

Contents

Chapter 1	5
Lit Review	5
Chapter 2	6
Experiment 1: Establishing PSEs in novel Remember/Know paradigm	12
Method	12
Participants	12
Materials	12
Design	14
Procedure	14
Data processing	15
Results	16
Picture superiority	16
PSE in rates of Recollection and Familiarity:	18
PSE and the availability of different response options:	23
Discussion	25
Chapter 3	28
Experiment: Development of a new set of standardised photographic stimuli	31
Method	32
Participants	32
Materials	32

Design	35
Procedure	36
Data processing	38
Analysis preparation	40
Results	41
<i>Naming</i>	41
<i>Mental imagery agreement</i>	42
<i>Familiarity</i>	43
<i>Visual complexity</i>	43
<i>Selection of final items</i>	43
Discussion	45
<i>The role of colour</i>	45
<i>Establishing a new set of stimuli</i>	47
Experiment: Effect of stimuli format (greyscale) and response option on recognition memory judgements.	48
Method	49
Participants	49
Materials	50
Design	50
Procedure	51
Data processing	52
Results	53
Stimuli distinctiveness	53

Chapter 4 (The role of colour)	57
Experiment: Effect of stimuli format (colour) and response option on recognition memory judgements	59
Method	59
Participants	59
Materials	59
Design	60
Procedure	61
Data processing	61
Results	62
Distinctiveness	62
Response option	63
Role of colour	64
Discussion	64
Chapter 5	65
Experiment: Manipulating word distinctiveness	65
Method	65
Results	66
Manipulation check	68
Elimination of the PSE: comparison with Ensor et al. (2019)	73
Recollection and Familiarity	76
References	82

Appendices	95
Appendix A: Spelling corrections / manipulations to naming responses.	95
Appendix B: Normative data for all photograph items.	98

Chapter 1

Lit Review

Hamilton and Geraci (2006)

IMPLICIT MEMORY: PSE results from conceptual processing of a picture's distinctive features (rather than semantic information). General semantic task: "What is a used car sometimes called?" No PSE. Distinctive conceptual information task: "What fruit is egg shaped?" PSE.

EXPLICIT RECOGNITION: PSE always evident?

aMCI Show larger PSE effects than controls.

Impaired REC, so this PSE must rely on FAM?

Mixed findings whether fam is intact in aMCI. Intact - generally use picture stim. Impaired - generally use verbal stim.

Is PSE in aMCI driven by intact FAM for pictures, but impaired FAM for words? Yes (Embree, Budson, & Ally, 2012): aMCI - Picture FAM - same as healthy OAs aMCI - Word FAM - impaired compared to healthy OAs

Ally, McKeever, 2009: Examined early frontal old/new effect (FAM) in aMCI: Intact for pictures. Impaired for words. BUT, P did not provide subjective Rec/Fam reports.

Embree, Budson, & Ally, 2012: Deep encoding (verbal like/dislike response). Modified Old/New (6-point rating scale): 6. Certain the item is old - to - 1. Certain the item is new.

Both used the same picture stim - colour photos.

###-----

Chapter 2

Dual-process theories of recognition memory suggest that two independent processes - recollection and familiarity - are implicated in the successful recognition of previously encountered material (Paivio, 1971, 1972). Recollection typically refers to the conscious recall of encoded information, whereby contextual details (usually obtained by mentally re-experiencing a previous encounter with the stimulus) facilitate successful recognition. Familiarity, on the other hand, describes the unsubstantiated *feeling* of having encountered the stimulus before, and despite the inability to retrieve any associated diagnostic information, is still able to produce accurate recognition (Schoemaker, Gauthier, & Pruessner, 2014). While single-process accounts of recognition memory have been proposed, with the view that such experiences can be understood simply as varying levels of memory strength (Dunn, 2008; Squire, Wixted, & Clark, 2007), the majority of memory researchers agree that multiple processes are necessary to account for a range of dissociable experimental findings (Yonelinas, 2002). Evidence from studies utilising event related potentials (ERPs; Curran & Doyle, 2011), functional magnetic resonance imaging (fMRI; Scalici, Caltagirone, & Carlesimo, 2017) and comparisons between healthy and clinical subject groups (e.g. Mild Cognitive Impairment; Belleville, Ménard, & Lepage, 2011) all implicate the existence of two functionally distinct processes. Despite this consensus, disagreement persists in the literature regarding the extent to which recollection and familiarity are independent, and the methods that should be used to measure them most effectively (Schoemaker et al., 2014; Yonelinas, 2002).

Experiments into recognition memory often focus on obtaining separate estimates of recollection and familiarity using process-estimation methods (Yonelinas, 2002). The most commonly used process-estimation method is the Remember/Know (RK) paradigm (Tulving, 1985) - a task endorsed by a wide body of literature (Gardiner, 2000; Jacoby, 1991; Jacoby, Yonelinas, & Jennings, 1997; Yonelinas & Jacoby, 1995). In a typical RK procedure, participants are generally tasked with making 'old' vs. 'new' recognition decisions toward a randomised list of items, many

of which were presented during an earlier encoding phase (targets) amongst novel items with highly similar characteristics (lures). When a subject recognises an item, and thus selects *Old*, a follow-up judgement probes how they arrived at this decision (*Remember* or *Know*). If the subject was able to recognise the item based on recollection (i.e. conscious recall of some diagnostic information: “I remember seeing this item earlier”), they should classify their recognition as *Remember*. If the subject arrived at their recognition decision due to familiarity (i.e. a feeling of certainty that the item was studied in the encoding phase, but unable to recall and details: “I know I saw this item earlier, but cannot determine why”), they should classify their recognition as *Know*. In addition to the literature endorsing the task in healthy samples, a large body of research also reports that the RK procedure produces reliable estimations of recollection and familiarity in clinical populations (Lombardi, Perri, Fadda, Caltagirone, & Carlesimo, 2016); for example, those with Mild Cognitive Impairment (MCI) typically produce results to suggest recollection impairments but intact familiarity compared to healthy older adults (Belleville et al., 2011; Hudon, Belleville, & Gauthier, 2009; Lombardi et al., 2016; Serra et al., 2010; Wang et al., 2013).

The RK procedure has been modified in a number of ways since its conception, and continues to adapt as understandings of recollection and familiarity processes evolve. An early development was the “independence correction” - a formula devised to ‘correct’ the inherent underestimation of familiarity processes within the mutually exclusive paradigm (Yonelinas & Jacoby, 1995). Participants are generally only instructed to select *Know* (a reflection of familiarity) when there is an absence of recollection, however, this approach does not allow for the possibility of recollection and familiarity co-occurring. Proportions of *Know* responses will likely always be lower than *Remember* if subjects do indeed perceive to experience both processes simultaneously, since the presence of recollection necessitates that they select the *Remember* option among the two choices. When the Yonelinas & Jacoby (1995) independence correction is applied, estimates of familiarity are determined by also taking into account the number of times *Remember* was selected when calculating the proportion of *Know* responses (Schoemaker et al., 2014). An alternative to this correction is to modify the response options available to subjects, so they are

able to individually determine the relative contributions of each process. Higham & Vokey (2004) proposed an independent ratings methodology whereby, instead of the binary *Remember/Know* options, subjects are provided with one rating scale to report the contribution of recollection and another to report the contribution of familiarity (RF-Ratings). Participants rate their recognition experience for each process accordingly: 1 = *definitely no*, 2 = *probably no*, 3 = *probably yes*, 4 = *definitely yes*. Such options allow for great variability in the way participants are able to respond, and for the possibility of both processes occurring conjointly: i) Recollection without Familiarity (high rating on R, low rating on F); ii) Familiarity without Recollection (high rating on F, low rating on R); iii) both Recollection *and* Familiarity (high rating on R and F); iv) neither R or F, i.e. a guess (1 rating on R and F). The methodology of Higham & Vokey (2004) has been used in numerous studies (Brown & Bodner, 2011; Kurilla & Westerman, 2008; Tousignant & Bodner, 2012), however, it could be argued that this rating task is somewhat removed from the original *judgement* task, and the extent to which the increased task complexity affects reports of recognition is unknown (Tousignant, Bodner, & Arnold, 2015).

Further modifications retain the original two binary response options, but avoid the mutual exclusivity issue by simply including a *Both* option (Tousignant et al., 2015). When calculating proportions of recollection and familiarity, the total proportion of *Both* responses can then be separately added to the totals for each process. Recent adaptations of the RK paradigm have also begun to include a *Guess* response option, allowing participants to report uncertainty in their recognition decision (Belleville et al., 2011; Eldridge, Sarfatti, & Knowlton, 2002; Larsson, Öberg, & Bäckman, 2006; Tunney & Fernie, 2007; Williams, 2019). Previous studies have found that subjects may falsely assign guesses to the *Know* option when there is no explicit *Guess* option available (Gardiner, Java, & Richardson-Klavehn, 1996; Gardiner & Ramponi, 1998; Gardiner, Ramponi, & Richardson-Klavehn, 2002), on the assumption that this option more closely resembles their state of low confidence (Tunney & Fernie, 2007). Responding in this manner may artificially inflate obtained estimates of familiarity (Tunney & Fernie, 2007). By including *Guess*, the likelihood of obtaining false *Know* responses (i.e. those that do not reflect underly-

ing familiarity processes) is reduced (Migo, Mayes, & Montaldi, 2012).

Despite its widespread use, the RK procedure has been criticized for its reliance on participants' subjective understanding of the provided instructions (Schoemaker et al., 2014), and the introspective nature of recognition judgements make it difficult to confirm whether all participants have understood the definitions (and thus responded) similarly (Lombardi et al., 2016). It is also difficult to determine whether subjects interpret the *Remember* and *Know* labels in the same way that researchers intend (Umanath & Coane, 2020), especially as there is evidence to suggest participants struggle to understand the distinction between the terms (Geraci, McCabe, & GUILORY, 2009; Rubin & Umanath, 2015; Williams & Moulin, 2014). Williams (2019) assessed the ways in which non-recollective subjective experiences were defined to participants, and found a great deal of inconsistency across a range of RK experiments. Some studies even changed the *Remember* and *Know* labels altogether; many exchanged the *Know* label with *Familiar* in an effort to reduce subjects defaulting to colloquial understandings of the word "know" which typically indicate high certainty; e.g. "I **know** I saw this item in the study phase" (Bastin, Van der Linden, Michel, & Friedman, 2004; Dobbins, KroU, & Liu, 1998; Donaldson, Mackenzie, & Underhill, 1996; Ingram, Mickes, & Wixted, 2012). Others also substitute *Remember* for *Recollection* (Harlow, MacKenzie, & Donaldson, 2010). Labels that accurately match the processes they intend to measure - *Recollection* and *Familiarity* - have been proposed in an effort to reduce the potentially misleading effects of the more colloquial *Remember* and *Know*, and thus make it easier for participants to 'map on' the definitions provided by researchers (Harlow et al., 2010; Mayes, Montaldi, & Migo, 2007).

In addition to the availability of different response options, and the labels used to describe the underlying processes, there is evidence to suggest that the format of to-be-remembered stimuli also plays a role in obtained estimates of recollection and familiarity. The Picture Superiority Effect (PSE) refers to a robust phenomenon whereby stimuli presented as pictures are markedly

better remembered on tests of recall or recognition than stimuli presented as words (Shepard, 1967). There is general agreement that, in recognition memory paradigms, picture superiority manifests as enhanced recollection rather than familiarity (Curran & Doyle, 2011; Rajaram, 1996a). Word stimuli, on the other hand, appear to produce increased familiarity ratings at test (Ally & Budson, 2007). Understanding this phenomenon could help to conceptualise how memory breaks down in healthy ageing, and in the earliest stages of amnestic Mild Cognitive Impairment (aMCI). For example, Ally et al. (2008) demonstrated that, despite similar levels of overall performance on a recognition task, healthy older adults showed greater picture superiority effects than younger adults. The memorial benefit of pictures was indeed evident in both the young and older groups, but the magnitude of this effect was greater in older adults, who only showed worse performance when responding to word stimuli. Interestingly, picture superiority also allows those with aMCI to show performance that is comparable to healthy older adult controls; despite exhibiting impaired performance overall, those with aMCI often show intact familiarity processes when pictures are utilised in recognition memory paradigms (and impaired familiarity when word stimuli are utilised; Embree, Budson, & Ally (2012); Ally et al. (2009a); Ally et al. (2009b); Wolk, Signoff, & DeKosky (2008); Algarabel et al. (2009); Anderson et al. (2008); Hudon et al. (2009); O'Connor & Ally (2010); Serra et al. (2010); Westerberg et al. (2006)].

The objective of the current programme of research is to better understand how different methodologies inform understandings about the underlying processes of recollection and familiarity. Across a number of experiments, the distinctiveness of to-be-remembered stimuli will be systematically examined to determine the level at which successful recognition is impacted, and which process(es) are most susceptible. The aim of the first experiment, outlined below, is to establish baseline PSE response patterns in a novel, modified RK paradigm. In a 2x3 mixed factorial design, a within-subjects variable of stimulus type (words / simple pictures) will be used to determine whether the magnitude of picture superiority effects (PSEs) is mediated by the particular response options available at test (between-subjects variable of response option: RFG, RFBG, RF-Ratings). In each condition, the labels *Recollection/Familiarity* will be used in place of the

standard *Remember/Know*, in an effort to reduce the impact of colloquial understandings on the current experimental definitions. To avoid guesses biasing estimations of familiarity (Belleville et al., 2011; Eldridge et al., 2002; Larsson et al., 2006; Tunney & Fernie, 2007; Williams, 2019), participants in all response-option conditions are also given the option to report that they are merely *Guessing* that an item is old. At test, subjects will be presented with either i) three response options (RFG); ii) four response options (RFBG; where a *Both* response option allows subjects to report the co-occurrence of R and F; iii) separate 0-5 rating scales for R and F (where subjects could report either process occurring alone, both processes occurring conjointly, or that they are guessing by providing a '0' rating on both scales). To establish whether a PSE is evident in the current paradigm, d' (d-prime) scores will be calculated for each participant. d' is a signal detection statistic, calculated by taking the standardised difference between the signal (i.e. correct hits) and signal+noise (i.e. false alarms); in other words, d' offers a representation of global recognition performance and participants' ability to distinguish target items from lures (Wixted, 2014). Higher d' scores demonstrate better overall performance on the memory task.

Based on the discussed research, the following results are hypothesised:

1. **Overall PSE:** a PSE will be evident within the current paradigm, manifesting as:

- i) higher overall d' scores for pictures compared to words;
- ii) higher proportion of correct hits;
- iii) lower proportion of false alarms;
- iv) better overall recognition.

2. **PSE in rates of Recollection and Familiarity:**

- i) pictures will produce a higher proportion of R hits and a lower proportion of R FAs than words.
- ii) words will produce a higher proportion of F hits and a higher proportion of F FAs than pictures.

3. **PSE and the availability of different response options:**

- i) comparable PSEs will be evident in each of the response option conditions (RFG, RFBG, RF-Ratings).
- ii) the availability of different response options will affect whether a PSE manifests as increased recollection, increased familiarity, or both (RFBG, RF-Ratings).

Experiment 1: Establishing PSEs in novel Remember/Know paradigm

Method

Participants A total of 186 subjects completed the online experiment ($M = 26.7$ years ($SD = 10.36$ years; see Table 1 for a comprehensive breakdown of the sample). The current sample was primarily comprised of participants sourced from voluntary participation websites such as Prolific Academic¹ (52.15%) (where they received payment at the rate of £5/hr) and via the in-school research participation system² (where they received course participation credits; 41.4%). A small number of participants were also recruited from social media and other online sources (Facebook: 3.76%; Call For Participants: 1.61%; Reddit: 0.54%; unspecified: 0.54%). To meet our YA requirements, all participants were required to be between 18-59 years of age (actual range: 18-59). As our experiment involved English word stimuli, we also asked subjects whether English was their first language; the vast majority (93.01%) reported that English was indeed their first language.

Table 1: Gender and age (SD) of the current sample.

Gender	N	Age	
Female	122	26.02	(10.04)
Male	60	28.10	(10.98)
Non-binary	2	19.50	(2.12)
Unspecified	2	39.00	(0)
Total	186	26.70	(10.36)

Materials Pictures of innocuous, everyday objects (e.g. clock, rabbit, shoe) and their written-word names were sourced from Rossion & Pourtois (2004). The picture stimuli consisted of

¹<https://www.prolific.co/>

²<https://keelepsychology.sona-systems.com/>

greyscale line-drawn illustrations (containing shaded surface details), while word stimuli were simply the written-word names of each object presented in a clear Sans-serif typeface. A total of 136 unique items were randomly selected for use in the current experiment, from a pool consisting of: i) items with a written name between 4 and 7 letters; ii) items that would conjure the same intended concept in our UK-based sample (e.g. “ladder” should be universally understood across English-speaking cultures, whereas “wagon” or “pants” can be interpreted differently); iii) items that were not unknown, or uncommon, for our sample (e.g. Americanisms such as “wrench”); and iv) non-specific concepts such as “bird” (since the pool of items already contained specific exemplars of birds, such as “peacock” and “penguin”). As the current experiment involved memorising word stimuli, a single item (“glass”) was also removed as it shared too many letters with another item (“glasses”). Selected items were split into four separate lists for counterbalancing purposes; using the normative data provided by Rossion & Pourtois (2004), each list was balanced based on the length of the written name, as well as scores of naming accuracy, familiarity, visual complexity, and mental imagery agreement. A series of independent samples t-tests confirmed that no list was significantly different from another on any of the aforementioned criteria.

The picture stimuli utilised in the current study were created in Photoshop CC (20.0.04 Release), by importing the greyscale, surface-shaded, line-drawings onto a plain 250x250px white canvas. Written word stimuli were created using the Calibri sans-serif typeface on the same size canvas (see Figure 1 for example stimuli). All items were exported as .pngs files for presentation by the online survey platform.

bottle	ladder	orange	shirt
			

Figure 1: Example word and picture stimuli from the current study.

Design The current study utilised a mixed design, with a 2-level within-subjects factor of stimuli format (words, drawings), and a 3-level between-subjects factor of response option (RFG, RFBG, RF-Ratings). Subjects completed two study blocks - one consisting only of word stimuli, the other consisting only of picture stimuli - before completing a single mixed format recognition test, where previously studied word and picture items were randomly shown among new, unseen items. Subjects passed through 2 levels of blocked randomization during the experiment (equally sized, predetermined blocks). First, subjects were randomly allocated into one of two study block orders, which determined the order in which they were presented with the picture and word blocks at study. Second, subjects were assigned into one of three possible recognition tests (identical aside from the response options available when categorising recognition experiences): 1) RFG: “Recollection”, “Familiarity”, “Guessing”; 2) RFBG: “Recollection”, “Familiarity”, “Guessing”, “Both”, or 3) RF-Ratings: two independent 0-5 rating scales to separately report the contribution of Recollection and Familiarity. These randomisation processes were completed automatically by the experiment software using balanced methods.

