configurations when compared to novel configurations during training, yet sometimes fail to explicitly recognize old configurations in a subsequent test, or accurately generate the correct location of the target when presented with the old distractor contexts (Chun & Jiang, 1998, 2003). While explicit knowledge is sometimes present (e.g., Smyth & Shanks, 2008), and several studies can be criticized on the basis of lacking sensitivity in the recognition test in comparison to the large number of trials involved in training (Vadillo, Konstantinidis, & Shanks, 2016), several researchers have concluded that learning via visual search in this paradigm is implicit (e.g., Chun & Jiang, 2003). Contextual cueing effects may reflect cueing of the target location by the surrounding visual context (Chun & Jiang, 1998), or facilitation in detection or responding due to learning associations between distractors for repeated stimuli (Beesley, Vadillo, Pearson, & Shanks, 2015; Kunar, Flusberg, Horowitz, & Wolfe, 2007).

While there is ongoing debate about the mechanisms responsible for contextual cueing effects as well as the role of awareness, it is clear that the learning that occurs is incidental to the task being performed. Thus, employing a visual search task that requires a correct response should sufficiently engage participants such that any learning that does occur can be confidently deemed to be incidental, in contrast to a task such as pointing to the middle dot, where there is usually no way to assess whether participants are performing the task properly.

Despite the theoretical significance of incidental learning conditions in establishing prototype effects as implicit, there have been few attempts to compare different methods of exposure in the prototype distortion task. One notable exception is a study by Gureckis, James, and Nosofsky (2011; see also Reber, Gitelman, Parrish, & Mesulam, 2003), who compared implicit and explicit learning conditions as well as an intentional encoding strategy against an incidental encoding strategy in a 2×2 between-subjects design. Participants were either aware (explicit group) or unaware (implicit group) of the existence of a category, and were either asked to study the stimulus as a configural whole (configural group) or imagine pointing to the center dot (dot group). At test, they found differential activation in response to exemplars and foils that varied as a function of encoding strategy, while telling participants about the existence of a category (implicit/explicit) had no consistent effect. They concluded that differential brain activation found in studies comparing implicit and explicit learning orientations (e.g., Reber et al., 2003) was better explained through different encoding strategies, rather than the implicit status of the learner or awareness of impending tests.

While this is an appropriate conclusion from Gureckis et al.'s (2011) results, it should be noted that all four groups produced an equivalent level of categorization accuracy (there were no significant main effects or interactions), suggesting that while the encoding manipulation was successful in encouraging participants to adopt different strategies, there was no evidence that this made a difference to actual test performance. Thus, while their study suggests that participants were obeying instructions to perform the pointing task when asked to, the issues of whether intentional encoding conditions have an advantage over incidental encoding conditions, and whether other incidental encoding conditions are conducive to prototype effects, remain unresolved.

The current study

The aims of this were twofold. The first was to assess the implicit status of learning in the A/Not-A version of the prototype distortion task by testing whether a prototypicality

gradient could result when participants performed a visual search task with the category exemplars. Appropriating a visual search paradigm similar to that used in contextual cueing allowed us to test this often-assumed property of implicit prototype effects in a novel way. The second aim was to directly compare two methods of exposure: intentional memorization of a set of prototype-centered stimuli for a subsequent memory test (Memorize group), and a visual search task using the same stimuli (Search group). This comparison is important since different studies use different methods of passive exposure, with the assumption that because participants are not informed about the existence of a category, then any learning that occurs must be incidental and in some sense equivalent. Comparing a visual search group to a group who are given direct instructions to memorize the stimuli allows the best chance of detecting potential differences between incidental and intentional encoding conditions, if they exist.

A novel aspect of our methodology was that instead of measuring categorization and recognition in separate tests, we chose to use a single continuous measure (familiarity ratings) to assess both. Learning about the category would thus be evident if participants show a generalization gradient (i.e., a prototypicality gradient), and recognition would be inferred if participants are able to discriminate between new and old test items at matched levels of distortion in their familiarity ratings. While this is a departure from the majority of studies on the prototype distortion task, it was motivated by the desire to test whether dissociations between generalization and recognition performance were still possible when participants were only making one kind of judgement.

For the following experiments, two novel sets of stimuli were constructed to mimic the statistical properties of the dot patterns with the aim of minimizing potential learning-at-test effects. Experiment 1 produced similar generalization gradients between participants who received a categorization test and participants who received a familiarity test, justifying the use of familiarity ratings only for the subsequent experiments. Experiment 1 also compared generalization gradients between groups who were not exposed to the stimuli (but led to believe that they were) against groups who did observe the stimuli. The magnitude of any prototypicality gradients exhibited after no exposure would indicate the degree of learning-at-test effects, allowing us to determine whether participants in subsequent experiments learned anything during visual search (see Smith, 2008, for a discussion of this subtraction logic). Experiments 2 and 3 compared prototypicality gradients for new and old test stimuli in a subsequent familiarity test. Experiment 3 doubled the length of exposure from Experiment 2 and added an additional visual search group, where the stimulus exposure terminated after the response was made to ensure that no explicit encoding could occur during the residual stimulus exposure duration.

Experiment 1

The primary aim of Experiment 1 was to test for any potential learning-at-test effects with our novel stimulus sets to use as a basis for comparison for the subsequent experiments. Previous studies have demonstrated that the dot pattern stimuli generate “false” prototype effects in normal participants who do not see the stimuli prior to the test phase (Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004). Participants in these studies undergo a mock-subliminal procedure, where they are led to believe that stimuli are being presented to them subliminally, but in fact never see any stimuli. Despite the stimuli being novel on test, they subsequently show a prototypicality gradient in their categorization judgements when presented with high- and low-distortion exemplars. One way that participants can show false prototype effects is by learning about the category during the test phase, since this usually involves informing participants about the existence of a category, and then exposing them to more category exemplars as well as the prototype. The degree of learning-at-test effects is especially pronounced if the test phase were to present a large number of low-distortion exemplars and foils on test, highlighting the contrast between categorical and noncategorical stimuli (Nosofsky et al., 2012), or when the prototype and low-distortion exemplars are presented multiple times during test, which would result in high repetition of prototypical features (Zaki & Nosofsky, 2004). For this reason, to minimize potential learning-at-test effects, we tested new and old exemplars at matched levels of distortion with no foils, and used the mock-subliminal procedure to test whether participants produced prototypicality gradients in the absence of exposure.

A potential explanation for the observed dissociations in amnesic patients discussed previously is that categorization and recognition tests have different task-specific variance. Usually, categorization accuracy is assessed by asking participants to make an A/Not-A category endorsement, and recognition is assessed by asking participants to make an old/new judgement about whether they have seen a particular stimulus before. Participants may have different response thresholds for categorization and recognition judgments, making an equivalent comparison between the two difficult. For example, participants may be more willing to classify a test stimulus as part of the seen category than to say that they recognize the stimulus based on the same level of uncertainty, or adopt a more stringent criterion for endorsing recognition since an exact match is required (Nosofsky, Little, & James, 2012). An equivalent comparison between recognition and categorization tasks is made more difficult when we consider that previous studies have typically measured the prototype effect using categorization judgements, and separately measured recognition in a completely different phase using different stimuli (e.g., Bozoki et al., 2006; Knowlton & Squire, 1993; Reber & Squire, 1999). For this reason, we sought to test whether a

dissociation could still occur when we equated not only the stimuli between groups but also the test measure. Thus, another aim of Experiment 1 was to assess category learning (as indexed by prototypicality gradients) and recognition (as indexed by discrimination between new and old exemplars) using both categorization and familiarity tests and to show that they were equivalent between test measures.

Method

Participants

Ninety-two ($M_{\text{age}} = 19.97$ years, $SD = 3.38$, 70 female) University of Sydney students participated in this experiment in exchange for partial course credit.

Apparatus

All experiments were programmed using PsychToolbox for MATLAB (Brainard, 1997; Pelli, 1997) and run on Apple Mac Mini desktop computers connected to 17-inch CRT monitors, refreshed at a rate of 85 Hz. A standard Apple keyboard and mouse were used, and testing was conducted in individual cubicles in groups of up to five. The apparatus was the same for all subsequent experiments.