Procedure Data collection was conducted via the online survey platform Qualtrics³. Subjects initially completed an encoding block, where target words and pictures were randomly presented one-at-a-time on-screen. To ensure attention was directed to the presented stimuli, participants

³<https://www.qualtrics.com/uk/>

were required to respond to a simple encoding question toward each item at study: “Is this a picture or a word?”. This question allowed for the assessment of performance during the study block (to determine whether participants were concentrating at study), whilst also avoiding potential levels-of-processing effects that can accompany deeper encoding judgements (e.g. pleasantness ratings). The encoding phase was followed by a short distractor task comprised of 20 multiplication sums. Finally, subjects completed the recognition task, where they were again randomly presented with word and picture items one-at-a-time on-screen, and were required to respond *Old/New* depending on whether they recognised the item or not. *Old* responses were succeeded by a follow-up screen whereby participants were asked to report their recognition experience for the current item; the response options available during this follow-up response page differed between participants, with random allocation into either the RFG, RFBG, or RF-Ratings response option conditions. Recollection and Familiarity were defined identically across conditions, and the only deviations in instructions were: i) to define the additional “Both” response option in the RFBG condition; and ii) explain how certain responses should be reported in the RF-Ratings condition (i.e. subjects could still report a “Guess” in this condition by providing a 0-rating on both of the scales).

Data processing Measured variables included the total number of hits and FAs, and the total number of hits and FAs assigned to each of the available response options (RFG, RFBG, and RF Ratings). In order to create a common dependant variable, proportions were calculated from these variables in slightly different ways depending on the response option group. In the RFG-judgement group, simple proportions were created from the total number of R responses and the total number of F responses. In the RFBG condition, similar proportions were calculated by separately adding the proportion of Both responses to the proportion of R and proportion of F responses. In the RF-Ratings group, proportions of R and F were calculated based on the number of responses scoring +>3; a response was classified R when subjects rated between 3-5 on the “Recollection” scale (regardless of the Familiarity rating), and a response was classified F when subjects rated between 3-5 on the “Familiarity” scale (regardless of the Recollection rating).

The scales therefore allowed for pure R responses ($R=3-5 + F=0-2$), pure F responses ($F=3-5 + R=0-2$), both responses ($R=3-5 + F=3-5$) and Guessing responses ($R=0 + F=0$). Additional DVs included: i) d' (d-prime, a signal detection measure of sensitivity); ii) c -value (a measure of response bias); iii) overall accuracy (hits / (hits + FAs)); iv) reaction times for all responses.

All analyses were conducted using R (R Core Team, 2020). d' (sensitivity) and c (bias) scores were calculated using the 'psycho' package (v0.5.0; Makowski, 2018). d' scores were calculated via: z-scores for correct hits minus z-scores for false alarms (Hautus, 1995 adjustments for extreme values were applied). c scores were calculated

A series of exclusion criteria were defined before analysis. First, subjects were to be excluded from analysis if they showed poor performance during the encoding task; the relative ease of reporting whether each item was shown as a word or picture prompted a performance cut off of 90% accuracy. This would allow for some accidental clicks, though subjects scoring less than 90% were to be excluded on the assumption they did not dedicate their full attention to the task. Second, subjects would be considered outliers (and thus excluded from analysis) if they presented extreme z-scores of +/- 3 for total hits, total FAs, or overall recognition (hits minus FAs). However, no subjects were found to meet any of these criteria.

Results

Picture superiority To establish baseline picture superiority effects in the current paradigm, and assess whether there were any interactions with the availability of different response options at test, a series of 2 (stimuli format: words, pictures) \times 3 (response option condition: RFG-judgements, RFBG-judgements, RF-ratings) mixed ANOVAs were conducted on a number of outcome variables. Namely, the signal detection measures of d' (sensitivity) and c (decision criterion), as well as the proportion of overall hits, false alarms (FAs), and overall recognition (hits - FAs) [see Table 2]. Significant main effects and interaction effects were followed-up with Bonferroni-adjusted pairwise comparisons.

Table 2: Mean d' (sensitivity), c (decision criterion), proportion of hits, FAs, and overall recognition by stimuli-format and response-option condition.

	d'	c	Hits	FAs	Overall recognition
Stimuli-format					
Words	0.86	0.53	0.47	0.21	0.27
Pictures	1.62	0.48	0.62	0.12	0.50
Response-option					
RFG	2.71	0.67	0.62	0.19	0.43
RFBG	2.41	1.00	0.54	0.16	0.38
RF Ratings	2.32	1.33	0.48	0.14	0.34

The ANOVA on d' scores demonstrated a significant main effect of stimuli-format, $F(1, 183) = 295.80$, $MSE = 0.18$, $p < .001$, $\hat{\eta}_G^2 = .223$; a PSE was evident, with pictures ($M = 1.62$) producing significantly better discrimination between hits and FAs than words ($M = 0.86$), $t(183) = -17.20$, $p < .001$. The ANOVAs on the proportion of hits, FAs, and overall recognition also produced findings consistent with a PSE. For hits, there was a significant main effect of stimuli-format, $F(1, 183) = 131.77$, $MSE = 0.01$, $p < .001$, $\hat{\eta}_G^2 = .092$, with pictures ($M = 0.62$) showing a higher number of overall hits compared to words ($M = 0.47$), $t(183) = -11.48$, $p < .001$. Similarly, the ANOVA on the proportion of FAs showed a significant main effect of stimuli-format $F(1, 183) = 61.18$, $MSE = 0.01$, $p < .001$, $\hat{\eta}_G^2 = .084$, with words ($M = 0.21$) producing more FAs than pictures ($M = 0.12$), $t(183) = 7.82$, $p < .001$. Overall recognition performance (a measure that takes into account both hits and FAs) offered further support for a PSE in the current paradigm; a significant main effect of stimuli-format $F(1, 183) = 409.20$, $MSE = 0.01$, $p < .001$, $\hat{\eta}_G^2 = .236$ showed pictures ($M = 0.50$) produced better overall recognition on the task compared to words ($M = 0.27$), $t(183) = -20.23$, $p < .001$. Finally, c scores showed no significant main effect of stimuli-format, $F(1, 183) = 2.31$, $MSE = 0.11$, $p = .130$, $\hat{\eta}_G^2 = .002$, suggesting response biases were similarly conservative between pictures and words. No interaction effects were found between stimuli format and response option for any of the variables.

Taken together, the findings demonstrate a replication of the PSE in the current memory paradigm, and suggest stimuli format plays a key role in memorability that is independent from the particular response options available to participants. The current findings support the hypotheses of a PSE manifesting as i) higher overall d' scores for pictures compared to words, ii) a higher proportion of correct hits, iii) lower proportion of false alarms, and iv) better overall recognition.

PSE in rates of Recollection and Familiarity: To determine the impact of stimuli format on the rates of R and F, additional 2 (words, pictures) x 3 (RFG-judgements, RFBG-judgements, RF-ratings) mixed ANOVAs were conducted on the mean proportion of hits (see Figure 2) and FAs (see Figure 3) assigned R, F, and G.

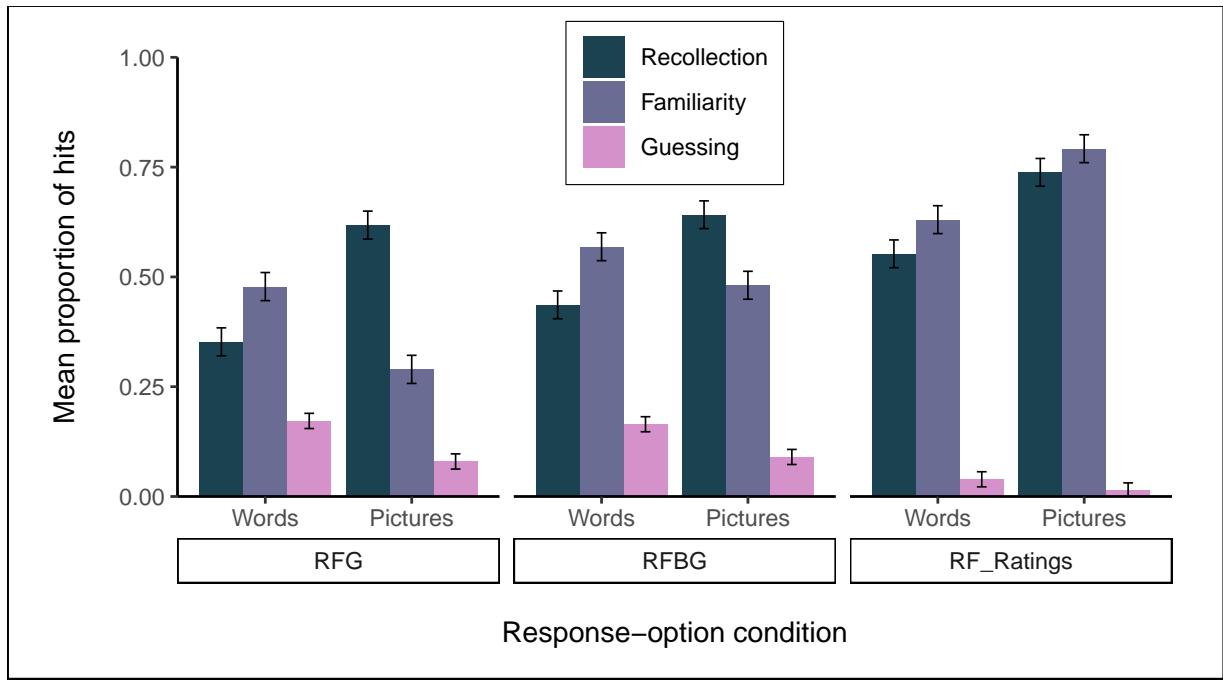


Figure 2: Proportion of hits assigned *Recollection*, *Familiarity*, and *Guessing*, by stimuli-format and response-option condition.

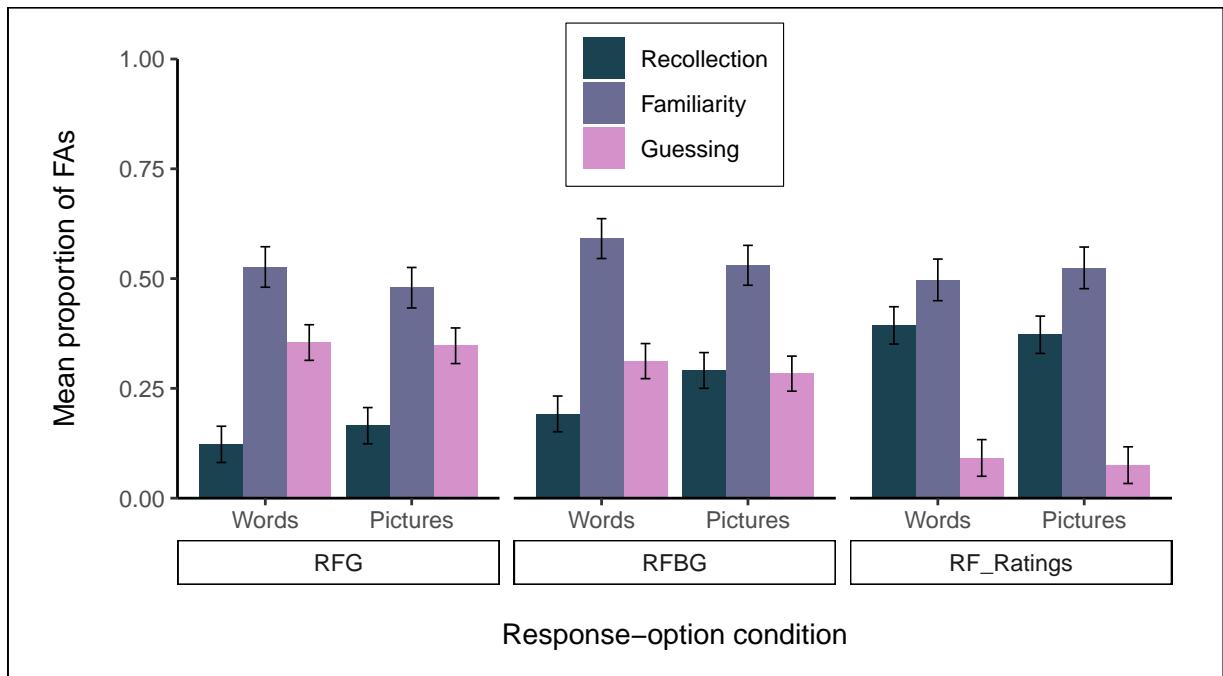


Figure 3: Proportion of FAs assigned *Recollection*, *Familiarity*, and *Guessing*, by stimuli-format and response-option condition.

Recollection: For R hits, there was a significant main effect of stimuli-format, $F(1, 178) = 158.42$, $MSE = 0.03$, $p < .001$, $\hat{\eta}_G^2 = .167$; pictures ($M = 0.67$) showed a higher proportion of Recollected hits than words ($M = 0.45$). For R FAs, there was no significant main effect of stimuli-format, $F(1, 136) = 2.78$, $MSE = 0.04$, $p = .098$, $\hat{\eta}_G^2 = .005$. There were no significant interaction effects between stimuli format and response option condition for R hits or FAs.

Familiarity: For F hits, there was a significant interaction between stimuli format and response option, $F(2, 178) = 34.42$, $MSE = 0.03$, $p < .001$, $\hat{\eta}_G^2 = .083$ (see Figure 4). Words resulted in more Familiarity hits than pictures in both the RFG group (words: $M = 0.48$; pictures: $M = 0.29$, $t(178) = 6.07$, $p < .001$) and RFBG group (words: $M = 0.57$; pictures: $M = 0.48$, $t(178) = 2.87$, $p = .005$). Conversely, pictures ($M = 0.79$) resulted in more Familiarity hits than words ($M = 0.63$) in the RF-Ratings group, $t(178) = -5.29$, $p < .001$. The ANOVA on Familiarity FAs did not yield any significant results; with no significant main effect of stimuli-format, $F(1, 136) = 1.12$, $MSE = 0.04$, $p = .292$, $\hat{\eta}_G^2 = .002$ or significant interaction effects, $F(2, 136) = 1.12$, $MSE = 0.04$, $p = .331$, $\hat{\eta}_G^2 = .004$.



Figure 4: Interaction plot between stimuli-format and response-option condition for the mean proportion of hits assigned *Familiarity*.

Guessing: The ANOVA on Guessing hits also showed a significant interaction between stimuli format and response option, $F(2, 178) = 4.17$, $MSE = 0.01$, $p = .017$, $\hat{\eta}_G^2 = .011$, (see Figure 5). Words resulted in more Guessing hits than pictures in both the RFG group (words: $M = 0.17$; pictures: $M = 0.08$), $t(178) = 5.38$, $p < .001$; and the RFBG group (words: $M = 0.16$; pictures: $M = 0.09$), $t(178) = 4.42$, $p < .001$. There was no difference in the number of Guessing hits between words $M = 0.04$ and pictures $M = 0.01$ in the RF-Ratings group, $t(178) = 1.51$, $p = .133$. For Guessing FAs, there was no significant main effect of stimuli-format or significant interaction effects.



Figure 5: Interaction plot between stimuli-format and response-option condition for the mean proportion of hits assigned ‘Guessing’.

Such findings mostly support the proposed hypotheses. Pictures indeed produced a higher proportion of R hits in comparison to words, though no picture superiority was evident in the number of R FAs. This suggests that, despite words showing a decreased level of memorability compared to pictures, they do not elicit high certainty false recognition at any higher rate. Words also produced a higher proportion of F hits compared to pictures, as predicted, but only in the RFG and RFBG conditions. It is unclear why the same pattern was not evident in the RF-Ratings group, aside from the possibility that participants avoided the more complex ratings screen (and instead more often chose *New*, inaccurately), unless they had a high certainty of their recognition (as evidenced for R hits). Again, there was no evidence of picture superiority in regard to the

number of F FAs. The hypotheses put forward for G responses were again mostly supported; there were more guesses made toward words than pictures, however, this again only applied to the RFG and RFBG conditions. Such findings align with the possible explanation outlined above, whereby participants avoided having to provide two separate ratings unless they were very certain they recognised the item.

PSE and the availability of different response options: To determine whether the availability of different options had an impact on picture superiority within the current paradigm, the main effects of response option were also examined in the aforementioned ANOVAs. Discriminability (d') between hits and FAs showed no significant main effect of response option, supporting the hypothesis that a PSE would be evident in each of the response option conditions. Response bias (c) scores did show a main effect of response option though, $F(2, 183) = 6.44$, $MSE = 0.51$, $p = .002$, $\hat{\eta}_G^2 = .054$; those in the RF-Ratings condition ($M = 0.67$) showed higher c-scores (and thus a more conservative response bias) than those in the RFG condition ($M = 0.34$), $t(183) = -3.59$, $p = .001$. This indicates subjects were less likely to respond *Old* when they were required to provide more detailed follow-up recognition judgements (i.e., using separate 0-5 scales for R and F), compared to simply selecting one of three options (R,F, or G). This again supports the notion of avoidance from participants when they were required to provide two separate ratings. The ANOVAs on the proportion of hits, FAs, and overall recognition also mostly supported our hypothesis that a PSE would be evident in each of the response option conditions. While the ANOVAs on the proportion of FAs and overall recognition showed no significant main effects of response option, there was a significant main effect for the proportion of hits, $F(2, 183) = 6.46$, $MSE = 0.09$, $p = .002$, $\hat{\eta}_G^2 = .057$, with the RFG group ($M = 0.62$) showing more hits than the RF-Ratings group ($M = 0.48$), $t(183) = 3.60$, $p = .001$. This unexpected result may also reflect the conservative response bias of those in the RF-Ratings group, whereby fewer total *Old* responses necessitates fewer hits as a result. However, a lack of differences between groups for FAs and overall recognition indicates a comparable PSE across each of the response option groups.

With regard to rates of R and F, there was a significant main effect of response-option for R hits, $F(2, 178) = 8.55$, $MSE = 0.09$, $p < .001$, $\hat{\eta}_G^2 = .069$; the RF-Ratings group ($M = 0.65$) showed significantly more R hits compared to both the RFG-group ($M = 0.49$), , and the RFBG-group ($M = 0.54$), . R FA's also exhibited an identical pattern; a significant main effect of response-option, $F(2, 136) = 10.70$, $MSE = 0.12$, $p < .001$, $\hat{\eta}_G^2 = .106$, showed that those in the RF-Ratings group ($M = 0.38$) produced more Recollection FAs than both the RFG-group ($M = 0.14$), , and RFBG-group ($M = 0.24$),). These findings align with those above to suggest a more conservative response bias in the RF-Ratings group; participants were more likely to respond *Old* in the RF-Ratings condition when they experienced high certainty in their recognition - regardless of whether or not that recognition was accurate or false.

The significant interaction for F hits showed that word stimuli produced significantly more F hits in the RF-Ratings group ($M = 0.63$) compared to the RFG group ($M = 0.48$), $t(276.78) = -3.37$, $p = .002$. An identical pattern was also observed for pictures between the RF-Ratings group ($M = 0.79$) and RFG group ($M = 0.29$), $t(276.78) = -11.13$, $p < .001$, however, F hits in the RF-Ratings group were also higher than the RFBG group ($M = 0.48$) when the stimuli were pictures, $t(276.78) = -6.94$, $p < .001$. The RFBG group ($M = 0.48$) also showed a significantly higher number of F hits compared to the RFG group ($M = 0.29$), $t(276.78) = -4.24$, $p < .001$. There was no significant main effect of response-option for F FAs. Such findings suggest that the proportion of F responses increases when participants are given the option to report *Both* - either explicitly, or from rating high on both rating scales.

Finally, the significant interaction for G hits demonstrated that word stimuli produced significantly fewer G hits in the RF-Ratings group ($M = 0.04$) compared to both the RFG group ($M = 0.17$), $t(281.42) = 5.44$, $p < .001$, and the RFBG group ($M = 0.16$), $t(281.42) = 5.18$, $p < .001$. Likewise, pictures also produced significantly fewer G hits in the RF-Ratings group ($M = 0.01$) compared to both the RFG group ($M = 0.08$), $t(281.42) = 2.70$, $p = .020$.

and RFBG group ($M = 0.09$), $t(281.42) = 3.15$, $p = .005$. For G FAs, a significant main effect of response-option, $F(2, 136) = 15.69$, $MSE = 0.11$, $p < .001$, $\eta^2_G = .144$ revealed that the RF-ratings group ($M = 0.08$) again showed significantly fewer G FAs than both the RFG ($M = 0.35$), , and RFBG groups ($M = 0.30$), . These findings again align with the previous results that suggest those in the RF-Ratings group responded conservatively overall, and were less likely to respond *Old* than the other groups unless they felt a high level of certainty.

Discussion

The aim of the current study was to establish baseline PSE response patterns in a novel, modified RK paradigm. Substituting the classic *Remember / Know* labels for *Recollection / Familiarity*, recognition for words and pictures was tested across three separate response option conditions (RFG, RFBG, RF-Ratings). Analysis of the behavioural data demonstrated a clear Picture Superiority Effect (PSE) in the current paradigm, with picture stimuli showing better discrimination, a higher number of overall hits, lower number of FAs, and better overall recognition performance than words. Taken together, these findings are consistent with the notion that pictures offer an enhanced memorability in comparison to words. When word stimuli were correctly identified, they were not recognised in the the same context-rich nature as pictures, evidenced by a higher proportion of F responses. The current findings also align with those from previous studies, with pictures showing enhanced recollection (Curran & Doyle, 2011; Rajaram, 1996a) and words showing enhanced familiarity (Ally & Budson, 2007). While most of the proposed hypotheses were supported, there were some unexpected results. Stimuli format had no effect on the obtained proportions of FAs; regardless of whether FAs were assigned R or F, there was no evidence of picture superiority. This finding does not refute the notion of a PSE in the current paradigm - the memorial advantage of pictures over words is evident, but it instead indicates that stimuli without this advantage (i.e. words) may produce more misses, but not increased levels of false recognition.

Many of the unexpected results are centred around the RF-Ratings response option condition. Word stimuli were hypothesised to produce more Familiarity and Guessing hits than pictures,

since it was expected that they would not be recognised in the same context-rich nature as pictures. This result was indeed obtained in both the RFG and RFBG response-option conditions, however, the RF-Ratings did not produce the same finding (no difference between stimuli formats was observed). Similarly, while the RFG and RFBG conditions showed comparable proportions of hits and levels of response bias (*c* scores), the RF-Ratings group again produced different findings, showing significantly fewer hits and significantly higher *c* scores (and thus a more conservative response bias) compared to the RFG group. As the proportion of hits and mean *c* scores were not significantly different between the RF-Ratings and RFBG response option groups, it indicates that these results may be attributable to participants having the ability to report that they experience *Both* recollection and familiarity processes conjointly. However, as performance differences were most notable in the RF-Ratings condition, it suggests these findings are attributable to the increased task complexity from RFG to RFBG, and RFBG to RF-Ratings. The option to report *Both* may be confusing to participants, especially to those who struggle to understand the distinction between recollection and familiarity to begin with (Geraci et al., 2009; Rubin & Umanath, 2015; Williams & Moulin, 2014). Providing the *Both* option in the form of two scales may exacerbate this confusion further, and thus lead to results that are significantly different from the condition with the least complexity (RFG). Such a hypothesis is supported by a number of other findings. First, the more conservative response bias exhibited by those in the RF-Ratings group demonstrates how subjects were less likely to respond *Old* when they were required to provide more detailed follow-up recognition judgements (i.e., using separate 0-5 scales for R and F), compared to simply selecting one of three options (R,F, or G). Second, despite *Guessing* responses being permissible in any of the response-option groups, participants were significantly less likely to report a guess when two independent ratings were required - a finding evident from the reduced number of *Guessing* hits and FAs compared to the other response option conditions. Third, the RF-Ratings group showed significantly more R hits and R FAs compared to both the RFG-group and the RFBG-groups, indicating participants were more likely to respond *Old* in the RF-Ratings condition when they experienced high certainty in their recognition - regardless of whether or not that recognition was accurate or false. Taken

together, these findings all support the notion of a certain level of avoidance from participants when they were required to provide more detailed reports of their recognition.

Establishing baseline PSEs in the current paradigm is important for allowing further experimental manipulations in the experiments that follow. While the independent ratings paradigm proposed by Higham & Vokey (2004) is undoubtedly useful in the discussion around the most effective methods of measuring recollection and familiarity, it does not suit the needs of current programme of research going forward, where comparisons between different stimuli formats is the primary concern. In the current experiment, picture stimuli consisted of simple greyscale illustrations (Rossion & Pourtois, 2004), though the extent to which stimuli of increasing levels of detail impacts recognition is unclear. The distinctiveness of to-be-remembered stimuli will be systematically compared across a number of experiments, following the conception of a new set of detailed realistic photograph stimuli. Following the unique results obtained from the RF-Ratings group in the current experiment, it is likely that this condition would exhibit further differences if included in the proposed experiments, which may become difficult to interpret when further stimuli formats are introduced. The current findings to suggest avoidant response patterns in the RF-Ratings group further highlight the unique results this condition might produce. Therefore, only the RFG and RFBG response option conditions will be taken forward into the proposed recognition experiments that focus on comparisons of stimuli distinctiveness.

#####-----

Chapter 3

The Picture Superiority Effect (PSE) is a highly robust and replicable phenomenon. In recognition memory paradigms, the PSE has been shown to manifest as both increased recollection and familiarity (Dewhurst & Conway, 1994; Rajaram, 1993, 1996b; Wagner, Gabrieli, & Verfaellie, 1997; Yonelinas, 2002). The effect is present in children, adolescents and healthy older adults (Whitehouse, Maybery, & Durkin, 2006), though perhaps more striking is the fact that patients with Alzheimer's disease or those presenting early isolated memory impairments, known as amnestic mild cognitive impairment (aMCI), also show memorial benefits toward pictures (Ally, 2012). This is supported by ERP studies demonstrating comparable enhancements to recollection-based ERP components between healthy older and aMCI groups when pictures, rather than words, are utilised (Ally et al., 2009a). There is debate within the literature attempting to characterise the nature of memory deficits in aMCI, whereby despite general agreement that recollection processes are impaired in such individuals, findings show great inconsistency with regard to familiarity (Algarabel et al., 2012; Belleville et al., 2011; Pitarque, 2016; Wolk, Dunfee, Dickerson, Aizenstein, & DeKosky, 2011; Wolk, Mancuso, Kliot, Arnold, & Dickerson, 2013). The PSE may have been largely overlooked as an area for further research in an effort to help settle this debate, despite recent reviews highlighting methodological differences across studies as the potential source of inconsistent findings (Koen & Yonelinas, 2014; Migo et al., 2012; Schoemaker et al., 2014). The level at which stimuli distinctiveness impacts successful recognition is currently unclear, and there is little consistency across studies with regard to what is considered a 'picture'.

Many experiments utilise illustrations for their picture stimuli (van der Meulen et al., 2012; Westerberg et al., 2013; Wolk et al., 2011), with a standardised set of items published by Snodgrass & Vanderwart (1980) among the most-used illustrated picture stimuli within the domain of memory research (Bermúdez-Margaretto, Beltrán, Cuetos, & Domínguez, 2018; Deason, Hussey, Flannery, & Ally, 2015; Hockley, 2008; Martins & Lloyd-Jones, 2006; McBride & Anne Dosher, 2002; Meade, Ahmad, & Fernandes, 2019; Schmitter-Edgecombe, Woo, & Greeley, 2009; van der Meulen et al., 2012; Wagner et al., 1997; Wammes, Meade, & Fernandes, 2016; Weldon, lii, &

Challis, 1989; Weldon & Roediger, 1987; Whitehouse et al., 2006). The set consists of 260 line drawings of common, everyday objects (in black ink), along with their written word counterpart (e.g. “shoe”). Items were selected on the basis of exemplifying a number of semantic categories, including animals, furniture, fruit, etc., and a range of normative data was collected for each item; indices of naming agreement, mental imagery agreement, visual complexity, and familiarity were all recorded for each drawing. The normative data for the Snodgrass & Vanderwart (1980) items has been continually revisited, with a number of studies gathering culturally-appropriate norms (e.g. in Spanish (Sanfeliu & Fernandez, 1996), Chinese (Yoon et al., 2004), and Russian (Tsaparina, Bonin, & Méot, 2011), and additional testing of the relationship between reaction time and naming agreement (Székely et al., 2003). There are multiple theories of object recognition; the recognition-by-components theory proposed by Biederman (1987) identifies shape as the most crucial factor for successful recognition, in which case, the object outlines found in the set by Snodgrass & Vanderwart (1980) should be more than sufficient for experimental cognitive research. Other theories, however, posit that surface details such as colour and texture are just as crucial in forming object representations (Tanaka, Weiskopf, & Williams, 2001; Tarr & Bühlhoff, 1998). The wide-ranging applicability of the Snodgrass & Vanderwart (1980) items throughout a number of cognitive disciplines has led to a more recent revision of the items by Rossion & Pourtois (2004). This revision consists of the exact same objects, digitally re-drawn to include surface textures and shading. Additionally, this set provides greyscale and colour versions for all items, as opposed to the greyscale-only items found in the Snodgrass & Vanderwart (1980) set (see Figure 6 for example items contained in the Snodgrass & Vanderwart (1980) and Rossion & Pourtois (2004) stimuli sets). The Rossion & Pourtois (2004) revision now appears to be favoured over the original Snodgrass & Vanderwart (1980) set among many cognitive researchers (Rollins & Riggins, 2018, p. @ensor2019b; Stenberg, 2006; Wolk et al., 2008), almost certainly attributable to the increased detail and ability to choose whether colour is a necessary condition.

Despite their widespread use, line drawings have been criticised for their relative simplicity and lack of realism (Viggiano, Vannucci, & Righi (2004)), with many researchers favouring the use of

photographs as experimental stimuli (Embree et al., 2012; Pitarque, 2016; Troyer et al., 2012; Troyer, Vandermorris, & Murphy, 2016; P. Wang et al., 2013). Photographs of faces are especially useful in research examining emotion and face recognition (Barba, 1997; Bowen, Fields, & Kensinger, 2019; Cui et al., 2016; Herzmann, Minor, & Curran, 2018), though a number of common-object photograph sets have also emerged as ecological alternatives to line-drawn items (Adlington, Laws, & Gale, 2009; Moreno-Martínez & Montoro, 2012; Viggiano et al., 2004). While the published sets of photographs are undoubtedly useful in a range of cognitive domains, they do not allow us to specifically examine stimuli format as a factor on its own, as the concepts depicted are unique to the set they derive from. In order to make such comparisons, and ensure any differences in performance (e.g. recognition memory ability) are indeed attributable to stimuli format, the objects depicted must be consistent across stimuli formats. The current study presents a new set of photographic stimuli that extend the set of words and drawings provided by Rossion & Pourtois (2004), wherein each of the concepts depicted has been carefully matched across formats. These new stimuli will be utilised throughout a number of planned recognition experiments that aim to systematically compare measures of recognition against different ‘levels’ of stimuli. The curation of a new set of photographs - carefully matched to other formats - allows investigation into whether picture superiority magnitudes are mediated by the format pictures are presented in. The inconsistent use of different formats across studies has previously made it difficult to reconcile effects obtained in response to drawings with those obtained in response to photographs - an inherent problem when concepts are not matched across format. Normative data for the new set of photographs is also presented, allowing others who also wish to use our photograph stimuli to filter items by measures of naming agreement, mental imagery agreement, familiarity, visual complexity, and colour diagnosticity.



Figure 7: Examples of matching pictures across Snodgrass & Vanderwart (1980), Rosson & Pourtois (2004), and photographs from the current study. Greyscale versions of the drawings and photographs are not presented in this example.

Experiment: Development of a new set of standardised photographic stimuli

Method

Participants A total of 377 subjects completed the online experiment (see Table 3 for a breakdown of the gender and age of the sample). This sample size provided 20 data points for each of the five response types, while also ensuring the experiment did not last too long for participants (approx 25-mins). Subjects were recruited from both voluntary participation websites such as Prolific Academic⁴ (where they received payment at the rate of £5/hr), and via the in-school research participation system⁵ (where they received course participation credits).

Table 3: Gender and age (*SD*) of the current sample.

Gender	N	Age	
Female	196	33.22	(11.28)
Male	171	33.15	(10.3)
Non-binary	2	23.50	(-)
Unspecified	5	29.40	(6.11)
Total	377	NA	NA

To meet our YA requirements, all participants were required to be aged between 18-59 years (actual obtained range: 18-59 years). As our experiment involved typing the English labels for a range of image stimuli, subjects were also asked whether English was their first language; all but one participant indicated that English was indeed their first language (99.2%).

Materials A pool of 136 line drawings (Rossion & Pourtois, 2004) - depicting common, everyday objects - were brought forward from the previous experiment. These items (along with their written-word labels) would form two of the unique stimuli formats that would be used in future recognition experiments (words and drawings). In this study, the drawings from Rossion & Pourtois (2004) were simply used as a reference in the photograph matching process. Corresponding photographs were obtained online with the aim of depicting the everyday objects in a similar manner to the drawings. The inherent subjectivity involved in this process may have led to images that were not a reliable 'match' to the concepts they were selected to depict (for

⁴<https://www.prolific.co/>

⁵<https://keelepsychology.sona-systems.com/>

example, the photograph chosen to depict the concept “bottle” may inadvertently provoke the majority of participants to give the label “wine”, thus indicating that this particular photograph fails to accurately depict the intended concept). To address this issue, and ensure all photographs more objectively depict the same concepts as the line drawings, three different photograph variations were found for each everyday object, with the aim of taking the best ‘match’ forward. An emphasis was placed on variety across these variations, with the aim of obtaining at least one photograph that very closely resembled the line-drawn depiction, and another offering a more modern depiction. Some items were substituted due to unique restrictions that meant they could not easily be translated into photographic format (for example, the shapes “arrow” and “star” can not be represented similarly as photographs). Photo stimuli were obtained by searching open-source, copyright-free image websites (e.g. Unsplash⁶; Pexels⁷) for photographs that depicted the same everyday objects as the line drawings (see Appendix B for the full list of image references).

The matching process produced a total of 408 unique photographs. All were imported into Adobe Photoshop (20.0.04 Release), where the background was removed to isolate the object of interest from other potentially distracting visual details. This was completed manually using the magnetic lasso and polygonal lasso tools (edges were either feathered by 1px or left unfeathered). The orientation of isolated objects was adjusted to ensure they matched as closely as possible with their line-drawn counterpart (e.g. all photograph variations of the item ‘boot’ were adjusted so the toe was facing left and the heel facing right, as in the line drawing); this was often achieved by flipping or mirroring the object to ‘correct’ the direction.

Despite isolating objects from their background, a small number of photographs still contained irrelevant and potentially distracting details. For example, in one photograph variation of the item ‘piano’, there was a sign on the object that may have impacted how the item was named or rated. Such details were removed as best as possible using the clone stamp and content-aware fill tools. Any obvious text (e.g. brand names) and numbers were also removed from photographs using the same method (see Figure 8). The primary aim of the current study was to obtain

⁶<https://unsplash.com/>

⁷<https://www.pexels.com/>

photographs that could be clearly distinguished as a unique stimuli format among words and line drawings; it is conceivable that combining these formats (i.e. inadvertently including photographs that also contain written words) might affect recognition performance in ways that are not directly comparable to items defined only by a single category. Any text in our photographs was therefore removed, apart from a couple of exceptions whereby such details happened to be integral to the depiction of the object (e.g. the numbers found on a ruler or clock).

All photographs were exported from Photoshop in “.png” format in both their original colour and in greyscale (by setting saturation levels to 0). Final edits were completed in Adobe Lightroom (Classic, 8.2 Release): exposure (brightness) adjustments were made on images that appeared too light or too dark; highlights were decreased if some areas were too bright compared to the rest of the photograph; shadows were raised if some areas were too dark compared to the rest of the photograph; noise reduction was applied to some items after isolating the subject had inadvertently made unwanted noise/grain more visible. The changes made to each image were systematically applied to both the colour and greyscale versions (e.g. if one variation of “shoe” had an exposure increase of .010 for the colour version, the greyscale version also received an exposure increase of .010). Some colour-specific adjustments were made to the colour photographs only, however; common photo artefacts such as chromatic aberration (purple fringing) were corrected, along with white balance normalisation. Finally, all photographs were placed on a 600x600 pixel white background, and made to fill this frame as much as possible (i.e. some items were restrained by height, whilst others were restrained by width).



Figure 8: Examples of background and text removal in photograph items.

Design This was a descriptive study; a mix of qualitative and quantitative data were gathered. Across three blocks, all participants provided five types of response toward photograph stimuli: i) Naming; ii) Familiarity; iii) Visual Complexity, iv) Colour Diagnosticity; and v) Mental Imagery Agreement. Excluding the Naming task (consisting of a typed single-word answer), all responses

were provided on a 5-point ordinal scale. Within participants, the maximum number of response type provided for any one item was two; Naming and Familiarity responses were paired in one block, Visual Complexity and Colour Diagnosticity responses were paired in another, and Mental Imagery Agreement responses were always presented in a separate block. The order of these three blocks was counterbalanced across participants. Toward each individual photograph, participants made only one or two types of response before moving on to the next item, and the same items were not repeated to participants. For each photograph, the five types of required data were obtained by counterbalancing between participants (e.g. for the first variation of the “cat” photograph, the Naming and Familiarity data was obtained from one participant, the Visual Complexity and Colour Diagnosticity data was obtained from another, and the Mental Imagery Agreement data was obtained from another).

Procedure Data collection was conducted via two online platforms; i) Qualtrics⁸ - a survey platform that allowed for straightforward collection of consent, demographics, and computer compatibility data, and ii) Pavlovia⁹ - an open-source experiment hosting platform for studies programmed in Javascript (Peirce et al., 2019).

In the Naming and Familiarity block, participants were first asked “What is the name of the item depicted?”. Subjects were instructed to name each photograph as briefly and unambiguously as possible, with one name only, and respond by typing their answer into the response box. If they did not know the name of an item, or had a tip-of-the-tongue experience, participants were instructed to type “no” for their answer (the term “don’t know” was avoided so as not to encourage subjects to deviate from single-word responses, as instructed). Following the naming judgement, with the same photograph still present on-screen, participants were next asked “How familiar is the item depicted?”. Subjects were instructed to judge each photo according to how usual or unusual the item was in their realm of experience; specifically, familiarity was defined as “the degree to which you come in contact with, or think about, the concept”, and encouraged participants to rate the concept itself rather than the particular way it was currently shown. Participants

⁸<https://www.qualtrics.com/uk/>

⁹<https://pavlovia.org/>

selected one value from the 5-point scale, ranging from very unfamiliar (1) to very familiar (5), and were encouraged to use the full range of the scale throughout the set of photographs.