Stimuli

The stimuli were constructed to be complex and have multiple dimensions so that focus on a particular feature or an attempt to derive a verbalizable rule to describe the category would be difficult and less useful than memorizing the whole pattern. There were two sets of stimuli constructed for these experiments, each containing 10 features (circles or lines). For the circle stimuli, there were 10 colored circles on a 600×600 pixel black square. Each circle had the following variable properties: hue (saturation and brightness were held constant at their maximum respective values), location (x and y coordinates) within the black square, line thickness, and size (see Fig. 1). For the line stimuli, there were 10 white lines (five oriented horizontally and five vertically) on a black square of the same size (see Fig. 1). Each line was defined by its starting and ending location on the border of the square, as well as line thickness. Maximum and minimum values were chosen for each property to constrain the possible exemplars that could be created and to avoid placing the circles and lines outside the stimulus boundary (see Table 1).

All variable feature properties listed above were varied from exemplar to exemplar except for line thickness, which was held constant at 3 pixels in both stimulus sets. Arbitrary distortion levels were chosen ranging from 0.1 to 1.0 in increments of 0.1, with the smaller levels of distortion indicating typical exemplars that were similar to the prototype and higher

levels indicating atypical category exemplars. While the distortion level numbers are arbitrary, they represent a ratio scale with, for example, a distortion level of 0.1 being 10 % of the distortion level of 1.0. Arbitrary dimension multipliers were chosen for each feature dimension which would create a dissimilar stimulus when multiplied by the higher distortion levels but a highly similar stimulus when multiplied by the lowest distortion level (see Fig. 1 for examples).

A different circle and line prototype stimulus was randomly generated according to a seed number, which was different for each participant number. Thus, participants with the same participant number were exposed to exactly the same stimuli throughout the experiment. Each category exemplar was created by distorting the relevant prototype stimulus on a feature-by-feature basis. This meant that the dimension multiplier (see Table 1) was multiplied by the distortion level (e.g., 0.1) for each dimension of the 10 features (circles or lines) separately. The direction of change (positive or negative) was determined randomly and independently for each dimension of each feature. For example, to create a circle exemplar at 0.1 distortion, the size multiplier was multiplied by 0.1, and then either added or subtracted (randomly chosen for each feature) to the prototype values of the first circle, with this process repeating for the remaining circles, and this process repeating again for each of the circle dimensions (location, hue, etc.). If any feature values extended beyond the minimum or maximum values, they were simply changed to either the minimum or maximum value they crossed (see Table 1).

Twenty circle exemplars and 20 line exemplars were shown during the initial exposure phase (two unique exemplars at each of 10 levels of distortion for each set). A further 20 novel circle exemplars and 20 novel line exemplars were also shown during the test phase (again, with two exemplars at each level of distortion). To ensure that participants could not rely on occasional salient features (e.g., clustering of circles in a fashion that leads to occlusion by overlapping circles) within the stimulus to aid memory, additional checks were implemented to ensure that there would be minimal overlap between the circles (at least 50 pixels between the center of all circles). If this check failed, then all circle locations were randomized until this condition had been met. Note that overlap and occlusion was a frequent and regular feature of the line patterns, and thus this check was not implemented for these stimuli because it would be difficult to use a specific conjunction to aid memory for a specific stimulus. In addition, to ensure that the average of the exposed exemplars was indeed the prototype, the average value of the 20 exemplars was computed for each dimension for each of the 10 features separately (e.g., x coordinate, size, starting position of lines, etc.). If the average was not within a predetermined maximum distance from the prototype, all stimulus values were randomized again until they were sufficiently close (see Table 1 for the maximum distance allowed for each feature dimension).

Fig. 1 Examples of exemplars from each stimulus set (*upper panels*: circle stimuli; *lower panels*: line stimuli) used throughout all experiments. From left to right: prototype (zero distortion), a low-

distortion exemplar (0.1 distortion), a medium distortion exemplar (0.5 distortion), and a high-distortion exemplar (1.0 distortion). (Color figure online)

Procedure

The experiment was a 2×2 between-subjects design. Participants were randomly allocated to either a mock “subliminal” (NoEx) or passive exposure (Ex) phase, and either a familiarity (Fam) or categorization (Cat) test phase. Participants allocated to the exposure phase saw 20 unique exemplars from each of the circle and line categories 4 times each, with presentation randomized within each block of 40 trials (four blocks, 160 trials in total). Participants allocated to the subliminal exposure phase were told that they would be presented with some subliminal stimuli, which would be quickly masked by a black square. They were told that because the presentation was so brief, all they might see is the screen flash before the mask covered the stimulus. In the (sham) exposure

phase that followed, no stimuli were presented, but on each trial, the screen would flash from white to black and after 10 ms return to white, along with the black stimulus background (i.e., the mask) presented for the same time (2 seconds) as in the actual exposure phase. In the two exposure groups, the stimuli appeared on screen for 2 seconds before disappearing. The interstimulus interval (ISI) was a blank white screen and was 2 seconds in both exposure conditions.

After the exposure phase, participants either completed a familiarity or categorization test. Participants allocated to the familiarity test were told that they would be presented with more stimuli, where some of the stimuli would be familiar (i.e., they had seen them before in the first phase) and others novel (they had not seen them in the first phase). They were then asked to rate their level of familiarity toward 84 new

Table 1 Minimum and maximum values and multiplier for each stimulus feature dimension

Dimension	Minimum value	Maximum value	Multiplier	Distortion to prototype
Circle location (x)	60 pixels	540 pixels	40	20
Circle location (y)	60 pixels	540 pixels	40	20
Circle hue	None	None	.05	.1
Circle radius	5 pixels	50 pixels	10	5
Circle line width	3 pixels	3 pixels	0	0
Line start (x)	50	550	50	20
Line start (y)	50	550	50	20
Line end (x)	50	550	50	20
Line end (y)	50	550	50	20
Line width	3 pixels	3 pixels	0	0

Note. Location values could range from zero to 600, and the saturation and brightness were always set at 1 and 255, respectively. For vertical lines, only the x values were varied (y values were held constant at zero and 600) and for horizontal lines, only the y values were varied (x values were held constant at zero and 600). Hue values wrapped around 0–1 since the scale was continuous. For each dimension of each feature, if the average of the exposed exemplars was not within the specified maximum distance to the prototype, all exemplar values were created again until this condition was met

exemplars (for each category: the prototype presented twice, 20 old and 20 new exemplars) using a visual analogue scale that ranged from *definitely not familiar* to *definitely familiar*. If participants were allocated to the categorization test, they were told that all of the stimuli they had seen formed part of a category and that they would be presented with more stimuli that may or may not be part of the same category. For each stimulus, participants were asked, “Is this stimulus part of the circle category you saw before?” or “Is this stimulus part of the line category you saw before?” It was made clear that participants should only consider the circle category when presented with a circle stimulus and similarly for the line category. Participants made their rating on a visual analogue scale that ranged from *definitely no* to *definitely yes*.

In both tests, each stimulus was presented for 2 seconds, and then disappeared. After 500 ms, a rating scale would appear along with the appropriate test question. Participants made their rating by clicking a point on the scale with the mouse and had unlimited time to make and change their rating. It was made clear that all questions referred to stimuli they had seen in the first part of the experiment. The midpoint of the scale and both endpoints were marked with ticks. All ratings were transformed to range from zero to 100. The familiarity and categorization tests ran the same way no matter what exposure phase participants were allocated to. For this, and all subsequent experiments, the same set of seed numbers for the random number generator were used in each group, such that the stimuli seen and tested were randomized but matched between groups of participants. This also meant that old and new test stimuli were dummy coded for the no-exposure groups. There was no feedback during the test phase.