In the Visual Complexity and Colour Diagnosticity block, participants were first instructed to respond to the question “How visually complex is this picture?” using a 5-point scale that ranged from “very simple” (1) to “very complex” (5). Complexity was defined to subjects as “the amount of detail in the picture”; in contrast to the familiarity ratings, participants were encouraged here to rate the complexity of the picture itself, rather than the real-life item. If the photograph shown was greyscale, subjects would simply move on to the next item. If the item shown was in colour, however, participants were also required to make a colour diagnosticity judgement. This concept was defined as “how typical / normal the colour of the item is”, instructing subjects to rate on a 5-point scale ranging from “Not at all diagnostic (i.e. this item could be in any other colour equally well)” (1) to ”Highly diagnostic (i.e. this item appears only in this colour in real life). Participants were instructed to utilise the full range of options on the scale when making visual complexity and colour diagnosticity judgements. After making these ratings, a fixation cross was presented during a 1s interstimulus interval.

Due to the slight change in procedure and increased task complexity, Mental Imagery Agreement ratings were always acquired in an individual block (i.e. not alongside any other response types). First, participants were presented with a written label for 3s (e.g. “cat”) and told to focus their attention on the word. Once the written word disappeared, a beep tone was played alongside the instruction “close your eyes and imagine this item” (subjects were encouraged to close their eyes and begin imagining the item as soon as they heard the tone, but the written instruction were included as a further prompt). After 3s a second beep tone sounded to alert subjects to open their eyes, where they were presented with a photograph of the item they had been instructed to imagine. On a 5-point scale, participants were asked to “rate the agreement between your mental image and the picture”, from “low agreement” (1) to “high agreement” (5). The degree of agreement was defined as “how similar your mental image of the item is to the picture shown”. A fixation cross was displayed for 1s before the next word item was shown.

All responses were self-paced; the timing was only controlled during the study/imagine section

of the Mental Imagery Agreement block.



Figure 9: Data collection procedure for Mental Imagery Agreement responses.

Data processing The naming responses for each photograph item were manually assessed for spelling and typing errors. Automatic spell checking software was avoided in an effort to avoid inadvertently introducing unique names that were not actually given by participants. The vast majority of errors were unambiguous and easy to correct (e.g. “anker” = “anchor”, “peguin” = “penguin”, “ssnowman” = “snowman”), or consisted of transforming plural words to singular (or vice versa, depending on the form of the intended label - e.g. “sock” to “socks”). Some responses were a little more ambiguous, and necessitated comparison to the photographs they were in response to for additional clarity (e.g. a photograph depicting a plug that would fit into North American electrical sockets was labelled as “usplug” - given the nature of our UK-based sample, it’s likely the subject was responding: “U.S. (i.e. United States) plug”).

There were instances where subjects provided a sensible and correctly spelled English word, but that were clearly typos when examined against the photograph they were in response to (e.g. “dock” for a photograph depicting a duck, “frock” for a frog, and “beer” for a “bear”, etc). The most ambiguous spelling error to correct was “bittle”, which was provided by more than one

participant and to more than one item; separate inspections of the photographs participants were responding to made this easy to correct though, with one participant clearly meaning to respond “bottle”, whilst the other meant to respond “beetle”. Though participants were instructed to only give a single label for each item, some multiple word responses were found (without spaces) during the spell checking process. On such occasions, a judgement was made regarding whether multiple words were retained, or whether the response could be shortened into a single word. A general rule was applied whereby if the other words provided additional information, they were retained (e.g. “maledeer” - presumably “male deer” - was kept as a two-word answer). Multiple word responses were generally shortened into a single word when the intended label for the item was clearly present, and no information was lost in the process (e.g. “haircomb” was shortened to the intended answer “comb”). It is noted that there was some inherent subjectivity in this process, though as such items were not common among straightforward responses, their overall effects are estimated to be negligible.

Finally, there were some responses that were changed to “no” as they were clearly intended to signify that the responder did not know the name of the item shown; the experiment instructed participants to type “no” in these instances, though the labels “none” and “idk” (common abbreviation for “I don’t know) were provided instead. There was also a single response that was manually changed to “no”, as the provided label was a single letter and thus entirely unclear what the intended answer should be (see Appendix A for full list of manipulations to naming responses). This process yielded data that could be used to determine which photograph variation best matched the intended concepts (e.g. 100% of participants labelled the object “bottle”, indicating a perfect match), and which did not (e.g. only 50% of participants labelled the item “bottle”, whilst the other 50% gave the label “wine”, indicating a poor match). Photographs showing poor agreement across participant-generated labels, or those where the majority of labels differed from the intended concept, could be replaced with the variation demonstrating the most accurate depiction.

Analysis preparation A number of variables were calculated prior to analysis. For familiarity, visual complexity, colour diagnosticity, and mental imagery agreement, mean ratings were calculated for each (see Appendix B). Mean reaction times (RTs) were also calculated for each photograph / response variable, including naming responses. For naming responses, accuracy was defined as the proportion of subjects reporting the correct/intended label for any given item (e.g. 80% of subjects correctly labelled a photograph of the moon as “moon”). Percentage agreement was also calculated (i.e. the proportion of subjects providing the most frequent name, regardless of whether it matched the correct/intended label) in order to compute H values for each item. The H statistic also reflects naming agreement, but it takes into account the total number of unique labels given for an item. This is especially useful for comparing similar items, as it captures information not provided by simple agreement proportions. For instance, if the first variation of the photo moon ('moon-1') demonstrated 90% naming agreement among subjects, and the second variation ('moon-2') also demonstrated 90% naming agreement, it would appear as if both versions offer the same level of agreement among participants. However, 'moon-1' may have received a total of 2 unique names (e.g. moon, planet), while 'moon-2' received a total of 4 unique names (e.g. moon, planet, earth, comet). H values utilise this useful information to determine which item shows the best naming agreement (in other words, the item with the least number of unique names). The original formula by Snodgrass & Vanderwart (1980) was used to calculate H values:

$$H = \sum_{i=1}^k p_i \log_2 \frac{1}{p_i},$$

A H value of 0 indicates perfect naming agreement (all subjects responded with the same label for that item). Items showing a H value of 1 signify two unique names were provided, with identical proportions (e.g. 10 subjects responded “moon” and 10 subjects responded “planet”). As the H value increases, overall naming agreement decreases.

Results

Summary statistics (mean and *SD*) for each of the measured variables are shown in Table 4. Data for the grey and colour photographs are presented alongside previously obtained normative values for a number of other stimuli formats (all obtained from Rossion & Pourtois (2004), who published revised norms for Snodgrass & Vanderwart (1980)'s (S&V) original line drawings, as well as their own re-drawn versions that contained shading and texture detail). The data from previous studies were not used in any statistical analyses. To examine whether the grey and colour photographs from the current study demonstrated any differences, a series of independent samples t-tests were run on each variable, as well as their corresponding reaction times (excluding scores of colour diagnosticity, which were obtained only in response to the colour items and thus cannot be compared). Mean (and *SD*) values for all x816 unique photograph items are presented in Appendix B.

Naming Naming accuracy was very high for all photographs ($M = 0.95$), indicating that overall, the selected items closely depicted the intended concepts. Compared with the other stimuli formats, there appears to be a steady increase in accuracy as items become more distinctive (see Table 4). Accuracy rates did not differ between the grey ($M = 0.94$) and colour ($M = 0.95$) versions of the photographs [$t(745.64) = -0.56, p = .576$].

H values were also low across all items ($M = 0.23$), showing that subjects generally agreed on how the items should be named. Similar to naming accuracy, naming agreement also appears to steadily increase as items become more distinctive (as indicated by decreasing *H* values - see Table 4. While Rossion & Pourtois (2004) observed significantly better naming agreement for their colour - rather than greyscale - items, this pattern did not reach significance with the current set of photographs; *H* values did not differ between the grey ($M = 0.24$) and colour ($M = 0.22$) photographs [$t(743.66) = 0.62, p = .537$].

A mean reaction time (RT) of (3.9s) was observed for naming responses. While this was of little interest on its own, and could not be compared to those obtained in response to the other stimuli formats as our methodology was slightly different (RTs were only recorded when subjects had

typed their response *and* clicked the mouse to signify they had finished), they were useful for marking comparisons between the grey and colour items (though no difference was observed [M grey = 4s, M colour = 3.8s, $t(651.86) = 1.57$, $p = .117$]). Overall, these analyses suggest that the current photographs closely resemble the drawings they were designed to match, with high levels of naming accuracy and agreement among subjects. The absence of any colour differences indicates there were no naming advantages when photographs were made even more distinctive through the addition of colour.

Table 4: Summary statistics for each of the measured variables. Mean values are presented in bold (SDs are shown in parentheses).

	Rossier & Pourtois (2004)			Current study	
	S&V lines	Grey shaded	Colour shaded	Grey photos	Colour photos
Naming accuracy	88.2 (17.1)	89.2 (17.2)	90.3 (16.9)	0.94 (0.08)	0.95 (0.08)
Naming agreement (H)	0.44 (0.56)	0.38 (0.52)	0.32 (0.46)	0.24 (0.33)	0.22 (0.31)
Mental imagery agreement	3.73 (0.48)	3.76 (0.55)	3.74 (0.63)	3.46 (0.56)	3.74 (0.65)
Familiarity	3.59 (0.94)	3.52 (1.01)	3.44 (1.01)	4.13 (0.56)	4.19 (0.54)
Visual complexity	2.76 (1.03)	2.88 (1.03)	2.7 (0.94)	2.87 (0.62)	3.16 (0.63)
Colour diagnosticity	-	-	-	-	3.22 (0.84)
	-	-	-	-	

Mental imagery agreement Scores of mental imagery agreement were moderate across all items ($M = 3.6$). While no colour differences were previously observed between stimuli formats, the grey ($M = 3.46$) photographs in the current study showed significantly lower mental imagery agreement scores than the colour ($M = 3.74$) items [$t(800.06) = -6.54$, $p < .001$].

Comparisons with previous normative data also highlight how the grey photographs exhibited uniquely poorer mental imagery agreement scores than any of the other stimuli formats (see Table 4). RTs between the grey ($M = 3.04$) and colour ($M = 2.81$) items did not significantly differ [$t(571.37) = 2.14, p = .033$].

Familiarity Familiarity scores were high overall ($M = 4.16$), and like previous findings, there was no difference between the grey ($M = 4.13$) and colour ($M = 4.19$) items [$t(813.19) = -1.63, p = .103$]. However, familiarity scores for the current set of photographs were higher than those obtained for any of the other stimuli formats, and while there previously appeared to be a decline in familiarity as stimuli become more distinctive (from line drawings, to grey shaded, to colour shaded), such a pattern was not evident with the current photographs (see Table 4). RTs between the grey ($M = 0.97$) and colour ($M = 0.98$) items did not significantly differ [$t(783.66) = -0.30, p = .762$].

Visual complexity Visual complexity ratings were moderate across all of the items ($M = 3.3$). Colour ($M = 3.16$) photographs showed significantly higher scores of visual complexity than grey ($M = 2.87$) photographs [$t(813.51) = -6.65, p < .001$]. This finding is further demonstrated when compared to the scores from the other stimuli formats (see Table 4); where grey photographs show comparable levels of visual complexity, the colour photographs show higher scores than all of the other formats. There was no significant difference between the RTs of grey ($M = 3.26$) and colour ($M = 3.35$) items [$t(754.08) = -1.21, p = .228$].

Selection of final items For each concept represented in the photographs, one variation (e.g. shoe-1, shoe-2, or shoe-3) was selected for inclusion in a final list of stimuli that would be taken forward into subsequent recognition experiments. The normative naming data was assessed to establish which version best matched the existing line-drawn depictions of the concepts (Rossion & Pourtois, 2004). Naming was favoured over all of the other variables as, if an item was found to primarily convey a different concept than was intended during the naming

task (e.g. if a photograph of the fruit ‘orange’ was labelled ‘grapefruit’ by the majority of subjects), then it could not be sufficiently compared to its line-drawn (and written-word) counterpart during recognition studies.

At least 20 unique naming responses were collected for each of the 816 photographs (408 grey items and 408 colour items). The proportion of ‘correct’ responses (i.e. names that were congruent with the intended concept) and the proportion of ‘don’t know’ responses were calculated for each item. Photographs were excluded if they:

1. received a high proportion of “don’t know” responses (20%; all of the photographs depicted common, everyday objects, and so if a number of subjects were unable to name the item, that particular photograph was considered to be a poor representation of the item);
2. were incorrectly named by the majority of subjects (i.e. if the proportion of correct responses equalled $\leq 50\%$, since it was essential for the photographs to depict the same concepts as those found in the line drawings and word stimuli);
3. had particularly poor naming agreement ($\leq 20\%$ subjects named the object similarly). Items may not have been flagged by the second criteria (e.g. if it received 4 different names, each with a 25% ratio), but could still be considered poor representations of the intended concepts.

54 photographs were found to meet at least one of the above criteria, and therefore excluded. Regardless of whether these items were grey or colour, it was also necessary to remove its grey or colour partner (since both versions were needed to make comparisons across recognition experiments). Thus, a total of 64 items (32 grey / 32 colour) were excluded at this stage (many items already had both grey and colour versions flagged by the original criteria).

Next, the proportion of correct responses were compared between grey and colour photographs in order to identify items showing the lowest difference. In order to manipulate colour in later recognition experiments, it was important to select items where naming was congruent across colour/grey items; in other words, it would be difficult to attribute particular recognition response patterns to the addition of colour (if a difference were found) when the grey version could not

be identified (or encoded) similarly. Variations exhibiting the least difference between colour and grey items (for the proportion of correct responses) were taken forward, while the rest were excluded. In a number of instances, multiple variations for the same object had the same ‘difference’ score. For example, all three variations of the item “balloon” exhibited perfect naming agreement, irrespective of whether they were presented in colour or grey (and thus “balloon1”, “balloon2”, and “balloon3” had a difference score of 0). For items where more than 1 variation remained, manual rankings were obtained from two of the researchers to determine which variation best depicted the intended concept. For each item, the researchers independently studied the remaining variations and provided a rank of which they thought was best (1) to worst (2 or 3, depending on the number of variations that remained). The ratings from both researchers were collated; items where there was agreement as to which variation best depicted the intended concept were selected for inclusion in the final stimuli list. For all the items where there was disagreement between the researchers rankings, one of the variations was simply selected at random.

Discussion

The role of colour For naming responses (accuracy, agreement [H], and RTs), no differences were observed between the grey and colour photographs. Such a result was expected for accuracy and agreement scores; the addition/absence of colour should not alter how participants identify (and thus label) items, except in rare instances whereby a lack of colour may lead to the misidentification of an object (e.g. incorrectly labelling a greyscale photograph of an orange as ‘grapefruit’). The data indicates, however, that this was not common, with the grey set of photographs exhibiting equally high levels of naming accuracy as the colour photographs. The absence of RT differences between the colour and greyscale sets was not expected for naming responses. It is reasonable to assume that colour photographs - with an additional layer of contextual information compared to grey items - would be identified (and therefore named) quicker than grey photographs (e.g. a colour photograph of an orange should avoid the potential ambiguity that might accompany a greyscale depiction, which could initially be confused for

another type of fruit). Indeed, Rossion & Pourtois (2004) demonstrated RTs consistent with this hypotheses, with colour drawings showing significantly quicker RTs than grey items. The lack of difference in the current data could be attributable to ceiling effects, whereby all photographs were sufficiently unambiguous, and were quickly identified irrespective of whether they were presented in greyscale or colour. Examination of the other naming data, showing similarly high levels of accuracy and agreement across grey and colour, supports this notion.

Scores of mental imagery agreement produced particularly interesting results between the grey and colour items. Grey photographs exhibited a significantly poorer match with subjects imagined presentation of the objects than the colour items. Colour differences were not observed previously between drawings (Rossion & Pourtois, 2004), and comparing the current data with that obtained in other studies (see Table 4) demonstrates how the greyscale photographs show uniquely lower mental imagery agreement scores compared with any of the other stimuli formats. To imagine the objects, it seems likely that subjects would conjure an image of how they naturally see the item in their everyday lives - which for the majority of subjects, would presumably be a colour representation. Therefore, when presented with greyscale depictions, subjects may have been more inclined to report that that item did not align quite as well as those presented in colour. However, it is unclear why a similar pattern is not also evident when comparing grey and colour drawings (Rossion & Pourtois, 2004). It may be that photographs promote stricter internal criteria when subjects must decide whether an item is a good match to their mental image. With line-drawn / illustrated items, subjects may simply accept that the items are baseline depictions, and that they will only able to match their real-world mental images to a certain degree - thus leading to a generally more liberal response bias throughout. The addition of colour may therefore do very little to further reconcile the match between the drawing and real-world mental representation. When subjects are responding only to photographs, the ecological nature of the items may facilitate deeper critical evaluation of whether they offer a good match to mental images, and thus promote a more conservative response bias. Colour may therefore be a far more important factor in photographs than it is in line drawings for allowing participants to decide whether an item matches well with their mental image.

There were no colour differences in familiarity scores. This result was expected - participants were asked to rate the degree to which they came in contact with, or think about, the concept itself rather than the particular depiction shown, and there is no apparent reason why colour should influence such ratings. Visual complexity, on the other hand, where participants were required to directly rate the amount of detail in the picture, did show an expected difference. Colour photographs were rated as significantly more visually complex than grey items, presumably due to their additional layer of contextual information. When compared to the previous data obtained for drawings, the greyscale photographs showed comparable levels of visual complexity, while the colour photographs showed higher levels than any of the other formats. It is unclear why the photographs of the current study showed colour differences, when grey and colour drawings did not differ, though it may tie in with the hypotheses proposed to explain the mental imagery agreement data. Subjects may apply stricter internal criteria when rating stimuli that are perceived as being closer to how they would be experienced in real life - when viewing a colour photograph of a rabbit, it is difficult to see how we could make the item any more visually complex than it already is (at least in a 2D medium). It's probable that subjects notice the absence of colour when viewing the greyscale items, since they depict the items in a way that they are not usually seen, and thus determine that these items could be made more complex if they were shown in colour (and so give lower visual complexity ratings as a result).

Establishing a new set of stimuli The objective of the current study was to establish a new set of ecological photograph stimuli to be taken forward into subsequent recognition memory experiments. Matching items with previously established drawings (and words) would allow for the effects of stimuli-format on recognition response patterns to be directly examined. A range of normative data was collected for 816 unique photograph items. These items may prove useful for a range of cognitive researchers that wish to utilise a set of high quality and realistic object stimuli, especially given the flexibility of items that can be filtered based on colour, naming agreement, familiarity, etc. For the needs of the current body of research, the naming data was used to determine which photographs best matched the intended concepts among a number of

possible variations. This allowed for the systematic comparison of recognition memory performance toward three distinct stimuli formats (words, drawings, and photographs) in the following study, in an effort to establish how stimuli of varying perceptual distinctiveness may affect recognition response patterns. Such comparisons might help to reconcile the inconsistencies present across recognition memory research, such as those attempted to determine whether familiarity processes are preserved in those with amnestic Mild Cognitive Impairment (aMCI).

Experiment: Effect of stimuli format (greyscale) and response option on recognition memory judgements.

For the recognition memory experiment, everyday objects were presented in three stimuli formats: i) words (written in simple, black ink); ii) drawings (shaded line-drawn illustrations); and iii) photographs (detail rich exemplars of the real world object). Rossion & Pourtois (2004) demonstrated that naming agreement could be improved by adding surface texture and shading to the original Snodgrass & Vanderwart (1980) items; however, it is unclear how manipulations to distinctiveness actually impact performance in recognition memory paradigms. As well as general inconsistencies regarding the type of stimuli used in recognition memory experiments, there is also much variability in the response options available to participants when reporting their recognition, for example: Remember/Know (Lombardi et al. (2016)), Recollection/Familiarity (???), or Low/Med/High confidence (???). In the current experiment, the availability of different response options when reporting recognition will also be examined by randomly assigning participants into a paradigm with three response options (Recollection / Familiarity / Guessing) or four response options (RFG + Both).

Based on the results of Experiment 1, which compared recognition to for words and drawings only, a number of hypotheses are proposed as to the potential effects of adding a third stimuli format (highly distinctive photograph stimuli). As stimuli become increasingly distinctive (from words, to drawings, to photographs), it seems likely that the number of hits (correctly recognised items) will increase, and the number of false alarms (FAs) will decrease. RFG responses are expected to show a similar pattern, with the most detailed stimuli showing the highest number

of hits assigned “Recollection”, while the less detailed formats show increasing levels of “Familiarity” and “Guessing” hits. Whilst we expect the overall number of FAs to increase as stimuli become less distinctive (i.e. words will show the highest rate of FAs), there is no reason to believe that these FAs will be biased toward any particular RFG judgement across formats. It is also hypothesised that the rates of reported Recollection and Familiarity will differ across response option conditions (RFG / RFBG), though the direction of this difference is currently unclear.

Method

Participants A total of 169 subjects completed the online experiment (see Table 5 for a breakdown of the gender and age of the sample). To meet our YA requirements, all participants were required to be between 18-59 years of age (actual range: 18-58). As our experiment involved English word stimuli, we also asked subjects whether English was their first language; the vast majority (95.86%) reported that English was indeed their first language. Subjects were recruited from voluntary participation websites such as Prolific Academic¹⁰ (73.37%), where payment at the rate of £5/hr was given, and via the in-school research participation system¹¹ (15.38%), where they received course participation credits. A small number of participants were also recruited from Psychological Research on the Net¹² (11.24%). In order to detect a medium effect size of Cohen’s $f = 0.25$ with 80% power ($\alpha = .05$, two-tailed), GPower indicated that we would need 79 participants per group ($N^* = 158$) in a 3x2 mixed ANOVA.

Table 5: Gender and age (SD) of the current sample.

Gender	N	Age	
Female	102	29.64	(10.22)
Male	63	30.98	(10.97)
Questioning	1	21.00	(0)
Unspecified	3	50.33	(4.93)
Total	169	30.46	(10.75)

¹⁰<https://www.prolific.co/>

¹¹<https://keelepsychology.sona-systems.com/>

¹²<https://psych.hanover.edu/research/exponnet.html>

Materials A total of 126 innocuous, everyday objects (e.g. clock, rabbit, shoe) were presented across three individual stimuli formats: written words, line drawings, and photographs. The line drawings were obtained from Rossion & Pourtois (2004), and consisted of greyscale shaded illustrations that contained some surface details. The word stimuli were simply the written word names of the line-drawn objects, presented in a clear Sans-serif typeface. The photograph stimuli were curated in the previous study; high quality photographs were sourced to similarly depict the same everyday objects as the line drawings. All objects in the photographs were isolated from their original background, converted to greyscale, and rotated to match the orientations shown in the line-drawn items.

Design The current study utilised a mixed design, with a 3-level within-subjects factor of stimuli format (words, drawings, photographs), and a 2-level between-subjects factor of response option (RFG, RFBG). Subjects passed through 2 levels of blocked randomization (equally sized, predetermined blocks); first, participants were randomly assigned one of six possible study lists (of equal length, and containing an even number of word, drawing, and photograph items) for counterbalancing purposes. Subjects were then either assigned into a recognition test with three possible response options (RFG: “Recollection”, “Familiarity”, “Guessing”), or four possible response options (RFBG: “Recollection”, “Familiarity”, “Guessing”, “Both”). These randomisation processes were completed automatically by the experiment software using balanced methods.

Words:	Drawings:	Photographs:
guitar		
mouse		
pumpkin		

Figure 10: Examples of the word, greyscale drawing, and greyscale photograph stimuli utilised in the current experiment.

Procedure Data was collected online using Gorilla¹³ - a platform for the building and hosting of online experiments. The experiment consisted of three self-paced phases: i) study phase, ii) distractor task, and iii) recognition test. In the study phase, an even mix of word, drawing, and photograph stimuli were presented one-at-a-time on the computer screen. Subjects were instructed to learn the items in preparation for a later memory test. To ensure attention was directed to the presented stimuli, subjects were required to report whether each item was shown as a word, drawing, or photograph using the computer mouse. Following the study phase, participants completed some simple multiple choice mathematical questions (e.g. $6 \times 4 = ?$) as a distractor. Finally, participants memory of the previously studied items was tested in the

¹³<https://gorilla.sc/>

recognition task. An even mix of word, drawing, and photograph stimuli were again presented one-at-a-time on the screen; half of the test items had been shown previously in the study phase, while the other half were new (and were not on the study list). For each item, subjects were instructed to press *Old* if they believed it was an item they had studied earlier, and *New* if they had not. *Old* responses led to a follow-up judgement, where participants reported whether they had experienced recognition through “Recollection”, “Familiarity”, or were simply taking an uninformed “Guess”. Participants that had been randomised into the RFBG test condition had a fourth option here, whereby they could report that they had experienced Recollection and Familiarity simultaneously (“Both”). Stimuli format was congruent across the study and test blocks (e.g. items presented as photos at study were also presented as photos at test). For each concept depicted across the three stimuli formats, subjects were only presented with one variation (in other words, if a subject saw a photograph for the item “shoe”, they did not see the word or line-drawn version of “shoe”).

Data processing Measured variables included the total number of hits and FAs, and the total number of hits and FAs assigned to each of the available response options (R/F/G and R/F/B/G). In order to create a common dependant variable, proportions were calculated from these variables slightly differently depending on the response option group. In the RFG-judgement group, simple proportions were created from the total number of R responses and the total number of F responses. In the RFBG condition, however, the proportion of Both responses was separately added to R proportions and F proportions. Additional DVs included: i) d' (d-prime, a signal detection measure of sensitivity); ii) c-value (a measure of response bias); iii) overall accuracy (hits / (hits + FAs)); iv) reaction times for all responses.

Participants were excluded from analysis if they showed poor performance during the encoding task; the relative ease of reporting whether each item was shown as a word, drawing, or photograph prompted a performance cut off of 90% accuracy. This allowed for some accidental clicks / incorrect responses toward potentially ambiguous items, though subjects scoring less than 90% were excluded on the assumption they did not dedicate their full attention to the task.

Subjects with extreme z-scores were also excluded from analysis; those presenting z-scores of +/- 3 (for total hits, total FAs, or overall recognition [hits minus FAs]) were considered outliers. These criteria resulted in the exclusion of 8 datasets, leaving a total of 161 in analyses.

Results

A series of 3x2 mixed ANOVAs were conducted on each of the DVs, using a within-subjects factor of stimuli format (words / grey drawings / grey photos) and a between-subjects factor of response option (RFG / RFBG). Significant main effects and interaction effects were followed-up with Bonferroni-adjusted pairwise comparisons. The Greenhouse-Geisser correction was applied when data was found to violate the assumption of sphericity (assessed using Mauchlys test statistic).

Stimuli distinctiveness Table 6: Mean proportion of hits, FAs, and mean d' scores, by stimuli format and response option condition.

	Hits	FAs	d'
Stimuli format			
Words	0.54	0.21	1.15
Grey drawings	0.76	0.09	2.38
Grey photographs	0.85	0.05	3.08
Response option			
RFG	0.74	0.13	2.25
RFBG	0.69	0.11	2.16

The mean proportion of hits and FAs, and mean d' scores are presented in Table 6. Results demonstrated a significant main effect of stimuli format in the ANOVA on the mean proportion of hits, $F(1.76, 280.25) = 225.67, p < .001, \eta_p^2 = .59$ Paired-samples t-tests showed that grey photographs ($M= 0.85$) produced a significantly higher proportion of hits than both words ($M= 0.54$; $t(160) = 18.56, p < .001$) and grey drawings ($M= 0.76$; $t(160) = -8.04, p < .001$). Grey drawings ($M= 0.76$) also produced a significantly higher proportion of hits compared to words ($M= 0.54$; $t(160) = 13.6, p < .001$). There were no significant interaction effects between stimuli format and response option condition $F(1.76, 280.25) = 0.58, p = .540, \eta_p^2 < .01$.

In the ANOVA on the mean proportion of FAs, there was again a significant main effect of stimuli format $F(1.43, 226.74) = 90.19, p < .001, \eta_p^2 = .36$. Photographs ($M= 0.05$) produced significantly fewer FAs in comparison to both words ($M= 0.21$), $t(160) = -11.84, p < .001$, and drawings ($M= 0.09$), $t(160) = 5.59, p < .001$. Compared to words ($M= 0.21$), drawings ($M= 0.09$) also showed a significantly lower proportion of FAs, $t(160) = -7.98, p < .001$. There were no significant interaction effects with response option condition, $F(1.43, 226.74) = 1.17, p = .299, \eta_p^2 < .01$.

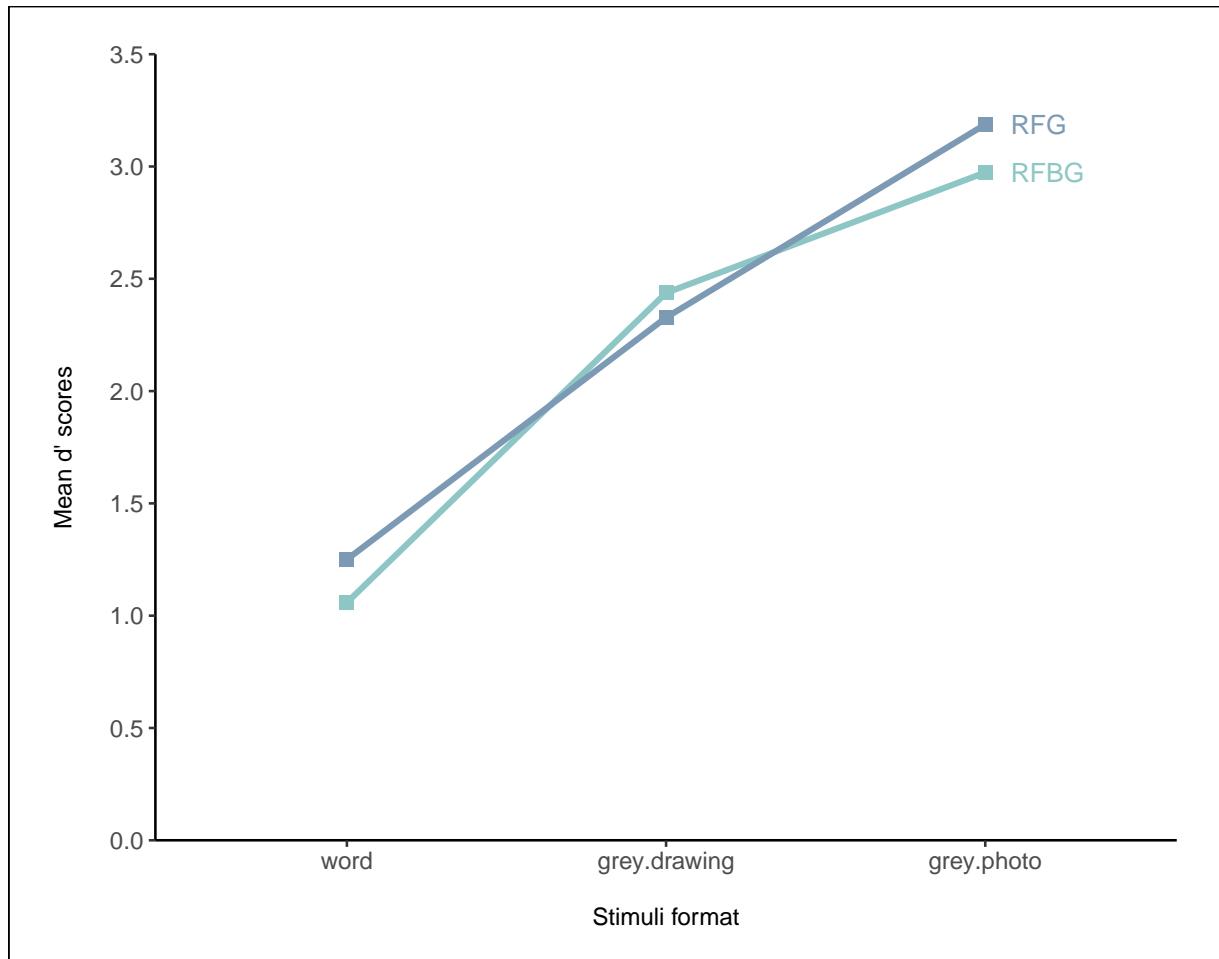


Figure 11: Interaction plot between stimuli format and response option for d' scores.

For mean d' scores, the ANOVA demonstrated a significant interaction between stimuli format and response option condition, $F(2, 318) = 3.34, p = .037, \eta_p^2 = .02$. While the RFG group produced numerically higher d' scores than the RFBG group for both words (RFG M

= 1.25, RFBG $M = 1.06$) and grey photographs (RFG $M = 3.19$, RFBG $M = 2.97$), neither comparison was statistically significant (words: $t(320.69) = -1.38, p > .999$; grey photos: $t(320.69) = -1.54, p > .999$. This pattern was reversed for grey drawings, where d' scores were numerically higher in the RFBG condition ($M = 2.44$) compared to the RFG condition ($M = 2.44$); though again, this comparison did not reach significance (grey drawings: $t(320.69) = 0.79, p > .999$).

#####-----

Chapter 4 (The role of colour)

In *Experiment 3*, this was demonstrated by comparing recognition toward stimuli of increasing levels of distinctiveness: words, line-drawn illustrated pictures, and detailed real-world photographs.

Indeed, colour images have previously been linked to enhanced memory performance in recognition tasks (Ally et al., 2009a), though it was difficult to disentangle the specific effects of colour from the complexity of detail featured.

Another inconsistency across recognition studies relates to whether picture stimuli are presented in greyscale or colour. The extent to which colour alters the distinctiveness of stimuli is unclear, though there is evidence to suggest it may play an important role in the memorability of stimuli. In a two-alternative forced choice recognition paradigm, Suzuki & Takahashi (1997) demonstrated that black & white images may not facilitate successful recognition in the same way as colour images. In an initial study phase, subjects were instructed to passively memorise a number of picture stimuli consisting of detailed real-world photographs of scenes (e.g. train stations, city streets etc.). At test, participants were presented with two images side-by-side (one target + one similar lure) and asked to select the item shown during the study phase. Manipulations were also made with regard to the congruency; photos were either presented in the same colour modality across study and test (1. b&w + b&w; 2. colour + colour) or different (3. b&w + colour; 4. colour + b&w). Participants were informed that old photos may be presented in a different colour format (e.g. a colour item shown at study might not be shown in b&w), and were also asked to make a source judgement about whether the colour format for each item was the same or different. Results showed that, of the four possible congruency conditions, recognition performance was best when items were congruently shown in colour at both study and test. This finding indicates colour information provided a recognition benefit at encoding and retrieval, though interestingly, this benefit was *only* evident in the congruent colour condition; if colour images were shown only at study, or only at test, there were no recognition benefits. Performance on the source judge-

ment question - regarding whether the colour of the item had changed since the study phase - was similar across all four conditions, but performance was particularly poor for items that had been presented in colour. This indicates recognition benefits were not a result of accessing memory for the colour information itself; Suzuki & Takahashi (1997) instead hypothesised that colour indirectly highlighted certain features within the photographs that were not otherwise noticed as prominently in b&w, and as a result, the colour photographs were overall more distinctive.

(colour = no effect) Stróżak

The current experiment aims to establish whether colour information facilitates recognition irrespective of the level of detail present in an image. The methodology of *Experiment 3* will be repeated using colour versions of both types of picture; a set of simple, line-drawn illustrations sourced from Rossion & Pourtois (2004), and a set of detailed real-world photographs established in *Experiment 2*. A secondary aim is to determine whether these response patterns are differentially affected by the availability of different response options at test (*Recollection/Familiarity/Guessing* or *Recollection/Familiarity/Both/Guessing*).

It is hypothesised that:

1. Distinctiveness: colour photographs will be more distinctive than colour drawings, and produce improved recognition performance manifesting as:

- i) higher overall d' scores;
- ii) higher proportion of correct hits;
- iii) lower proportion of false alarms;
- iv) better overall recognition.

2. The availability of different response options:

- i)

Experiment: Effect of stimuli format (colour) and response option on recognition memory judgements.

Method

Participants 161 participants completed the experiment online (see Table 7 for a breakdown of the age/gender of the current sample). All participants were required to be between the age of 18-59 years in order to meet our YA criteria (actual range: 18-57). As our experiment involved written words as to-be-remembered stimuli, we also asked that subjects first language be English; the vast majority (96.95%) reported that English was indeed their first language. Subjects were recruited from the voluntary participation website Prolific Academic¹⁴ (85.98%), where payment at the rate of £5/hr was given, and via the in-school research participation system¹⁵ (14.02%), where they received course participation credits. *G*Power* software was used to calculate an appropriate sample size; to detect a medium effect size of Cohen's $f = 0.25$ with 80% power ($\alpha = .05$, two-tailed), 79 subjects were necessary per group ($N = 158$) in a 3x2 mixed ANOVA.

Materials Stimuli were the same as those utilised in *Experiment 3*, except the greyscale drawings and photographs were substituted for their colour versions. Items consisted of 126 innocuous, everyday objects (e.g. clock, rabbit, shoe), presented across three individual stimuli formats: written words, line drawings, and photographs. Words and line drawings were sourced from Rossion & Pourtois (2004); the drawings consisted of shaded, colour illustrations, and the words were simply the written names of the depicted objects (these were presented in a clear Sans-serif typeface in the current experiment). A matching set of photograph stimuli were curated in *Experiment 2*; high quality photographs were sourced to similarly depict the same everyday objects as those found in the Rossion & Pourtois (2004) line drawings. In each photograph, the object of interest was isolated from its original background and rotated to match the orientations shown in the line-drawn items. See Figure 12 for examples of each stimuli format.

¹⁴<https://www.prolific.co/>

¹⁵<https://keelepsychology.sona-systems.com/>

Design A mixed 3x2 design was utilised, consisting of a within-subjects factor of stimuli format (words, drawings, photographs) and a between-subjects factor of response option (RFG, RFBG). Blocked randomisation presented participants with: 1) one of six possible study lists (equal length, with the same number of words, drawings, and photographs); 2) one of two possible recognition tests (either RFG: “Recollection”, “Familiarity”, “Guessing”), or RFBG: “Recollection”, “Familiarity”, “Guessing”, “Both”). All routes were of equal length, and subjects were randomly assigned into blocks via balanced methods.