Results and discussion

For all subsequent analyses, Greenhouse–Geisser corrections to p values are reported for violations of sphericity, and the results reported combine both circle and line stimulus sets.¹ Figure 2 shows the mean ratings for the test stimuli in each of the four groups. To analyze whether the prototype effect, and recognition, were dependent on exposure and varied with test question, a 2 (exposure group: exposure vs. no exposure) $\times$ 2 (test group: categorization vs. familiarity) $\times$ 2 (novelty: new vs. old) $\times$ 10 (distortion level) ANOVA was performed on the exemplar ratings (excluding the prototype) in the test phase. In all subsequent analyses we report the linear trend contrasts (and not the main effect) in distortion level and associated interactions, since only the linear trend addresses our hypotheses about prototypicality gradients.

¹ Analyses on the familiarity ratings were also performed on the circle and line stimulus sets separately, and produced similar results for our critical findings. We have noted where the results differed between stimulus sets.

The main effect of distortion level was significant in all cases where the linear trend was significant.

There was a main effect of novelty, $F(1, 88) = 18.85$, $p < .001$, $\eta_p^2 = .176$, which did not interact with test group, $F < 1$, indicating that overall, participants rated old test stimuli higher on category endorsement and familiarity than new test stimuli. Unsurprisingly, novelty interacted with exposure group, $F(1, 88) = 17.97$, $p < .001$, $\eta_p^2 = .170$, such that participants who had been exposed to the exemplars could discriminate between new and old test items (see Fig. 2a, c), but participants who had not seen the stimuli could not (Fig. 2b, d). This was confirmed in a set of separate analyses where there were significantly higher familiarity ratings, $F(1, 22) = 10.18$, $p = .004$, $\eta_p^2 = .316$, and category endorsements, $F(1, 22) = 15.24$, $p = .001$, $\eta_p^2 = .409$, for old exemplars in the two exposure groups, and no differences between new and old exemplars in the two no-exposure groups, largest $F < 1$.

There was a significant linear trend for distortion level, $F(1, 88) = 121.9$, $p < .001$, $\eta_p^2 = .581$, since overall ratings declined as the level of distortion increased, indicating the presence of a prototypicality gradient. The linear trend for distortion interacted with exposure group, $F(1, 88) = 33.48$, $p < .001$, $\eta_p^2 = .276$, but not test group, $F(1, 88) = 1.85$, $p = .177$, $\eta_p^2 = .021$, nor novelty, $F(1, 88) = 3.58$, $p = .062$, $\eta_p^2 = .039$. Thus, it appears that the prototypicality gradient was stronger (i.e., there was a steeper gradient) when participants were exposed to the stimuli, but did not vary according to the type of test conducted, nor between new and old exemplars.

+None of the three-way interactions or the four-way interaction were significant, largest $F(1, 88) = 1.31$, $p = .256$, $\eta_p^2 = .015$. There was a significant main effect of exposure group, $F(1, 88) = 20.37$, $p < .001$, $\eta_p^2 = .188$, and test group, $F(1, 88) = 5.13$, $p = .026$, $\eta_p^2 = .055$, since overall ratings were higher for participants who had been exposed to the exemplars, and ratings for the categorization test were generally higher than for the familiarity test. This may have been because a few participants in the no-exposure group rated all test stimuli as zero on familiarity (*definitely not familiar*), while an equivalent number of participants in the categorization test rated all stimuli as 50, indicating a noncommittal response. Ratings for the category prototype were significantly higher for the exposed groups than for the nonexposed groups, $F(1, 88) = 35.50$, $p < .001$, $\eta_p^2 = .287$, but did not differ according to test, $F(1, 88) = 2.22$, $p = .140$, $\eta_p^2 = .025$. There was also no significant interaction between exposure group and test group, $F < 1$.

In summary, it appears that overall, participants were able to discriminate between new and old exemplars, and produce prototypicality gradients, and these effects depended on whether participants underwent the mock-subliminal or exposure phase, while there seemed to be no effect of test question.

Fig. 2 Category endorsement (upper panels, **a** and **b**) and familiarity ratings (lower panels, **c** and **d**) for new and old test stimuli for groups of participants who were either exposed to the stimuli (left panels, **a** and **c**) or were not exposed to the stimuli (right panels, **b** and **d**) in Experiment 1

Exposure groups

To further explore what participants in the no-exposure groups learned and whether equivalent prototypicality gradients were obtained for categorization and recognition tests, the two no-exposure groups were analyzed separately to the two exposure groups in two 2 (test group: categorization vs. familiarity) $\times$ 2 (novelty) $\times$ 10 (distortion level) ANOVAs. For the two groups who were exposed to the exemplars (Ex-Fam and Ex-Cat), there was a significant main effect of novelty, $F(1, 44) = 25.09$, $p < .001$, $\eta_p^2 = .363$, and a significant linear trend for distortion level, $F(1, 44) = 153.9$, $p < .001$, $\eta_p^2 = .778$. None of these effects interacted with test group, largest $F < 1$, and there was no main effect of test group, $F < 1$, and no three-way interaction, $F < 1$. Thus, it is reasonable to conclude that equivalent prototypicality gradients were found for both familiarity and categorization tests, and participants were equally able to discriminate between new and old exemplars, giving higher familiarity ratings and category endorsements to old exemplars (see Fig. 2a, c). The high level of similarity in prototypicality gradients for the categorization and familiarity tests suggests that using familiarity ratings to assess the prototypicality gradient is appropriate for the subsequent experiments.

² The false prototypicality gradient was present in the first quarter, and increased from the first to the second quarter when collapsed over groups, $F > 10$. Therefore, learning-at-test effects appeared quite early.

No-exposure groups

A similar analysis was performed on the ratings for the two no-exposure groups. There was a significant linear trend for distortion level, $F(1, 44) = 12.79$, $p = .001$, $\eta_p^2 = .225$, and a significant main effect of test group, $F(1, 44) = 5.34$, $p = .026$, $\eta_p^2 = .108$. Importantly, the linear trend for distortion did not interact with test group, $F(1, 44) = 2.90$, $p = .095$, $\eta_p^2 = .062$. All other main effects and interactions were not significant, largest $F(1, 44) = 1.97$, $p = .167$, $\eta_p^2 = .043$. It appears that our mock subliminal procedure produced a false prototypicality gradient in familiarity ratings, similar to previous studies (e.g., Palmeri & Flanery, 1999). Because an effort was made to ensure that no intrinsic features within our stimuli would enable participants to discriminate between low- and high-distortion exemplars, the most likely explanation was that participants were able to learn about the category on test.² Participants may have either deliberately based their judgments on stimuli seen during the test phase to have some basis for varying their responses, or found it difficult to discount the stimuli they had seen throughout the test phase.

The mock subliminal procedure appears to have been quite convincing. When asked at the end of the experiment, participants allocated to the categorization test claimed to have seen, on average, 12.04 % of the stimuli ($SD = 23.93$, min = 0, max = 99.2) during the mock subliminal phase, and participants allocated to the familiarity test claimed to have seen 15.00 % of the stimuli on average ($SD = 29.87$, min = 0, max = 99.60). While it is possible that a small number of participants

misinterpreted the question, the manipulation check indicates that on the whole, participants were convinced that stimuli had been presented to them in the first phase and this should have provided sufficient motivation to make a serious attempt at judging the stimuli presented in the test phase.

In summary, the categorization and familiarity tests produced very similar prototypicality gradients and differences between ratings for old and new exemplars. Thus, the following experiments used familiarity ratings only to assess both of these effects. The familiarity test was chosen over the categorization test to more closely approximate other implicit learning paradigms that use recognition to assess explicit awareness of the learned material, and because the test was directly related to the instructions given to the Memorize group (see [Supplementary Materials](#)).

Experiment 2

Experiment 2 compared prototypicality gradients in familiarity ratings between intentional exposure conditions where participants were told to memorize a set of stimuli for a subsequent test (Memorize group), and incidental exposure

conditions where participants performed a visual search task using the same stimuli (Search group). Participants were required to search for a singleton target, which was defined by line width (see Fig. 3 for examples). To ensure that participants were attending to the stimuli in the Search group, they were required to respond according to the identity of the line width of the singleton (i.e., was the “odd one out” thicker or thinner than the others?).