Words:	Drawings:	Photographs:
cloud		
lock		
penguin		

Figure 12: Examples of the word, colour drawing, and colour photograph stimuli utilised in the current experiment.

Procedure The procedure was identical to that of *Experiment 3*; data collection was conducted online using the experiment platform Gorilla¹⁶. All subjects completed three self-paced phases: i) study phase, ii) distractor task, and iii) recognition test. At study, subjects were instructed to learn each of the word, drawing, and photograph items (shown at random, one-at-a-time) in preparation for a later memory test. For each item, participants were required to report whether the current format was a word, drawing, or photograph - an encoding judgement that ensured attention was directed toward the to-be-remembered stimuli. Next, subjects completed some simple multiple choice mathematical questions (e.g. $6 \times 4 = ?$) as a distractor task. Finally, participants were presented with the recognition test; word, drawing, and photograph items were once again shown one-at-a-time at random. Half of the test items had been shown previously in the study phase, while the other half were new (not shown at study). Subjects were first required to make an *Old/New* judgement, based on whether they believed they had studied the item earlier or not. While *New* judgements simply led to the next item, *Old* judgements led to a follow-up screen where participants were asked whether they had recognised the item via *Recollection*, *Familiarity*, or were simply *Guessing* that it was old. Those in the RFBG response option condition had an additional *Both* option at this stage, where they could report that they had experienced recollection and familiarity simultaneously. Stimuli format stayed the same across study and test (e.g. if the item “penguin” was shown in word format at study, it was also shown as a word at test), and the same concepts were not repeated across the other formats within-subjects (e.g. if the item “penguin” was shown as a word, that subject would not view the drawing or photo version).

Data processing The primary DVs consisted of d' (d-prime, a signal detection measure of sensitivity), the proportion of hits and false alarms (FAs), overall performance accuracy (hits / (hits + FAs)), and the total number of hits and FAs assigned to each of the available response options (R/F/G and R/F/B/G). Proportions of Recollection and Familiarity were calculated slightly differently depending on the response option condition; in the RFG-judgement group, simple

¹⁶<https://gorilla.sc/>

proportions were created from the total number of R responses and the total number of F responses. In the RFBG condition, Both responses was separately added to R proportions and F proportions. Additional DVs included: *c* scores (a measure of response bias), and reaction times for all responses.

All analyses were conducted using R (R Core Team, 2020). *d'* (sensitivity) and *c* (bias) scores were calculated using the 'psycho' package (v0.5.0; Makowski, 2018). *d'* scores were calculated via: z-scores for correct hits minus z-scores for false alarms (Hautus, 1995 adjustments for extreme values were applied).

Subjects were excluded according to two key criteria; 1) less than 90% accuracy during the encoding task ("Is this a word, drawing, or photograph?"); 2) extreme z-scores (those presenting z-scores of +/- 3 for total hits, total FAs, or overall recognition [hits minus FAs]). A total of 3 participants were found to meet (at least) one of these criteria, and were thus considered outliers and excluded from analysis.

Results

A series of 3x2 mixed ANOVAs were conducted on each of the DVs, with a within-subjects factor of stimuli format (words / colour drawings / colour photos) and a between-subjects factor of response option (RFG / RFBG). Significant main effects and interaction effects were followed-up with Bonferroni-adjusted pairwise comparisons. The Greenhouse-Geisser correction was applied when data was found to violate the assumption of sphericity (assessed using Mauchlys test statistic).

Distinctiveness Since *Experiment 3* already established the current set of photograph stimuli as being more distinctive than words when presented in greyscale, it seemed unnecessary to make this comparison again. Instead, the following analyses specifically aimed to deter-

mine whether colour photographs were more distinctive than colour drawings, or if distinctiveness ‘peaked’ such that colour drawings and photographs resulted in comparable recognition performance. A series of 2 (stimuli format: words, pictures) x 3 (response option condition: RFG-judgements, RFBG-judgements, RF-ratings) mixed ANOVAs were conducted on the signal detection measures of d' (sensitivity) and c (decision criterion), as well as the proportion of overall hits, false alarms (FAs), and overall recognition (hits - FAs) [see Table 8]. Significant main effects and interaction effects were followed-up with Bonferroni-adjusted pairwise comparisons.

Visual inspection of the data (Table 8) shows some expected patterns with regard to stimuli distinctiveness. As the intended distinctiveness increases across the formats (from words, to drawings, to photographs), d' scores also increase, along with the mean proportion of hits and overall scores of recognition (hits - FAs), while the mean proportion of FAs decreases.

In determining whether the colour photos were more distinctive than the colour drawings, all of these variables showed significant main effects of stimuli format (see Table 9). Photos showed significantly higher d' scores (thus better discrimination between hits and FAs), proportion of hits, and overall recognition than drawings. Photos also produced significantly fewer FAs and showed lower c scores (and thus a more liberal response bias) than drawings.

Table 9: ANOVA results for the main effects of stimuli format on a number of outcome variables.

Taken together, the findings indicate that the colour photographs were indeed more distinctive than the colour drawings, and that the current set of stimuli appear to effectively capture different ‘levels’ of stimuli distinctiveness.

Response option In each of the aforementioned ANOVAs, the role of response option was also examined to determine whether there would be any differential effects as a result of using colour stimuli, as opposed to the greyscale items used in *Experiment 3*. The results showed no significant interaction effects between stimuli format and response option for any of the above variables. The only main effects of response option were in the ANOVA on the mean proportion of FAs, (with the RFG group showing more FAs than the RFBG group, RFG: , RFBG: ,) and the

ANOVA on c scores, (with the RFG group showing a more liberal response bias than the RFBG group, RFG: , RFBG: ,).

Role of colour Visual comparison of the data from *Experiment 3* (greyscale drawings and photographs) and that obtained in the current study indicates that colour has a negligible impact on stimuli distinctiveness (see Figure 13). While there were clear performance differences between the different levels of stimuli distinctiveness in both studies,

Performance, in terms of the proportion of hits, FAs, and overall recognition accuracy, appears the same irrespective of whether greyscale or colour pictures were used.

In turn, this further implicates the stimuli format as the

Discussion

#####-----

Chapter 5

Experiment: Manipulating word distinctiveness

Method

All stimuli were presented as 500x500px image files at their actual size (i.e. without scaling) to ensure consistency across participants. Adobe Photoshop 2021 (22.0.0 Release) was used to create the image files, all on a plain white canvas. The Rossion and Pourtois (2004) line-drawings were sourced directly from the authors, and were trimmed of any preexisting background before being resized to fit the 500x500px canvas (for some items this was determined by a maximum width of 500px, whilst for others it was determined by a maximum height of 500px).

word stimuli were created in Roboto (light, 54pt)

1) Manipulation check (1st Bonferroni pool)

- Are distinctive words actually more distinctive?
- If its not significant, might as well stop analysis.
- low word vs high word.

2) Have I established a PSE? (2 + 3 in same Bonferroni pool)

- Must demonstrate a PSE first before we can even begin to check whether any of the sub-conditions can eliminate it.
 - Main effect of stim format + contrasts:
 - all 3(?)
- Main effect might not be the only indication of this though (if I
- Have to do planned contrasts to see if High > Low relationship holds to words.

(wouldn't do all 3 contrasts. only ever do 2 comparisons (or at least start with 2). if words bigger than pics)

3) Have I managed to eliminate PSE in the same way Ensor has?

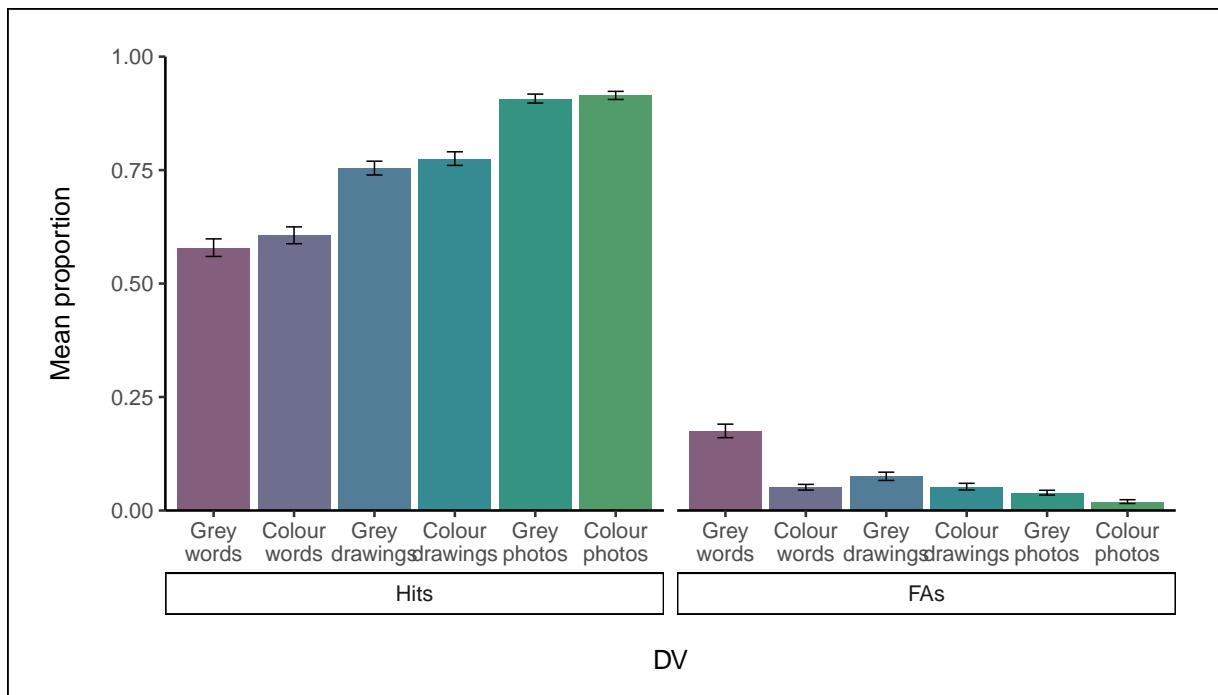
- Distinctive words vs. grey drawings (reverse PSE)
- Distinctive words vs. colour drawings (PSE - reduced)

4) Have I extended it to photos? (3rd Bonferroni pool)

- Do they behave differently to each other? Are photos more distinctive than drawings?
 - drawings high vs photos high

Results

Comparisons are made in relation to the first two experiments of Ensor et al. (2019), the methodologies of which were identical to one another, apart from the number of trials involved (one experiment consisting of 40 trials at study and 80 at test, and the other consisting of 80 trials at study and 160 at test). With regard to trial numbers, the current study sits between the two Ensor et al. (2019) experiments, with 60 trials at study and 120 at test.



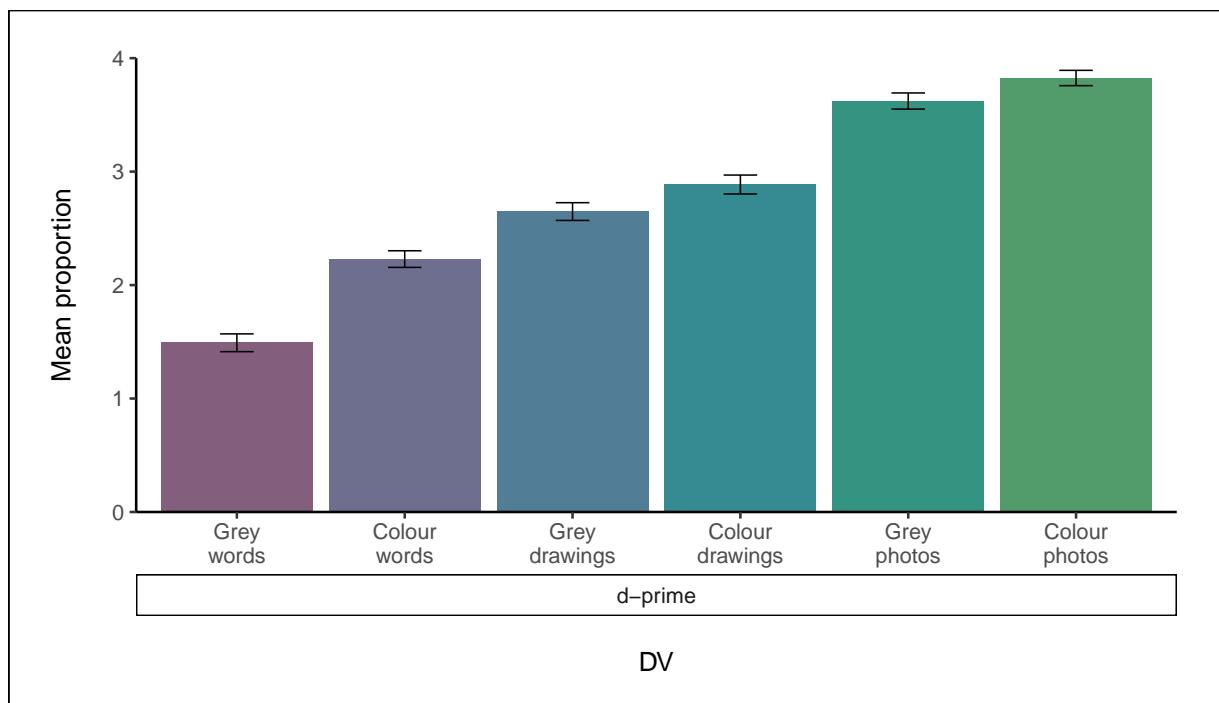


Figure 14: Mean proportion of hits, FAs, and d' scores by stimuli format.

Manipulation check To establish whether the distinctive word stimuli were indeed more distinctive than regular words, a series of one-way repeated measures analyses of variance (ANOVA) - using six factors (grey & distinctive words, grey & colour drawings, grey & colour photographs) - were conducted on the mean scores of d' , mean proportions of hits, and mean proportion of FAs. Sphericity was assessed using Mauchly's Test; when the assumption of sphericity was violated, Greenhouse-Geisser corrections were applied.

Throughout each of the conducted ANOVAs, there was consistently a significant main effect of stimuli format: mean proportion of hits [$F(3.95, 619.49) = 137.94, p < .001, \eta_p^2 = .47$], mean proportion of FAs [$F(2.83, 443.69) = 52.71, p < .001, \eta_p^2 = .25$], d' -scores [$F(4.74, 744.81) = 190.22, p < .001, \eta_p^2 = .55$]. Planned comparisons following these significant results, however, did not always demonstrate that the distinctive word stimuli were indeed any more distinctive than the regular word stimuli. Primarily, there was no difference in the proportion of overall hits between distinctive ($M = 0.61$) and regular words ($M = 0.58$), $t(157) = 1.26, p = .211, d = 0.10$ - recognition performance (i.e. the number of items subjects were able to correctly identify) was the same regardless of the format word stimuli were presented in. Subjects did, however, make significantly fewer FAs as a result of the manipulation (distinctive words $M = 0.05$, regular words $M = 0.18$; $t(157) = -8.70, p < .001, d = -0.69$), offering some support to the manipulation, as subjects were better able to correctly reject lure items when they were presented in a distinctive format, rather than regular format, highlighting some performance benefits. Furthermore, distinctive words ($M = 2.23$) also showed significantly higher d' scores than regular words ($M = 1.49$), $t(157) = 7.96, p < .001, d = 0.63$; subjects were better able to accurately discriminate between hits and FAs when responding to the distinctive words over regular words.

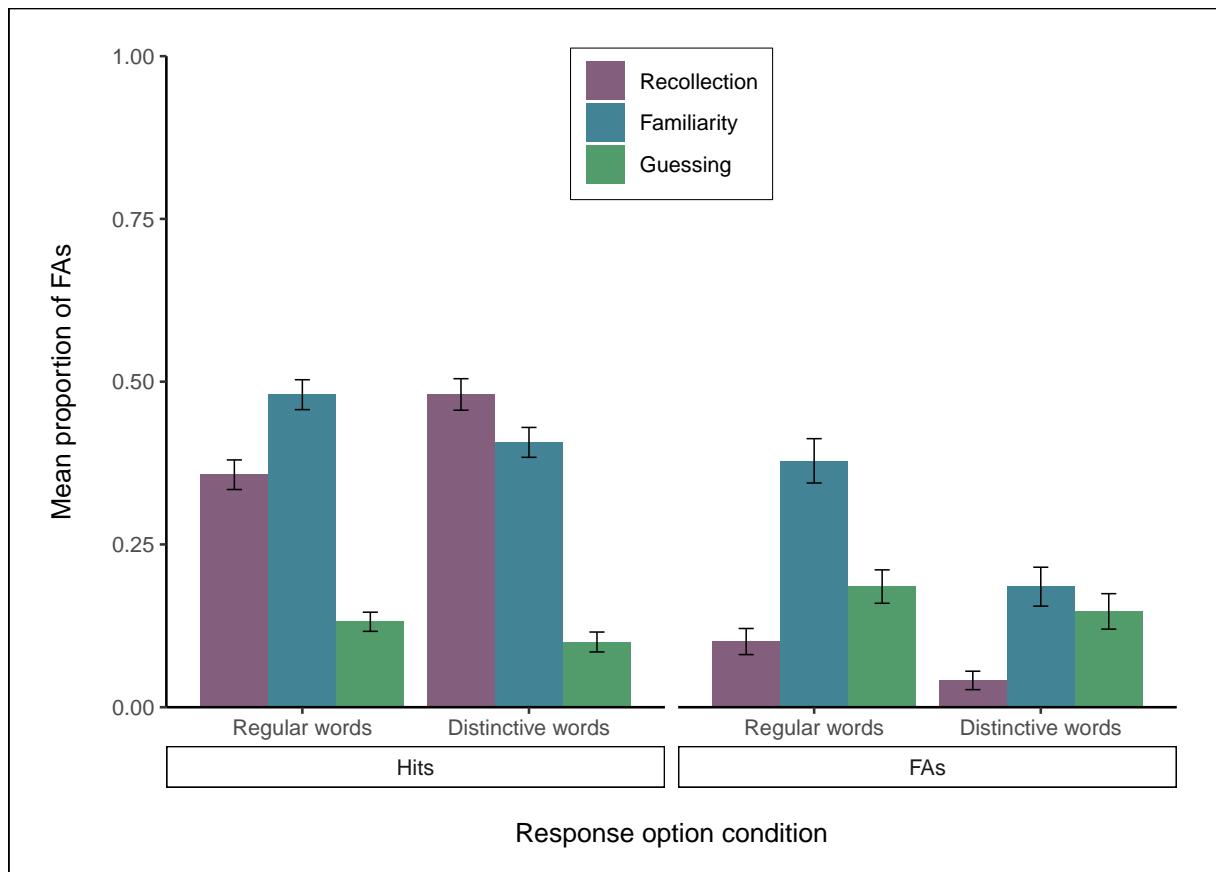


Figure 15: Mean proportion of hits and FAs by stimuli format.

Raw data provided by Ensor et al. (2019) was re-analysed to allow for direct comparisons with data from the current study. For evidence of the manipulation check:

For mean d' scores, there were significant main effects of stimuli format in both the first [$F(3, 87) = 10.99, p < .001, \eta_p^2 = .27$] and second [$F(3, 87) = 19.01, p < .001, \eta_p^2 = .40$] experiments. Follow-up comparisons showed evidence of the distinctiveness manipulation, whereby distinctive words produced significantly higher d' scores compared to regular words in both experiments (see Table 10). Visual inspection of the data shows that the d' scores obtained in the current experiment sit between the Ensor et al. (2019) experiments - in line with reducing scores as the number of trials increases.