Experiment 2 tested whether the prototypicality gradient would result under more cognitively engaging incidental exposure conditions. If the reason why prototypicality gradients were obtained in healthy participants in previous studies was because of an opportunity for explicit encoding, then we may not expect to see any evidence of learning using a visual search task. Furthermore, comparing the Search group to the Memorize group allowed us to test whether altering the exposure conditions makes any difference to the prototypicality gradient and ability to discriminate between new and old exemplars. As mentioned previously, the choice of an intentional memorize group was to maximize any potential group differences that may exist because previous studies (Gureckis et al., 2011) have showed that manipulating encoding strategies does not lead to differences in test performance.

Fig. 3 Examples of circle and line stimuli seen during the visual search task. Stimuli on the left contain a circle/line randomly chosen to have a thinner line width; stimuli on the right contain a circle/line randomly

chosen to have a thicker line width. Participants responded according to the identity of the singleton (whether it was thicker or thinner)

Method

Participants

Fifty participants (40 female, $M_{\text{age}} = 19.58$ years, $SD = 5.14$) took part in Experiment 2 in exchange for partial course credit. All participants were first-year psychology students at the University of Sydney.

Procedure

Participants were randomly allocated to the Memorize group or the Search group. The stimuli during exposure and test were generated in the same way as in Experiment 1.

Exposure phase If participants were allocated to the Search group, the task was framed as a visual cognition task (see [Supplementary Materials](#) for exact instructions). Their task was to search for an “odd one out” on each trial, which was defined using line thickness for both the circle and line stimuli. The singleton was created by either adding or subtracting from the line width of a randomly chosen feature in each stimulus, and participants had to respond by saying whether the odd one out was thicker or thinner than the other circles/lines by pressing either *A* (thicker) or *L* (thinner) on the keyboard as quickly as possible (see Fig. 3 for examples of the singleton). Participants were also given a sheet with a visual example using squares, where one square was thicker and another was thinner, to refer to. It was emphasized verbally that it did not matter where the odd one out was located within the stimulus, and only the identity of the odd one out mattered for their response. There was no feedback given as to the accuracy of each response to equate the exposure conditions between groups as much as possible. However, participants were given feedback (“Too slow”) if they failed to respond in the given timeframe (3 seconds from stimulus onset) on each trial, and they were told that they could still respond after the stimulus had disappeared. If participants were allocated to the Memorize group, they were told to memorize the stimuli as a whole for a subsequent memory test, and the task was framed as a visual memorization task (see [Supplementary Materials](#) for exact instructions). Note that because participants in the Memorize group would also see a singleton on each trial, the instructions explicitly discouraged participants from attending to specific features. Participants in both groups saw each stimulus for 2 seconds, after which there was a blank ISI of 2 seconds. As in Experiment 1, within each stimulus set there were four repetitions of each of the 20 unique exemplars that amounted to 160 trials in total. The identity of the singleton was not consistent across the repeat presentations of each exemplar, so that its location could not be predicted. The entire exposure phase was presented in one continuous block with no breaks.

Familiarity ratings test The test phase was the same as for the familiarity condition in Experiment 1, except that participants in the Search group were warned that the “odd one out” would no longer be present and that they should not look for it to aid their judgements. Participants in the Memorize group were not given any instructions about the “odd one out” since participants’ responses on a post-experimental questionnaire in a pilot experiment did not suggest that participants noticed the singleton during exposure.³

Results and discussion

Overall, participants were quite good at the visual search task, performing at 88.8 % accuracy ($SD = 6.82$, min = 75.0 %, max = 99.4 %). The test data were analyzed in the same way as in Experiment 1, with the presence of a prototypicality gradient inferred by the presence of a significant linear trend in the familiarity ratings, and recognition evident if the familiarity ratings for old exemplars were significantly higher than that for new exemplars. There was a significant linear trend for distortion, $F(1, 48) = 131.3$, $p < .001$, $\eta_p^2 = .732$, indicating an overall prototypicality gradient (see Fig. 4). The linear trend in distortion interacted with group, $F(1, 48) = 21.85$, $p < .001$, $\eta_p^2 = .313$, and novelty, $F(1, 48) = 10.76$, $p = .002$, $\eta_p^2 = .183$. This was due to the steeper overall generalization gradient in the Memorize group than the Search group, and also the steeper overall gradient for new exemplars than old exemplars (see Fig. 4).

There was an overall main effect of novelty, with significantly higher ratings for old over new exemplars, $F(1, 48) = 25.76$, $p < .001$, $\eta_p^2 = .349$, but the interaction with group fell short of conventional levels of statistical significance, $F(1, 48) = 3.21$, $p = .079$, $\eta_p^2 = .063$. Analyses within each group revealed that both groups were able to distinguish between new and old exemplars in their familiarity ratings, smallest $F(1, 24) = 5.95$, $p = .022$, $\eta_p^2 = .199$. The three-way interaction between group, novelty and distortion level was not significant, $F < 1$. The Memorize group rated the prototype as significantly more familiar than the Search group, $F(1, 48) = 12.47$, $p = .001$, $\eta_p^2 = .206$, and the Memorize group also produced higher overall familiarity ratings than the Search group, $F(1, 48) = 5.47$, $p = .024$, $\eta_p^2 = .102$.

It appears that memorizing the stimuli had a more reliable effect on improving the prototypicality gradient (i.e., increasing the slope of the generalization gradient) than improving recognition. However, because we used a visual analogue scale, the range in which participants used to make their familiarity ratings could have affected the strength of their prototypicality gradient. To illustrate this point, consider a participant from the Memorize group and a participant from the Search group who have equivalent category knowledge. It is possible that despite both

³ In a pilot experiment, one participant out of 14 in the Memorize group reported noticing a singleton in response to the question, “Did you notice anything peculiar about the stimuli?”

Fig. 4 Familiarity ratings in the Memorize (a) and Search (b) groups for new and old test stimuli in Experiment 2

participants showing a significant prototypicality gradient, a participant who had been intentionally memorizing the stimuli may rate low-distortion exemplars higher on familiarity, and high-distortion exemplars lower on familiarity, than their counterpart in the Search group due to higher confidence in their knowledge and therefore greater willingness to use the full range of the scale. The two participants would show the same degree of differentiation between low- and high-distortion exemplars, but the slope of their generalization gradient would differ. Thus, to test whether the two groups differed on their sensitivity in distinguishing high and low-distortion exemplars, we conducted a signal detection analysis.

Signal detection analysis

The distribution of responses for each participant was split into sextiles, creating five thresholds. For the prototypicality index, for each threshold, higher ratings for low-distortion exemplars (distortion level ≤ 0.5) counted as a hit, whereas higher ratings for high-distortion exemplars (distortion level ≥ 0.6) counted as a false alarm (old and new exemplars were collapsed). Figure 5 shows the receiver operating characteristic (ROC) curve for this analysis, and the resulting sensitivity index (dA) calculated from the linear transformation of the curve (see Simpson & Fitter, 1973). The index of sensitivity in discriminating between high and low-distortion exemplars was .859 ($SD = .436$) in the Memorize group and .419 ($SD = .369$) in the Search group, with higher sensitivity in the Memorize group, $F(1, 48) = 14.84$, $p < .001$, $\eta_p^2 = .236$.

We performed a similar analysis (counting higher ratings for old stimuli as hits, and higher ratings for new stimuli as false alarms, collapsed over all distortion levels) for discrimination between new and old exemplars (see Fig. 5). The Memorize group ($dA = .232$, $SD = .239$) showed numerically better recognition than the Search group, but again, this difference was marginally non-significant ($dA = .096$, $SD = .269$), $F(1, 48) = 3.55$, $p = .065$, $\eta_p^2 = .069$. Because both tests of this group difference fell just short of statistical significance, we will reserve discussion of whether this group difference is meaningful until after reporting the equivalent analysis for Experiment 3.

Comparison to Experiment 1

To test whether the prototypicality gradient observed in the Search group was significantly larger than would be expected on the basis of learning at test alone, the Search group was compared to the NoEx-Fam group in Experiment 1. Although this is a between-experiments comparison, the two experiments were run at the same time on the same participant pool. There were significantly higher familiarity ratings overall in the Search group, $F(1, 46) = 7.51$, $p = .009$, $\eta_p^2 = .140$, significantly higher familiarity ratings overall for old exemplars,⁴ $F(1, 46) = 4.38$, $p = .042$, $\eta_p^2 = .087$, and a significant interaction between novelty and group,⁵ $F(1, 46) = 4.23$, $p = .045$, $\eta_p^2 = .084$, because only participants in the Search group were able to discriminate between new and old exemplars (see Fig. 6a). Importantly, while the linear trend for distortion level was significant, $F(1, 46) = 35.50$, $p < .001$, $\eta_p^2 = .436$, it did not interact with group, $F < 1$. There was a significant interaction between novelty and the linear trend in distortion, $F(1, 46) = 5.24$, $p = .027$, $\eta_p^2 = .102$, such that the slope of the generalization gradient was steeper for new exemplars. The three-way interaction was not significant, $F < 1$. There was also no difference in sensitivity in detecting low and high-distortion exemplars between the Search group ($dA = .419$, $SD = .369$) and the no-exposure group from Experiment 1 ($dA = .528$, $SD = 1.26$), $F < 1$ (see Fig. 6b).

To provide stronger evidence for the equivalence of the prototypicality gradients between the Search group and the NoEx-Fam group in Experiment 1, we conducted a Bayes factor (BF) test comparing the slopes and sensitivity index between the two groups. As suggested by Rouder, Speckman, Sun, Morey, and Iverson (2009), a Bayes factor was calculated from the results of a t test using a JZS prior for the alternative hypothesis, which assumes a Cauchy distribution of effect sizes. This distribution assumes a high likelihood of smaller effect sizes, but a larger likelihood of medium-to-large effect sizes than a normal distribution. A BF_{01} of 3 is usually considered the threshold for concluding moderate evidence in favor of the null over the alternative

⁴ This analysis was significant for the circle stimuli only, but did not differ according to stimulus set, $F < 1$.

⁵ This analysis was significant for the circle stimuli only, but there was no interaction with stimulus set, $F(1, 46) = 2.492$, $p = .121$, $\eta_p^2 = .051$.

Fig. 5 Results from the signal detection analysis in Experiment 2. *Upper panels* show ROC curves for discrimination between new and old exemplars (recognition, **a**) and discrimination between low- and high-distortion exemplars (category learning, **b**), with the dotted diagonal line

indicating zero sensitivity. *Lower panels* show dA (sensitivity) measures for recognition (**c**) and category learning (**d**) calculated from a linear transformation of the ROC curve, with error bars representing the standard error of the mean

hypothesis (that the means are different). Using this technique, a Bayes factor of 4.1 was found when comparing the slopes, and a Bayes factor of 4.3 when comparing the sensitivity index (dA). Thus, the null hypothesis was more than 4 times more likely than the alternative, suggesting that the magnitude of the prototypicality gradients was the same for participants who searched through the stimuli, and participants who produced a false prototypicality gradient after no exposure to the stimuli.

Although it is possible that the prototypicality gradient in the Search group in Experiment 2 resulted for different reasons than in the mock-subliminal condition in Experiment 1, the magnitude of the prototypicality gradient (ignoring the overall higher ratings in Experiment 2) is indistinguishable from the prototypicality gradient that emerges in the absence of any exposure, and therefore could be entirely attributed to learning-at-test effects. Nevertheless, it is possible that rather than there being a

qualitative difference between exposure conditions, the difference might be quantitative, and the Search group simply learned less about the stimuli than the Memorize group did. Therefore, Experiment 3 doubled the number of trials in the exposure phase to increase the opportunity for incidental category learning.

Interestingly, participants were still able to discriminate between new and old exemplars even after incidental exposure to the stimuli through visual search. One potential explanation is that the visual search task was too easy and participants were able to encode the stimuli in an explicit way during the exposure phase. In addition to the high levels of accuracy, participants took, on average, 1.47 seconds to respond ($SD = 0.188$, $min = 1.16$, $max = 1.85$) in the visual search task, meaning that they could have been using the residual exposure time to study the stimulus. Accordingly, Experiment 3 aimed to replicate Experiment 2 with double the number of exposure trials and

Fig. 6 **a** Familiarity ratings for new and old test stimuli in NoEx-Fam group in Experiment 1 and Search group in Experiment 2. **b** dA measures for NoEx-Fam group and Search group. NoEx-Fam = no exposure, familiarity test phase

ensure that any effects obtained in the Search group were actually due to incidental learning, by adding an additional group where the stimulus disappeared after a response was made in the visual search task.

Experiment 3

Experiment 3 was a replication of Experiment 2, where the length of the exposure phase was doubled. Since the prototypicality gradients in Experiment 2 in the Search group were found to be no different to those displayed after no exposure in Experiment 1, increasing exposure to the stimuli may also increase the magnitude of the prototypicality gradient, if it can indeed result from incidental learning conditions. Furthermore, although Experiment 2 found no significant difference between Memorize and Search groups in terms of their ability to distinguish new from old exemplars, small numerical differences in the predicted direction were apparent, and statistical tests of these differences were marginal (e.g., in Fig. 5c). Any differential recognition between groups that may exist may become clearer with greater exposure to the exemplars.

Experiment 3 also sought stronger evidence that learning in the Search group was actually due to incidental exposure conditions. Because it was necessary to equate the exposure time per trial between groups, it is possible that participants in the Search group were not searching through the stimulus for the entirety of the 2-sec exposure duration. In Experiment 3, a second Search group was added (Search-Terminate) where the stimulus exposure terminated after a response was made, ensuring that participants would only be exposed to the stimuli while searching for and responding to the target. Although the total exposure time between the two Search groups would not be equated, a comparison between the two groups would enable us to determine whether the residual post-response time makes any difference to test performance.

Method

Participants

Fifty-seven University of Sydney first-year psychology students ($M_{\text{age}} = 19.61$ years, $SD = 3.06$, 38 female) participated in this experiment in exchange for partial course credit.

Procedure

The procedure was identical to Experiment 2 except for the following changes. The number of trials in the exposure phase was doubled to 320 trials (eight presentations of each individual exemplar for both circle and line stimulus sets). Participants could now be allocated to the Memorize group, Search group (same as the Search group from Experiment 2), or Search-Terminate group. Participants in the Search-Terminate group were given instructions identical to the Search group, except they were told

that the stimulus would disappear once their response was made. If a response was not made within 2 seconds, then the stimulus disappeared, regardless. Because the total trial time remained the same between the two Search groups (4 seconds), the blank ISI could range from 2 to 4 seconds and depended on how long participants took on each trial to make a response. As in the previous experiments, the test phase consisted of 84 trials that included the 20 old exemplars, 20 new exemplars, and the prototype presented twice for each stimulus set.

Results and discussion

After excluding one participant from the Search group because of poor performance (45.0 % accuracy), the average accuracy in the Search group was 92.3 % ($SD = 5.06$, min = 79.4, max = 94.7), which was significantly higher than in the Search-Terminate group ($M = 83.9$ %, $SD = 8.82$, min = 61.2, max = 94.7), $t(34) = 3.51$, $p = .001$, $SED = .024$. The average reaction time (RT) in the visual search task in the Search group was 1.42 seconds ($SD = .227$, min = 0.89, max = 1.75), which was significantly slower than in the Search-Terminate group, ($M = 1.25$ seconds, $SD = .153$, min = 0.91, max = 1.54), $t(34) = 2.55$, $p = .016$, $SED = .064$. Thus, it appears that making the stimulus disappear after participants responded resulted in poorer accuracy, but faster RTs.