Table 10: Mean d' scores for regular and distinctive word stimuli, across the current experiment and Ensor et al. (2019).

	$*d'$		
	Grey words	Distinctive words	Planned comparisons
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	2.18	2.97	$*t^*(29) = 3.69, *p^* < .001, *d^* = 0.67$
Trials: 80 study / 160 test:	1.09	1.89	$*t^*(29) = 7.35, *p^* < .001, *d^* = 1.34$
Current study			
Trials: 60 study / 120 test:	1.49	2.23	$*t^*(157) = 7.96, *p^* < .001, *d^* = 0.63$

The ANOVA on the mean proportion of hits also showed evidence of the distinctiveness manipulation; significant main effects of stimuli format in their first [$F(3, 87) = 10.91, p < .001, \eta_p^2 = .27$] and second [$F(3, 87) = 8.37, p < .001, \eta_p^2 = .22$] experiments produced follow-up comparisons that consistently showed distinctive words resulted in significantly more hits than regular words (see Table 11). This is in contrast to the results of the current study, whereby no difference was found as a result of the distinctiveness manipulation. Visual inspection of the data does not indicate the number of trials had an impact on the number of hits; while Ensor et al. (2019)

did observe a reduction in the proportion of hits for both word formats when trial numbers were increased, the current study still showed a lower proportion of hits than both of the Ensor et al. (2019) experiments (including the experiment with more trials).

Table 11: Mean proportion of hits for regular and distinctive word stimuli, across the current experiment and Ensor et al. (2019).

	Hits		
	Grey words	Distinctive words	Planned comparisons
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	0.72	0.84	*t*(29) = 3.14, *p* = .004, *d* = 0.57
Trials: 80 study / 160 test:	0.63	0.78	*t*(29) = 4.16, *p* < .001, *d* = 0.76
Current study			
Trials: 60 study / 120 test:	0.58	0.61	*t*(157) = 1.26, *p* = .211, *d* = 0.10

For the mean proportion of FAs, the second experiment of Ensor et al. (2019) showed a significant main effect of stimuli format [$F(3, 87) = 6.05, p < .001, \eta_p^2 = .17$], much like the current study. However, while the current study showed distinctive words produced significantly fewer FAs than regular words, this comparison was not significant for Ensor et al. (2019) (see Table 12). In their first experiment, there was no significant main effect of stimuli format [$F(3, 87) = 1.41, p = .245, \eta_p^2 = .05$], or follow-up comparison. Visual inspection of the data shows that subjects generally made more FAs in the Ensor et al. (2019) experiments compared to the current study; this is especially apparent when comparisons are made toward the experiment with more trials.

Table 12: Mean proportion of FAs for regular and distinctive word stimuli, across the current experiment and Ensor et al. (2019).

FAs			
	Grey words	Distinctive words	Planned comparisons
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	0.15	0.08	*t*(29) = -1.94, *p* = .062, *d* = -0.35
Trials: 80 study / 160 test:	0.29	0.23	*t*(29) = -1.48, *p* = .149, *d* = -0.27
Current study			
Trials: 60 study / 120 test:	0.18	0.05	*t*(157) = -8.70, *p* < .001, *d* = -0.69

Elimination of the PSE: comparison with Ensor et al. (2019) A key finding in Ensor et al. (2019) data was an elimination of the PSE; distinctive word stimuli and grey drawings produced similar d' scores, indicative of comparable distinctiveness between the two stimuli formats. Such findings were replicated in the current reanalysis of their data (see Table 13). In the current experiment, however, this finding was not replicated, with drawings showing significantly higher d' than the distinctive words, suggesting the manipulation was not effective.

Table 13: Mean d' scores for distinctive word stimuli and grey drawings, across the current experiment and Ensor et al. (2019).

	d'		Planned comparisons
	Distinctive words	Grey drawings	
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	2.97	2.96	* $t^*(29) = 0.06$, * $p^* = .956$, * $d^* = 0.01$
Trials: 80 study / 160 test:	1.89	1.90	* $t^*(29) = -0.07$, * $p^* = .945$, * $d^* = -0.01$
Current study			
Trials: 60 study / 120 test:	2.23	2.65	* $t^*(157) = -4.70$, * $p^* < .001$, * $d^* = -0.37$

Similarly, Ensor et al. (2019) also showed no difference in the number of hits between the distinctive words and grey drawings, across both experiments (see Table 14). This again suggests their distinctiveness manipulation was successful in eliminating the PSE, however, such a finding was not apparent in the current study, where grey drawings produced significantly more hits than the distinctive words.

Table 14: Mean proportion of hits for distinctive word stimuli and grey drawings, across the current experiment and Ensor et al. (2019).

	Hits		
	Distinctive words	Grey drawings	Planned comparisons
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	0.84	0.86	*t*(29) = -0.69, *p* = .495, *d* = -0.13
Trials: 80 study / 160 test:	0.78	0.75	*t*(29) = 0.98, *p* = .333, *d* = 0.18
Current study			
Trials: 60 study / 120 test:	0.61	0.75	*t*(157) = -7.28, *p* < .001, *d* = -0.58

Finally, Ensor et al. (2019) also showed no difference in the number of FAs between the distinctive words and grey drawings, across both experiments (see Table 15) - an expected result if the distinctiveness manipulation resulted in similar levels between the distinctive words and grey drawings. Again, the current experiment produced different results; interestingly though, there were significantly more FAs for the grey drawings, as opposed to the distinctive word stimuli. This is somewhat unexpected, since the rest of the findings identify the grey drawings as having increased memorability compared to the distinctive words.

Table 15: Mean proportion of FAs for distinctive word stimuli and grey drawings, across the current experiment and Ensor et al. (2019).

	FAs		
	Distinctive words	Grey drawings	Planned comparisons
Ensor, Surprenant, & Neath (2019)			
Trials: 40 study / 80 test:	0.08	0.12	*t*(29) = -1.38, *p* = .178, *d* = -0.25
Trials: 80 study / 160 test:	0.23	0.19	*t*(29) = 1.06, *p* = .297, *d* = 0.19
Current study			
Trials: 60 study / 120 test:	0.05	0.08	*t*(157) = -2.48, *p* = .014, *d* = -0.20

Recollection and Familiarity

To determine if recollection and familiarity response patterns were mediated by the format of word stimuli, further one-way repeated measures analyses of variance (ANOVA) were conducted on the mean proportion of hits and false alarms (FAs) assigned *Recollection* (R), *Familiarity* (F), and *Guessing* (G).

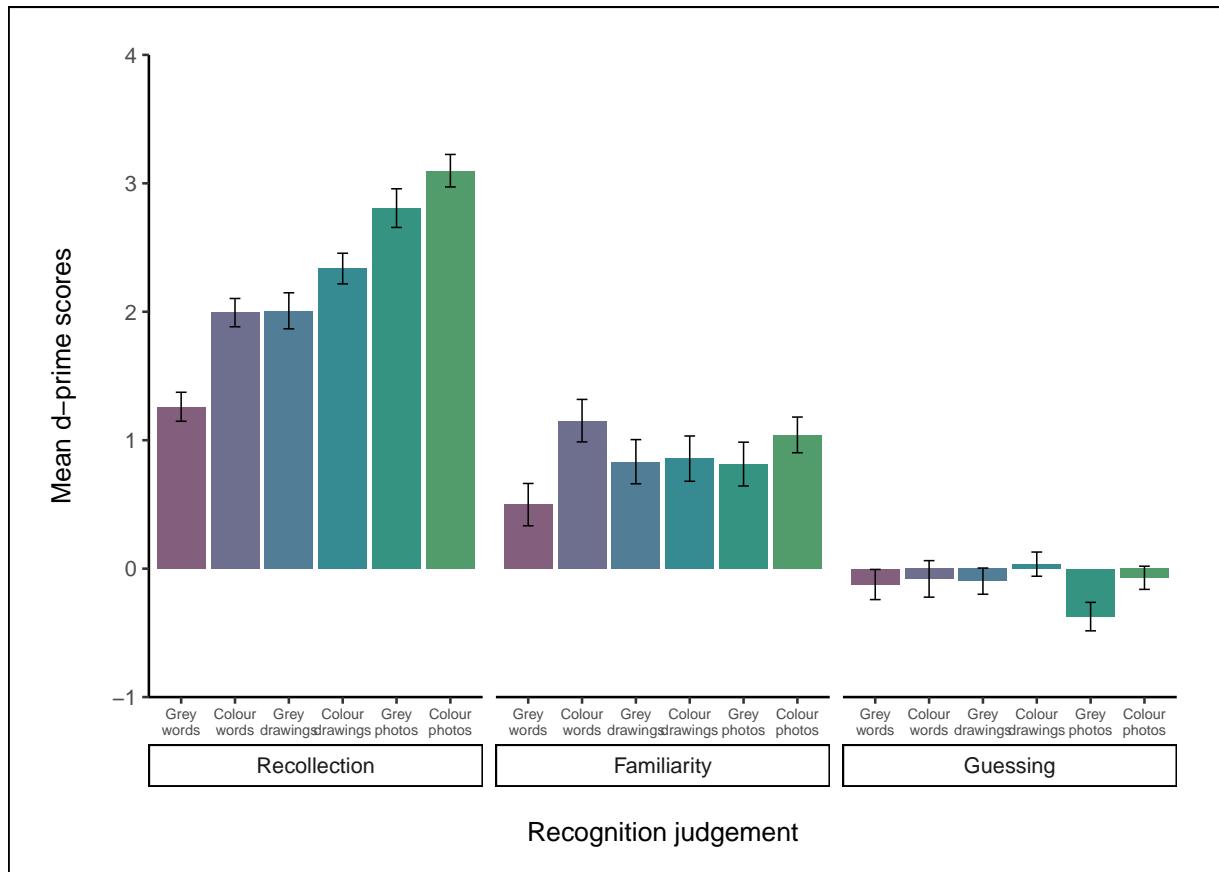


Figure 16: Mean d' scores by stimuli format and recognition judgement.

For responses assigned *Recollection*, there was a significant main effect of stimuli format for the mean proportion of hits [$F(3.10, 486.70) = 73.92, p < .001, \eta_p^2 = .32$], mean proportion of FAs [$F(4.38, 687.03) = 3.89, p = .003, \eta_p^2 = .02$], and mean d' scores [$F(4.40, 690.83) = 40.88, p < .001, \eta_p^2 = .21$].

Distinctiveness manipulation (regular words vs. distinctive words): Across each of the variables, the distinctive word stimuli indeed appeared to be more distinctive than the regular words. Planned comparisons showed that distinctive words ($M = 0.48$) produced significantly more R hits than regular words ($M = 0.36$), $t(157) = 4.74, p < .001, d = 0.38$, and also significantly higher d' scores (distinctive words $M = 1.99$, regular words $M = 1.26$, $t(157) = 5.25, p < .001, d = 0.42$. Conversely, regular words ($M = 0.1$) produced a significantly higher proportion of FAs assigned R than distinctive words ($M = 0.04$), $t(157) = -2.65, p = .009, d = -0.21$.

Elimination of the PSE (distinctive words vs. greyscale drawings): Grey drawings ($M = 0.56$) produced a significantly higher number of hits assigned R compared to the distinctive words ($M = 0.48$), $t(157) = -3.64, p < .001, d = -0.29$.

On the other hand, the distinctive words ($M = 0.04$) produced significantly fewer FAs assigned R than grey drawings ($M = 0.11$), $t(157) = -2.41, p = .017, d = -0.19$. This somewhat extended to the colour drawings, where there was no difference between the distinctive words ($M = 0.04$) and colour drawings ($M = 0.06$), $t(157) = 0.71, p = .478, d = 0.06$. Comparisons with grey photos, however, showed a return of the PSE, as the distinctive words ($M = 0.04$) produced a significantly higher number of R FAs than grey photo ($M = 0.07$), $t(157) = -2.41, p = .017, d = -0.19$.

For d' scores, there was an elimination of the PSE; distinctive words ($M = 1.99$) and grey drawings ($M = 2.01$) showed no significant difference in mean d' scores, $t(157) = -0.09, p = .932, d < 0.01$.

Extension to colour drawings: Elimination of the PSE for d' scores did not extend to colour drawings. Colour drawings ($M = 2.34$) showed significantly higher d' scores than the distinctive words ($M = 1.99$), $t(157) = 2.30, p = .023, d = 0.18$.

Extension to photos: Grey photos ($M = 2.81$) showed significantly higher d' scores than the

distinctive words ($M = 1.99$), $t(157) = -4.95$, $p < .001$, $d = -0.39$.

Colour photos ($M = 3.1$) showed significantly higher d' scores than the distinctive words ($M = 1.99$), $t(157) = 8.26$, $p < .001$, $d = 0.66$.

For responses assigned *Familiarity*, there was again a significant main effect of stimuli format across all of the variables; the mean proportion of hits [$F(3.09, 484.95) = 25.45, p < .001, \eta_p^2 = .14$], mean proportion of FAs [$F(4.54, 713.44) = 15.62, p < .001, \eta_p^2 = .09$], and mean d' scores [$F(4.52, 709.38) = 2.52, p = .033, \eta_p^2 = .02$].

Distinctiveness manipulation (regular words vs. distinctive words):

Regular words ($M = 0.48$) produced significantly more F hits than the distinctive words ($M = 0.41$), $t(157) = -2.53, p = .012, d = -0.20$, and significantly more F FAs (regular words $M = 0.38$, distinctive words $M = 0.19$, $t(157) = -4.56, p < .001, d = -0.36$). d' scores, however, were significantly higher for the distinctive words ($M = 1.15$) compared to the regular words ($M = 0.5$), $t(157) = 2.93, p = .004, d = 0.23$.

Elimination of the PSE (distinctive words vs. greyscale drawings):

For the proportion of F hits, the PSE was eliminated, with no significant difference between distinctive words ($M = 0.41$) and grey drawings ($M = 0.38$), $t(157) = 0.84, p = .403, d = 0.07$. Likewise, there was no significant difference for the number of FAs assigned F between distinctive words ($M = 0.19$) and grey drawings ($M = 0.23$), $t(157) = -1.24, p = .219, d = -0.10$. For d' scores, there was again an elimination of the PSE; distinctive words ($M = 1.15$) and grey drawings ($M = 0.83$) showed no significant difference in mean d' scores, $t(157) = 1.54, p = .125, d = 0.12$.

Extension to colour drawings: Elimination of the PSE for d' scores also extended to colour drawings. Mean d' scores were not significantly different between distinctive words ($M = 1.15$) and colour drawings ($M = 0.86$), $t(157) = -1.33, p = .184, d = -0.11$.

Extension to photos: Elimination of the PSE for d' scores also extended to grey photos. Mean d' scores were not significantly different between distinctive words ($M = 1.15$) and grey photos ($M = 0.81$), $t(157) = 1.58, p = .116, d = 0.13$.

Elimination of the PSE for d' scores also extended to colour photos. Mean d' scores were not significantly different between distinctive words ($M = 1.15$) and colour photos ($M = 1.04$), $t(157) = -0.63, p = .527, d = -0.05$.

For responses assigned *Guessing*, there was a significant main effect of stimuli format for the mean proportion of hits [$F(2.99, 468.71) = 24.58, p < .001, \eta_p^2 = .14$] and mean proportion of FAs [$F(4.69, 736.23) = 4.66, p < .001, \eta_p^2 = .03$]. However, the main effect of mean d' scores did not reach significance [$F(4.51, 708.09) = 1.57, p = .172, \eta_p^2 < .01$]; visual inspection of this data shows that across the stimuli formats, d' scores were generally negative for *Guessing* responses, an expected pattern that shows poor discrimination and often more FAs than hits.

Distinctiveness manipulation (regular words vs. distinctive words): There was no difference in the number of hits assigned G between regular words ($M = 0.13$) and distinctive words ($M = 0.1$), $t(157) = -1.75, p = .082, d = -0.14$, no difference in the number of FAs assigned G between regular words ($M = 0.19$) and distinctive words ($M = 0.15$), $t(157) = -1.10, p = .275, d = -0.09$, and no difference in mean d' scores between regular words ($M = -0.12$) and distinctive words ($M = -0.08$), $t(157) = 0.23, p = .815, d = 0.02$.

Elimination of the PSE (distinctive words vs. greyscale drawings):

For the proportion of G hits, the distinctive words ($M = 0.1$) produced a significantly higher proportion compared to the grey drawings ($M = 0.05$), $t(157) = 3.28, p = .001, d = 0.26$. There was no significant difference for the number of FAs assigned G between distinctive words ($M = 0.15$) and grey drawings ($M = 0.1$), $t(157) = 1.41, p = .161, d = 0.11$, or mean d' scores (distinctive words $M = -0.08$, grey drawings $M = -0.1$), $t(157) = 0.10, p = .920, d < 0.01$.