Memorize versus search

Because there were three groups in Experiment 3, two separate ANOVAs were run in a similar manner to Experiment 2, one comparing the Memorize group to the two Search groups, and another comparing the two Search groups to each other. This was done so that the effects of exposure conditions and exposure time during visual search could be examined separately. For the comparison between the Memorize group against the two Search groups, there was a significant main effect of novelty,⁶ $F(1, 54) = 33.39$, $p < .001$, $\eta_p^2 = .382$, which did not interact with group, $F < 1$. This indicates that there was an overall ability to discriminate between new and old exemplars that did not vary according to exposure condition, similar to Experiment 2 (see Fig. 7). There was an overall linear trend for distortion, $F(1, 54) = 139.4$, $p < .001$, $\eta_p^2 = .721$, which did not interact with novelty, $F < 1$, suggesting an overall prototypicality gradient that was equivalent for new and old exemplars. Critically, the linear trend for distortion interacted with group, $F(1, 54) = 5.21$, $p = .026$, $\eta_p^2 = .088$, replicating Experiment 2. The three-way interaction was not significant, $F < 1$, and the main effect of group was also not significant, $F < 1$. Overall, familiarity ratings for the prototype also did not differ according to group, $F < 1$. Therefore,

⁶ This result was significant for the line stimuli only, but did not differ according to stimulus set, $F < 1$.

Fig. 7 Familiarity ratings in the Memorize group (a), Search group (b), and Search-Terminate group (c) for new and old test stimuli in Experiment 3

similarly to Experiment 2, the prototypicality gradient was stronger overall in the Memorize group than the two Search groups, but this time there was no clear evidence of an advantage in recognition for the Memorize group despite the extension of the exposure phase.

Search versus search-terminate

Comparing the two Search groups to each other, there was again a significant main effect of novelty, $F(1, 34) = 17.41, p < .001, \eta_p^2 = .339$, that did not interact with group, $F < 1$. There was a significant linear trend for distortion, $F(1, 34) = 69.74, p < .001, \eta_p^2 = .672$, which did not interact with novelty, $F < 1$, or group, $F < 1$. The three-way interaction was not significant, $F < 1$, and the main effect of group was also not significant, $F < 1$. Thus, there was no evidence to suggest that the residual exposure time in the Search group benefitted overall recognition or the prototypicality gradient compared to the Search-Terminate group when the stimulus disappeared after a response was made. Ratings for the prototype alone also did not differ between the two Search groups, $F(1, 34) = 3.15, p = .085, \eta_p^2 = .085$.

Signal detection analysis

A signal detection analysis was conducted in a similar manner to Experiment 2, comparing the Memorize group against the two Search groups and then the two Search groups to each other (see Fig. 8). For recognition, there was no difference in sensitivity (dA) comparing the Memorize group with the two Search groups, $F < 1$, nor when comparing the two Search groups to each other, $F < 1$. The Memorize group did have an advantage in discriminating high- and low-distortion exemplars when compared to the two Search groups, $F(1, 54) = 6.05, p = .017, \eta_p^2 = .101$, but there was no difference in sensitivity between the two Search groups, $F < 1$ (see Fig. 8). Therefore we can conclude that the advantage in the prototypicality gradient for the Memorize group was not due to participants using the scale differently to the other groups.

Comparison to Experiment 1

To test whether the increased exposure resulted in a prototypicality gradient that was greater than a learning-at-test effect, we combined the two Search groups in this experiment and compared them with the NoEx-Fam group in Experiment 1. As in Experiment 2, there was an overall ability to discriminate between new and old exemplars,⁷ $F(1, 57) = 8.99, p = .005, \eta_p^2 = .133$, and this interacted with group, $F(1, 57) = 8.74, p = .004, \eta_p^2 = .141$, since participants in Experiment 1 did not see any stimuli and thus could not recognize them in the subsequent test phase. The linear trend for distortion was significant, $F(1, 57) = 53.48, p < .001, \eta_p^2 = .484$, but again, importantly, the linear trend did not interact with group, $F(1, 57) = 1.70, p = .198, \eta_p^2 = .029$. The signal detection analysis comparing dA between experiments was also not significant, $F < 1$, with similar levels of sensitivity in the two Search groups in this Experiment ($dA = .482, SD = .330$) and the no-exposure group in Experiment 1 ($dA = .528, SD = .126$). While we acknowledge that this is a between-experiments comparison, the results are consistent with Experiment 2, and between two methods of analysis. To provide stronger evidence for the null hypothesis, we conducted a Bayes factor test in a similar manner to Experiment 2, following Rouder et al.'s (2009) technique. This resulted in a Bayes factor of 2.04 for the comparison of the gradient slopes, and a Bayes factor of 4.9 for the comparison of the sensitivity index (dA) between groups, both in favor of the null. While the BF for the slopes cannot be considered good evidence in favor of the null, the BF for the sensitivity index suggests that, consistent with Experiment 2, participants in the Search groups discriminated between high-distortion and low-distortion exemplars to an equivalent extent as those in the no-exposure group.

The results of Experiment 3 suggest that doubling the number of exposure trials did not exacerbate any potential group differences in recognition ability suggested in Experiment 2, but the advantage for the Memorize group for the prototypicality

⁷ This result was significant for the line stimuli only, but did not differ according to stimulus set, $F < 1$.

Fig. 8 Results from the signal detection analysis in Experiment 3. *Upper panels* show ROC curves for discrimination between new and old exemplars (recognition, **a**) and discrimination between low- and high-distortion exemplars (category learning, **b**), with the dotted diagonal line

indicating zero sensitivity. *Lower panels* show *dA* (sensitivity) measures for recognition (**c**) and category learning (**d**) calculated from a linear transformation of the ROC curve, with error bars representing the standard error of the mean

gradient remained. Increasing exposure still did not result in a prototypicality gradient greater in magnitude than that expected on the basis of learning-at-test alone in the Search groups, implying that the prototypicality gradient is not implicit in the sense of resulting automatically from incidental exposure. On the other hand, the lack of any group differences between the two Search groups suggests that the ability of participants to recognize old exemplars at test does arise from incidental exposure, and not from using the residual exposure time after responding to explicitly encode the stimuli.

General discussion

This study tested the implicit status of learning in the prototype distortion task by introducing a visual search task as a means of incidental exposure, and tested the effect of manipulating encoding conditions by comparing a visual search group to a group who were simply asked to memorize the stimuli. Our methodology was novel in the sense that we used the same measure (familiarity ratings) and also the same test stimuli to assess both the prototypicality gradient and ability to discriminate between new and old exemplars. Surprisingly, we found no evidence that participants learned about the similarity structure of the stimuli during visual search, and a dissociation of the opposite nature to those commonly reported in amnesia studies

(e.g., Knowlton & Squire, 1993) and healthy participants (Nosofsky & Zaki, 1998), with intentional memorization improving prototypicality gradients but having a much less reliable effect on improving participants' ability to discriminate between new and old exemplars.

Experiment 1 showed that the magnitudes of the prototypicality gradients and discrimination between old and new exemplars were equivalent for categorization and familiarity tests for participants who were exposed to the stimuli, justifying our use of familiarity ratings for the subsequent experiments to assess the prototypicality gradient. It can be argued that in our decision to use a single measure in Experiments 2 and 3, we have not directly tested participants' ability to categorize the stimuli. While we acknowledge that familiarity judgements and category endorsements are potentially different in terms of the decision rules they engage (Nosofsky, 1988), and the responses they can produce (Squire & Knowlton, 1995), it is clear that category structure was reflected in the prototypicality gradients for familiarity ratings, and these prototypicality gradients were sensitive to our group manipulations. It remains to be seen whether our results replicate using a categorization test in place of the familiarity test. Presumably, in our particular task where participants are exposed to a single category, visual similarity would be the primary determinant of both category endorsements and familiarity ratings because the categories are not defined using binary features or rules.

Experiment 1 also addressed an issue brought up by previous studies, that prototypicality gradients can result in the absence of exposure. By comparing generalization gradients

between a group of participants who were misled into believing that they had been shown the stimuli in a mock-subliminal procedure (Palmeri & Flanery, 1999), against a group of participants who actually were exposed to the stimuli, we showed that our stimuli produced false prototypicality gradients and thus concluded that participants were learning about the category on test. This result, along with the false prototypicality gradients displayed in other studies (e.g., Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004), exemplifies a general problem with assessing A/Not-A category learning in the prototype distortion task. It is difficult to ensure that effects displayed on test are due to exposure alone, and participants who have not learned anything during the initial exposure phase may feel that they need to provide some variation in their responses and thus seek out information to enable them to do so on test. If this is indeed an unavoidable problem with any stimulus set, a similar procedure to that of Experiment 1 should be employed in future studies before claiming that a prototypicality gradient exists.

Experiments 2 and 3 compared ability to discriminate between new and old exemplars and prototypicality gradients using familiarity ratings between a group who searched through the category exemplars for a singleton, and a group who attempted to memorize the exemplars for a subsequent familiarity test. Experiments 2 and 3 showed that the Memorize groups produced steeper prototypicality gradients than the Search groups, while there was no strong evidence for a similar advantage in discriminating new and old exemplars. While the Search groups in Experiments 2 and 3 displayed an ability to discriminate between new and old exemplars that was above chance, the prototypicality gradient (i.e., generalization gradient) displayed after performing the visual search task was found to be no different to the false prototypicality gradient displayed in Experiment 1 after no exposure. Doubling the length of exposure from Experiment 2 to Experiment 3 did not have any effect on the general pattern of results and if anything, weakened the (nonsignificant) advantage for the Memorize group in recognition in Experiment 2. Furthermore, Experiment 3 showed that there were no differences between a visual search group who were exposed to the stimuli for the set duration (2 seconds), and a visual search group for whom the stimulus disappeared after a response was made. This allowed us to conclude that the ability to recognize old exemplars in the Search group was in fact due to learning that occurred during visual search, and not due to deliberate encoding that may have occurred after a response was made.

Learning during visual search

Contrary to claims in the literature (e.g., Smith, 2008; Smith & Grossman, 2008), the current study found no support for the idea that learning in the prototype distortion task is implicit in the sense of resulting from an automatic learning process. If this were the case, then the prototypicality gradient should have

resulted as a consequence of exposure to the stimuli during the visual search task, despite learning about the stimuli being incidental to the requirements of the task (searching for, and responding to the identity of the singleton). Because the magnitude of the prototypicality gradient displayed in the Search groups across Experiments 2 and 3 was not found to be greater than the false prototypicality gradient displayed in Experiment 1, we cannot claim to have strong evidence of learning on the basis of the presence of a prototypicality gradient. Although this conclusion rests on a null result, the sample sizes used in our experiments were sufficient to detect significant group differences in prototypicality gradients, and the Memorize groups in Experiments 2 and 3 consistently showed prototypicality gradients that were substantially larger in magnitude than those displayed in Experiment 1. Bayes factor analyses also support the null hypothesis when comparing the slope of the prototypicality gradient as well as the sensitivity index between groups.

Our failure to find prototypicality gradients after visual search may seem surprising given that there is evidence from other paradigms, such as contextual cueing that learning reliably occurs as a consequence of repeated exposures in visual search (e.g., Chun & Jiang, 1998; 2003; Colagiuri & Livesey, 2016). There are, however, several notable differences between the two paradigms. For example, in our visual search task, although features of the *category* were certainly repeated throughout the exposure phase, and each unique exemplar was repeated a number of times, the stimulus configurations did not predict the location of the singleton, and our assessment of learning took place after the exposure phase, meaning that it may have been harder to detect incidental learning in general.

Another reason why only a weak prototypicality gradient was observed (i.e., equivalent to no exposure) could be that the Search groups focused their attention on finding a singleton during the exposure phase, which was subsequently removed in the test phase. While the exposure and test stimuli were equated between groups, and the Search groups were explicitly instructed not to look for the singleton on test, it is possible that the Search groups were affected more than the Memorize groups by this small change between the exposure and test stimuli. If there was indeed greater generalization decrement in the Search group, this should have lowered familiarity ratings for the old stimuli more than the new stimuli, weakening the level of discrimination between old and new stimuli. The interaction between group and novelty was not significant in Experiment 2 or Experiment 3, suggesting that familiarity ratings were lower in the Search group by an equivalent magnitude for old *and* new stimuli. Also, because the Search groups were able to discriminate between old and new exemplars on test, it appears that any generalization decrement suffered by the Search group was minimal.

The failure to find a prototypicality gradient in the Search group is more puzzling because participants were able to give higher

familiarity ratings to old stimuli, demonstrating that they had learned *something* about the stimuli. One way to explain these results is that the prototypicality gradient displayed after the mock-subliminal procedure in Experiment 1 and the prototypicality gradient displayed after visual search were not the consequence of the same learning process despite being of comparable magnitude (see Smith, 2008). In other words, perhaps the latter was a genuine consequence of learning during training and was not further influenced by learning on test. If the visual search task forces participants to encode the features of the stimuli serially, on test, when participants are asked to judge familiarity of the stimuli as a whole, the Search group may resort to searching through the stimuli for specific features that they recognize. This technique might be sufficient to allow a prototypicality gradient to emerge, but one that is not larger in magnitude to that in Experiment 1. Participants who underwent the mock-subliminal procedure, in contrast, may have used the whole stimulus at test to make their categorization or familiarity judgments. However, because they have not been exposed to any stimuli in the exposure phase, what they can learn on test is obviously limited, and thus their prototypicality gradient is similarly small in magnitude to the Search groups. Unfortunately, our study provides no means to determine whether this account of the failure to obtain prototypicality gradients after visual search is true. Clearly, further studies are needed to clarify whether incidental learning is conducive to producing prototypicality gradients under different encoding conditions (e.g., serial vs. configural feature encoding).

If we assume that the prototypicality gradients obtained after visual search were due to incidental learning, another way in which the prototypicality gradient can be interpreted as implicit is if it results in the absence of explicit recognition, consistent with the suggestion that prototype effects are not dependent on declarative memory (Knowlton & Squire, 1993; Reber & Squire, 1999). Again, this was not supported by our data, with the Search groups in Experiments 2 and 3 showing higher familiarity ratings for old exemplars, consistent with contextual cueing studies that have found above-chance recognition (e.g., Smyth & Shanks, 2008).

The effect of encoding conditions

Importantly, Experiments 2 and 3 found that varying the nature of the exposure conditions does affect the ability of participants to show a prototypicality gradient, with participants in the Memorize groups consistently showing an advantage over the Search groups. Our results stand in contrast to those of Gureckis et al.'s (2011), who found no differences in test performance between groups of participants who were asked to memorize the stimuli as a configural whole, and participants who were asked to imagine pointing to the center dot. As mentioned previously, one explanation for their results is that their incidental task is not cognitively demanding, and thus participants may have been able to explicitly encode the stimuli while performing the

task. An alternative explanation is that the task itself (pointing to the middle dot) might result in a similar deployment of visual attention as the memorization task. To speculate, pointing to the middle dot might also involve configural processing of the stimuli as participants would have to encode spatial relations between the features (in effect processing the stimulus as a whole) in order to determine where the center of the pattern was located, thus making the two encoding strategies similar in terms of what participants attend to. Strong prototypicality gradients that have been obtained in the past under incidental exposure conditions may therefore be largely due to the nature of the encoding conditions facilitating later performance, rather than any incidental learning that may have occurred.

In contrast, in our study, we used a more cognitively engaging visual search task, demonstrating that test performance can vary between different exposure conditions. However, our visual search task requires participants to search through individual features of the stimuli to find the target, and thus differs both in terms of the requirements of the task and what features (configural vs. specific) of the stimuli participants focus on. As mentioned above, it is possible that the Search group did learn incidentally, but only about the specific features of the stimuli. The Memorize groups, on the other hand, were told to memorize the stimuli as a whole, and thus were encouraged to encode the configural as well as specific features of the stimuli. A prediction concerning the impact of these encoding strategies on test can be derived by considering how these changes in encoding may affect generalization to previously unseen high- and low-distortion exemplars.

Generalization increases as the similarity between the test stimulus and the seen exemplars increases.⁸ The low-distortion exemplars are very similar to the prototype, and therefore also very similar to each other, whereas the high-distortion exemplars are relatively dissimilar to each other. Thus, the prototypicality gradient in the old test stimuli can be seen to result in the following way: generalization is high (and therefore high familiarity ratings are given) for low-distortion exemplars because of a high degree of similarity to other seen low-distortion exemplars. Generalization is reduced for high-distortion exemplars because these exemplars are not as similar to the seen exemplars (whether they be high or low distortion). If we assume that the Memorize group are better able to detect overall similarity (e.g., due to encoding of configural features of the stimuli), then generalization will be higher in this group for the low-distortion exemplars and thus produce a steeper prototypicality gradient. In contrast, the Search group may have more difficulty detecting similarity between stimuli (e.g., because they have only encoded the specific features of the stimuli in a serial fashion, and the exemplars are created by distorting each dimension of each feature individually). As mentioned above, if the Search group were to adopt

⁸ This explanation can also be framed in terms of generalization from an extracted prototype.

the strategy at test of searching for individual features that they recognize, this might allow them to discriminate between old and new exemplars to the same degree as the Memorize group, but would not enable them to produce as strong a prototypicality gradient. Thus, a speculative conclusion is that our visual search task may have resulted in incidental learning, but through a manner of encoding that was very different to the Memorize group.

An alternative way to explain the group differences in Experiments 2 and 3 is through awareness of the subsequent familiarity test, as this was necessarily part of the encoding instructions in the Memorize group. Although it is certainly possible that any group differences observed could be attributed to this awareness and not the difference in encoding conditions, the results of Gureckis et al. (2011) and Reber et al. (2003) do not support this idea. Both studies failed to find a difference in categorization performance when comparing groups that were aware of the existence of a category against a group who were unaware of the category (but note the small sample size in both of these studies). This suggests that awareness of a category either does not lead participants to look for similarities between stimuli to aid them in their category judgments, or alternatively that the nature of the stimuli makes this difficult. Although our study uses a familiarity test rather than a categorization test, participants in the Memorize group knew that they would be tested on their memory, and yet in both Experiments 2 and 3 there was little evidence that this facilitated their performance in discriminating between new and old exemplars. Thus, if anything, we would need to conclude counterintuitively that awareness of a subsequent memory test facilitates prototypicality gradients, but not recognition ability. It seems that what matters for learning in the prototype distortion task is not whether participants are aware of the existence of a category or an impending memory test, but what demands the particular encoding conditions make on the participant.

The dissociation between categorization and recognition

Finally, our manipulation of encoding conditions in Experiments 2 and 3 increased the magnitude of the prototypicality gradient, but did not reliably improve discrimination between new and old exemplars. Although the results of Experiment 2 were suggestive of a similar advantage for the Memorize group in both the prototypicality gradient and ability to discriminate between old and new exemplars, the results of Experiment 3 showed the opposite of what has been found when comparing amnesic patients to healthy controls, where there is usually a discrepancy in recognition but similar categorization performance (e.g., Knowlton & Squire, 1993). This result might be surprising given that our use of a single measure should have eliminated potential differences in response thresholds and task-specific variance. However, there is still a difference in the statistical measures we used for the prototypicality gradient (linear trend in ratings) and recognition (difference in ratings

between old and new stimuli). Thus, there is still a parameter difference between these chosen indices. For example, it may be that they are not equally easy effects to obtain using familiarity ratings alone.

It is also worth emphasizing that if participants are learning about the category on test, as Experiment 1 suggests, then discriminating new from old stimuli may become more difficult, since participants would be explicitly encoding both old *and* new exemplars on test. This would presumably strengthen, or make no difference to the prototypicality gradient but may contribute to a weaker level of recognition. This potential problem also applies to most of the prototype effect literature because test stimuli are typically a combination of old and new stimuli and usually many (if not all) are structured around the prototype to some extent. Thus, for the same reasons, their presentation on test should strengthen the prototypicality gradient but weaken participants' ability to recognize exemplars they have seen. Ensuring that difficulty level is equated across assessment of the prototypicality gradient and recognition is thus important for interpretation of dissociations but also difficult to implement.

To summarize, our results call into question the logic of interpreting single dissociations as evidence for implicit category learning due to the inherent difficulty in ensuring a fair comparison between tests of categorization and recognition. Because we have shown a pattern of results that is the reverse of the majority of studies comparing amnesiacs to controls (e.g., Knowlton & Squire, 1993), it seems that the way in which the prototype effect and recognition are assessed has a large influence on the pattern of results obtained, and even when attempts are made to minimize differences between tasks by using the same test stimuli and test measure, dissociations are still possible.

Conclusion

In conclusion, this study does not support the characterization of learning in the prototype distortion task as implicit in the sense of resulting automatically from incidental exposure during visual search, and found a dissociation opposite to that commonly reported in the literature when the same test measure was used to assess prototypicality gradients and recognition. Our findings highlight the need to test for potential learning-at-test effects before claiming that learning exists, and to exercise caution in interpreting dissociations between categorization and recognition due to the difficulty in eliminating potential parameter differences that can cause dissociations. Our study emphasizes the importance of studying the encoding strategies and exposure conditions that are required for so-called implicit learning effects.

Acknowledgments The authors declare that they have no conflicts of interest in conducting this research.

This research was supported by Discovery Grant DP120102190 awarded to E.J.L. from the Australian Research Council.

The authors would like to thank Ian McLaren for helpful comments on an earlier version of this manuscript.

References

- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, 56, 149–178.
- Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, 1224, 147–161.
- 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, 442–481.
- Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 348–362.
- Berry, C. J., Henson, R. N. A., & Shanks, D. R. (2006). On the relationship between repetition priming and recognition memory: Insights from a computational model. *Journal of Memory and Language*, 55, 515–533.
- Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia*, 44, 816–827.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Chun, M. M., & Jiang, Y. (1998). Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28–71.
- Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 224–234.
- Cleeremans, A., & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M. Stadler & P. Frensch (Eds.), *Handbook of implicit learning* (pp. 323–364). Newbury Park: SAGE.
- Colagiuri, B., & Livesey, E. J. (2016). Contextual cuing as a form of nonconscious learning: Theoretical and empirical analysis in large and very large samples. *Psychonomic Bulletin and Review*. doi:10.3758/s13423-016-1063-0
- Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 495–532). Thousand Oaks: SAGE.
- Gureckis, T. M., James, T. W., & Nosofsky, R. M. (2011). Re-evaluating dissociations between implicit and explicit category learning: An event-related fMRI study. *Journal of Cognitive Neuroscience*, 23, 1697–1709.
- Kéri, S., Kálmán, J., Rapcsak, S. Z., Antal, A., Benedek, G., & Janka, Z. (1999). Classification learning in Alzheimer's disease. *Brain*, 122, 1063–1068.
- Kéri, S., Kálmán, J., Kelemen, O., Benedek, G., & Janka, Z. (2001). Are Alzheimer's disease patients able to learn visual prototypes? *Neuropsychologia*, 39, 1218–1223.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747–1749.
- Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, 3, 172–179.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399–1402.
- Kunar, M. A., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, 33, 816–828.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 700–708.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, 9, 247–255.
- Nosofsky, R. M., Denton, S. E., Zaki, S. R., Murphy-Knudsen, A. F., & Unverzagt, F. W. (2012). Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 860–880.
- Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences*, 109, 333–338.
- Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, 10, 526–530.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Poldrack, R. A., & Forde, K. (2008). Category learning and the memory-systems debate. *Neuroscience and Biobehavioral Reviews*, 32, 197–205.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Posner, M. I., Goldsmith, R., & Welton, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, 73, 28–38.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience*, 113, 235–242.
- Reber, P., Gitelman, D., Parrish, T., & Mesulam, M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, 15, 574–583.
- Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience*, 113, 411–419.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16, 225–237.
- Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, 80, 481–488.
- Smith, E. E. (2008). The case for implicit category learning. *Cognitive, Affective, & Behavioral Neuroscience*, 8, 3–16.
- Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews*, 32, 249–264.
- Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory and Cognition*, 36, 403–415.
- Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 12470–12474.
- Stadler, M. A., & Frensch, P. A. (1994). Whither learning, whither memory? *Behavioral and Brain Sciences*, 17, 423–424.

- Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, 23, 87–102.
- Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin and Review*, 11, 1048–1054.
- Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory and Cognition*, 32, 390–398.
- Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, 9, 394–406.