#####-----

References

- Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, 31(6), 731–753. <https://doi.org/10.1080/13803390802488103>
- Algarabel, S., Escudero, J., Mazón, J. F., Pitarque, A., Fuentes, M., Peset, V., & Lacruz, L. (2009). Familiarity-based recognition in the young, healthy elderly, mild cognitive impaired and Alzheimer's patients73. *Neuropsychologia*, 47(10), 2056–2064. <https://doi.org/10.1016/j.neuropsychologia.2009.03.016>
- Algarabel, S., Fuentes, M., Escudero, J., Pitarque, A., Peset, V., Mazón, J.-F., & Meléndez, J.-C. (2012). Recognition memory deficits in mild cognitive impairment. *Aging, Neuropsychology, and Cognition*, 19(5), 608–619. <https://doi.org/10.1080/13825585.2011.640657>
- Ally, B. A. (2012). Using Pictures and Words To Understand Recognition Memory Deterioration in Amnestic Mild Cognitive Impairment and Alzheimer's Disease: A Review. *Current Neurology and Neuroscience Reports*, 12(6), 687–694. <https://doi.org/10.1007/s11910-012-0310-7>
- Ally, B. A., & Budson, A. E. (2007). The worth of pictures: Using high density event-related potentials to understand the memorial power of pictures and the dynamics of recognition memory. *NeuroImage*, 35(1), 378–395. <https://doi.org/10.1016/j.neuroimage.2006.11.023>
- Ally, B. A., Gold, C. A., & Budson, A. E. (2009a). The picture superiority effect in patients with Alzheimer's disease and mild cognitive impairment. *Neuropsychologia*, 47(2), 595–598. <https://doi.org/10.1016/j.neuropsychologia.2008.10.010>
- Ally, B. A., McKeever, J. D., Waring, J. D., & Budson, A. E. (2009b). Preserved frontal memorial processing for pictures in patients with mild cognitive impairment. *Neuropsychologia*, 47(10), 2044–2055. <https://doi.org/10.1016/j.neuropsychologia.2009.03.015>
- Ally, B. A., Waring, J. D., Beth, E. H., McKeever, J. D., Milberg, W. P., & Budson, A. E. (2008). Aging memory for pictures: Using high-density event-related potentials to understand the

- effect of aging on the picture superiority effect. *Neuropsychologia*, 46(2), 679–689. <https://doi.org/10.1016/j.neuropsychologia.2007.09.011>
- Anderson, N. D., Ebert, P. L., Jennings, J. M., Grady, C. L., Cabeza, R., & Graham, S. J. (2008). Recollection- and familiarity-based memory in healthy aging and amnestic mild cognitive impairment. *Neuropsychology*, 22(2), 177–187. <https://doi.org/10.1037/0894-4105.22.2.177>
- Barba, G. D. (1997). Recognition Memory and Recollective Experience in Alzheimer's Disease. *Memory*, 5(6), 657–672. <https://doi.org/10.1080/741941546>
- Bastin, C., Van der Linden, M., Michel, A.-P., & Friedman, W. J. (2004). The effects of aging on location-based and distance-based processes in memory for time. *Acta Psychologica*, 116(2), 145–171. <https://doi.org/10.1016/j.actpsy.2003.12.014>
- Belleville, S., Ménard, M.-C., & Lepage, É. (2011). Impact of novelty and type of material on recognition in healthy older adults and persons with mild cognitive impairment. *Neuropsychologia*, 49(2011), 2856–2865.
- Bermúdez-Margaretto, B., Beltrán, D., Cuetos, F., & Domínguez, A. (2018). Brain Signatures of New (Pseudo-) Words: Visual Repetition in Associative and Non-associative Contexts. *Frontiers in Human Neuroscience*, 12, 354. <https://doi.org/10.3389/fnhum.2018.00354>
- Biederman, I. (1987). *Recognition-by-Components: A Theory of Human Image Understanding*. 33.
- Bowen, H. J., Fields, E. C., & Kensinger, E. A. (2019). Prior Emotional Context Modulates Early Event-Related Potentials to Neutral Retrieval Cues. *Journal of Cognitive Neuroscience*, 31(11), 1755–1767. https://doi.org/10.1162/jocn_a_01451
- Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. Independent ratings. *Journal of Memory and Language*, 65(2), 98–108. <https://doi.org/10.1016/j.jml.2011.04.003>
- Cui, L., Shi, G., He, F., Zhang, Q., Oei, T. P. S., & Guo, C. (2016). Electrophysiological Correlates

- of Emotional Source Memory in High-Trait-Anxiety Individuals. *Frontiers in Psychology*, 7.
- <https://doi.org/10.3389/fpsyg.2016.01039>
- Curran, T., & Doyle, J. (2011). Picture Superiority Doubly Dissociates the ERP Correlates of Recollection and Familiarity. *Journal of Cognitive Neuroscience*, 23(5), 1247–1262. <https://doi.org/10.1162/jocn.2010.21464>
- Deason, R. G., Hussey, E. P., Flannery, S., & Ally, B. A. (2015). Preserved conceptual implicit memory for pictures in patients with Alzheimer's disease. *Brain and Cognition*, 99, 112–117. <https://doi.org/10.1016/j.bandc.2015.07.008>
- Dewhurst, S. A., & Conway, M. A. (1994). *Pictures, Images, and Recollective Experience*. 11.
- Dobbins, I. G., KroU, N. E. A., & Liu, Q. (1998). Confidence-Accuracy Inversions in Scene Recognition: A Remember-Know Analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1306–1315.
- Donaldson, W., Mackenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review*, 3(4), 486–490. <https://doi.org/10.3758/BF03214551>
- Dunn, J. C. (2008). The Dimensionality of the RememberKnow Task: A State-Trace Analysis. *Psychological Review*, 115(2), 426–446.
- Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, 9(1), 139–145. <https://doi.org/10.3758/BF03196270>
- Embree, L. M., Budson, A. E., & Ally, B. A. (2012). Memorial familiarity remains intact for pictures but not for words in patients with amnestic mild cognitive impairment. *Neuropsychologia*, 50(9), 2333–2340. <https://doi.org/10.1016/j.neuropsychologia.2012.06.001>
- Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account.

- Memory & Cognition*, 47(1), 182–193. <https://doi.org/10.3758/s13421-018-0858-9>
- Gardiner, J. M. (2000). On the objectivity of subjective experiences of autonoetic and noetic consciousness. In E. Tulving (Ed.), *Memory, consciousness, and the brain: The Tallinn Conference* (pp. 159–172). Psychology Press.
- Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 50(1), 114–122. <https://doi.org/10.1037/1196-1961.50.1.114>
- Gardiner, J. M., & Ramponi, C. (1998). Experiences of Remembering, Knowing, and Guessing. *CONSCIOUSNESS AND COGNITION*, 7, 1–26.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, 10(2), 83–98.
- Geraci, L., McCabe, D. P., & Guillory, J. J. (2009). On interpreting the relationship between rememberKnow judgments and confidence: The role of instructions. *Consciousness and Cognition*, 18(3), 701–709. <https://doi.org/10.1016/j.concog.2009.04.010>
- Harlow, I. M., MacKenzie, G., & Donaldson, D. I. (2010). Familiarity for associations? A test of the domain dichotomy theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1381–1388. <https://doi.org/10.1037/a0020610>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 27(1), 46–51. <https://doi.org/10.3758/BF03203619>
- Herzmann, G., Minor, G., & Curran, T. (2018). Neural evidence for the contribution of holistic processing but not attention allocation to the other-race effect on face memory. *Cognitive, Affective, & Behavioral Neuroscience*, 18(5), 1015–1033. <https://doi.org/10.3758/s13415-018-0583-0>

018-0619-z

- Higham, P. A., & Vokey, J. R. (2004). Illusory Recollection and DualProcess Models of Recognition Memory. *The Quarterly Journal of Experimental Psychology Section A*, 57(4), 714–744. <https://doi.org/10.1080/02724980343000468>
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, 36(7), 1351–1359. <https://doi.org/10.3758/MC.36.7.1351>
- Hudon, C., Belleville, S., & Gauthier, S. (2009). The assessment of recognition memory using the Remember/Know procedure in amnestic mild cognitive impairment and probable Alzheimer's disease. *Brain and Cognition*, 9.
- Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 325–339. <https://doi.org/10.1037/a0025483>
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541. [https://doi.org/10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F)
- Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Mahwah, NJ: Erlbaum.
- Koen, J. D., & Yonelinas, A. P. (2014). *The Effects of Healthy Aging, Amnestic Mild Cognitive Impairment, and Alzheimer's Disease on Recollection and Familiarity: A Meta-Analytic Review*. 41.
- Kurilla, B. P., & Westerman, D. L. (2008). Processing fluency affects subjective claims of recollection. *Memory & Cognition*, 36(1), 82–92. <https://doi.org/10.3758/MC.36.1.82>
- Larsson, M., Öberg, C., & Bäckman, L. (2006). Recollective experience in odor recognition: Influences of adult age and familiarity. *Psychological Research Psychologische Forschung*,

- 70(1), 68–75. <https://doi.org/10.1007/s00426-004-0190-9>
- Lombardi, M. G., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2016). Forgetting of the recollection and familiarity components of recognition in patients with amnestic mild cognitive impairment. *Journal of Neuropsychology*, 12(2), 231–247. <https://doi.org/10.1111/jnp.12114>
- Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 3(22), 470. <https://doi.org/10.21105/joss.00470>
- Martins, C. A. R., & Lloyd-Jones, T. J. (2006). Preserved Conceptual Priming in Alzheimer's Disease. *Cortex*, 42(7), 995–1004. [https://doi.org/10.1016/S0010-9452\(08\)70205-3](https://doi.org/10.1016/S0010-9452(08)70205-3)
- Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11(3), 126–135. <https://doi.org/10.1016/j.tics.2006.12.003>
- McBride, D. M., & Anne Dosher, B. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition*, 11(3), 423–460. [https://doi.org/10.1016/S1053-8100\(02\)00007-7](https://doi.org/10.1016/S1053-8100(02)00007-7)
- Meade, M. E., Ahmad, M., & Fernandes, M. A. (2019). Drawing pictures at encoding enhances memory in healthy older adults and in individuals with probable dementia. *Aging, Neuropsychology, and Cognition*, 27(6), 880–901. <https://doi.org/10.1080/13825585.2019.1700899>
- Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition*, 21(3), 1435–1455. <https://doi.org/10.1016/j.concog.2012.04.014>
- Moreno-Martínez, F. J., & Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLoS ONE*, 7(5), e37527. <https://doi.org/10.1371/journal.pone.0037527>
- O'Connor, M. K., & Ally, B. A. (2010). Using stimulus form change to understand memorial

- familiarity for pictures and words in patients with mild cognitive impairment and Alzheimer's disease. *Neuropsychologia*, 48(7), 2068–2074. <https://doi.org/10.1016/j.neuropsychologia.2010.03.027>
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
- Paivio, A. (1972). Symbolic and sensory modalities of memory. In M. E. Meyer (Ed.), *The third Western symposium on learning: Cognitive Learning*. Western Washington State College.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pitarque, A. (2016). *The effects of healthy aging, amnestic mild cognitive impairment, and Alzheimer's disease on recollection, familiarity and false recognition, estimated by an associative process-dissociation recognition procedure*. 7.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21(1), 89–102. <https://doi.org/10.3758/BF03211168>
- Rajaram, S. (1996a). Perceptual Effects on Remembering: Recollective Processes in Picture Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(2), 365–377.
- Rajaram, S. (1996b). *Perceptual Effects on Remembering: Recollective Processes in Picture Recognition Memory*. 13.
- R Core Team. (2020). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria: R Foundation for Statistical Computing.
- Rollins, L., & Riggins, T. (2018). Age-related differences in subjective recollection: ERP studies of encoding and retrieval. *Developmental Science*, 21(3), e12583. <https://doi.org/10.1111/desc.12583>
- Rosson, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set:

- The Role of Surface Detail in Basic-Level Object Recognition. *Perception*, 33(2), 217–236. <https://doi.org/10.1068/p5117>
- Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, auto-biographical, and fictional events. *Psychological Review*, 122(1), 1–23. <https://doi.org/10.1037/a0037907>
- Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, 28(4), 537–555. <https://doi.org/10.3758/BF03200541>
- Scalici, F., Caltagirone, C., & Carlesimo, G. A. (2017). The contribution of different prefrontal cortex regions to recollection and familiarity: A review of fMRI data. *Neuroscience & Biobehavioral Reviews*, 83, 240–251. <https://doi.org/10.1016/j.neubiorev.2017.10.017>
- Schmitter-Edgecombe, M., Woo, E., & Greeley, D. R. (2009). Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*, 23(2), 168–177. <https://doi.org/10.1037/a0014186>
- Schoemaker, D., Gauthier, S., & Pruessner, J. C. (2014). Recollection and Familiarity in Aging Individuals with Mild Cognitive Impairment and Alzheimer's Disease: A Literature Review. *Neuropsychol Rev*, 19.
- Serra, L., Bozzali, M., Cercignani, M., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2010). Recollection and familiarity in amnesic mild cognitive impairment. *Neuropsychology*, 24(3), 316–326. <https://doi.org/10.1037/a0017654>
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156–163. [https://doi.org/10.1016/S0022-5371\(67\)80067-7](https://doi.org/10.1016/S0022-5371(67)80067-7)
- Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name

- Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, 8(11), 872–883. <https://doi.org/10.1038/nrn2154>
- Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, 18(6), 813–847. <https://doi.org/10.1080/09541440500412361>
- Suzuki, K., & Takahashi, R. (1997). Effectiveness of color in picture recognition memory. *Japanese Psychological Research*, 39(1), 25–32. <https://doi.org/10.1111/1468-5884.00033>
- Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, 35(4), 621–633. <https://doi.org/10.3758/BF03195542>
- Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *TRENDS in Cognitive Sciences*, 5(5), 211–215.
- Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-20).
- Tousignant, C., & Bodner, G. E. (2012). Test context affects recollection and familiarity ratings: Implications for measuring recognition experiences. *Consciousness and Cognition*, 21(2), 994–1000. <https://doi.org/10.1016/j.concog.2012.01.009>
- Tousignant, C., Bodner, G. E., & Arnold, M. M. (2015). Effects of context on recollection and familiarity experiences are task dependent. *Consciousness and Cognition*, 33, 78–89. <https://doi.org/10.1016/j.concog.2014.11.011>
- Troyer, A. K., Murphy, K. J., Anderson, N. D., Craik, F. I. M., Moscovitch, M., Maione, A.,

- & Gao, F. (2012). Associative recognition in mild cognitive impairment: Relationship to hippocampal volume and apolipoprotein E. *Neuropsychologia*, 50(14), 3721–3728. <https://doi.org/10.1016/j.neuropsychologia.2012.10.018>
- Troyer, A. K., Vandermorris, S., & Murphy, K. J. (2016). Intraindividual variability in performance on associative memory tasks is elevated in amnestic mild cognitive impairment. *Neuropsychologia*, 90, 110–116. <https://doi.org/10.1016/j.neuropsychologia.2016.06.011>
- Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*, 43(4), 1085–1099. <https://doi.org/10.3758/s13428-011-0121-9>
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, 26(1), 1–12. <https://doi.org/10.1037/h0080017>
- Tunney, R. J., & Fernie, G. (2007). Repetition priming affects guessing not familiarity. *Behavioral and Brain Functions*, 3(1), 40. <https://doi.org/10.1186/1744-9081-3-40>
- Umanath, S., & Coane, J. H. (2020). Face Validity of Remembering and Knowing: Empirical Consensus and Disagreement Between Participants and Researchers. *Perspectives on Psychological Science*, 15(6), 1400–1422. <https://doi.org/10.1177/1745691620917672>
- van der Meulen, M., Lederrey, C., Rieger, S. W., van Assche, M., Schwartz, S., Vuilleumier, P., & Assal, F. (2012). Associative and Semantic Memory Deficits in Amnestic Mild Cognitive Impairment as Revealed by Functional Magnetic Resonance Imaging: *Cognitive and Behavioral Neurology*, 25(4), 195–215. <https://doi.org/10.1097/WNN.0b013e31827de67f>
- Viggiano, M. P., Vannucci, M., & Righi, S. (2004). A New Standardized Set of Ecological Pictures for Experimental and Clinical Research on Visual Object Processing. *Cortex*, 40(3), 491–509. [https://doi.org/10.1016/S0010-9452\(08\)70142-4](https://doi.org/10.1016/S0010-9452(08)70142-4)
- Wagner, A. D., Gabrieli, J. D. E., & Verfaellie, M. (1997). *Dissociations Between Familiarity*

- Processes in Explicit Recognition and Implicit Perceptual Memory.* 19.
- Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, 69(9), 1752–1776. <https://doi.org/10.1080/17470218.2015.1094494>
- Wang, P., Li, J., Li, H., Li, B., Yang Jiang, Bao, F., & Zhang, S. (2013). Is emotional memory enhancement preserved in amnestic mild cognitive impairment? Evidence from separating recollection and familiarity. *Neuropsychology*, 27(6), 691–701. <https://doi.org/10.1037/a0033973>
- Wang, Li, H., Liang, Y., Zhang, J., Li, X., Shu, N., ... Zhang, Z. (2013). Amnestic Mild Cognitive Impairment: Topological Reorganization of the Default-Mode Network. *Radiology*, 268(2), 501–514. <https://doi.org/10.1148/radiol.13121573>
- Weldon, M. S., Iii, H. L. R., & Challis, B. H. (1989). *The properties of retrieval cues constrain the picture superiority effect.* 11.
- Weldon, M. S., & Roediger, H. L. (1987). *Altering retrieval demands reverses the picture superiority effect.* 12.
- Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M.-M., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology*, 20(2), 193–205. <https://doi.org/10.1037/0894-4105.20.2.193>
- Westerberg, C., Mayes, A., Florczak, S. M., Chen, Y., Creery, J., Parrish, T., ... Paller, K. A. (2013). Distinct medial temporal contributions to different forms of recognition in amnestic mild cognitive impairment and Alzheimer's disease. *Neuropsychologia*, 51(12), 2450–2461. <https://doi.org/10.1016/j.neuropsychologia.2013.06.025>
- Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, 24(4), 767–773. <https://doi.org/10.1080/07475620600713117>

- org/10.1348/026151005X74153
- Williams, H. L. (2019). *Different definitions of the nonrecollection-based response option(s) change how people use the “remember” response in the remember/know paradigm.* 16.
- Williams, H. L., & Moulin, C. J. A. (2014). Know versus Familiar: Differentiating states of awareness in others' subjective reports of recognition. *Memory*, 23(7), 981–990. <https://doi.org/10.1080/09658211.2014.945460>
- Wixted, J. T. (2014). Signal detection theory. In *Wiley StatsRef: Statistics reference online*. American Cancer Society. <https://doi.org/10.1002/9781118445112.stat06743>
- Wolk, D. A., Dunfee, K. L., Dickerson, B. C., Aizenstein, H. J., & DeKosky, S. T. (2011). A medial temporal lobe division of labor: Insights from memory in aging and early Alzheimer disease. *Hippocampus*, 21(5), 461–466. <https://doi.org/10.1002/hipo.20779>
- Wolk, D. A., Mancuso, L., Kliot, D., Arnold, S. E., & Dickerson, B. C. (2013). Familiarity-based memory as an early cognitive marker of preclinical and prodromal AD. *Neuropsychologia*, 51(6), 1094–1102. <https://doi.org/10.1016/j.neuropsychologia.2013.02.014>
- Wolk, D. A., Signoroff, E. D., & DeKosky, S. T. (2008). Recollection and familiarity in amnestic mild cognitive impairment: A global decline in recognition memory. *Neuropsychologia*, 46(7), 1965–1978. <https://doi.org/10.1016/j.neuropsychologia.2008.01.017>
- Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, 46(3), 441–517. <https://doi.org/10.1006/jmla.2002.2864>
- Yonelinas, & Jacoby. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, Volume 34(5), 622–643. <https://doi.org/10.1006/jmla.1995.1028>
- Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H.-Y. M., ... Park, D. C. (2004). A cross-culturally standardized set of pictures for younger and older adults: Amer-

ican and Chinese norms for name agreement, concept agreement, and familiarity. *Behavior Research Methods, Instruments, & Computers*, 36(4), 639–649. <https://doi.org/10.3758/BF03206545>

–drawing-word.png –photo-word.png

Appendices

Appendix A: Spelling corrections / manipulations to naming responses.

Response	Correction	Response	Correction	Response	Correction
;ashtray	ashtray	draw	drawers	lobster	lobster
a	no	drawer	drawers	onions	onion
acoop	scoop	draws	drawers	onions	onion
ancher	anchor	drums	drum	osterich	ostrich
anchore	anchor	eagal	eagle	ostrage	ostrich
ancor	anchor	eclipes	eclipse	ostridge	ostrich
anker	anchor	eclipses	eclipse	ostrigie	ostrich
aparagus	asparagus	eclipsse	eclipse	ostrisge	ostrich
apricorte	apricot	eeagle	eagle	ostritch	ostrich
ashtry	ashtray	eele	seal	pair	pear
ballon	balloon	eyeglass	eyeglasses	paper	pepper
ballone	balloon	falg	flag	peacck	peacock
balloone	balloon	feet	foot	pecock	peacock
ballun	balloon	fencing	fence	peguin	penguin
baloon	balloon	footstall	footstool	peneut	peanut
bamnna	banana	fott	foot	pengiuin	penguin
bananaa	banana	frock	frog	pengiun	penguin
bananna	banana	frog/	frog	penguine	penguin
bannan	banana	geese	goose	penquin	penguin
bannana	banana	giaffee	giraffe	peper	pepper
barel	barrel	giaraffe	giraffe	pestleandmorter	pestleandmortar

(continued)

Response	Correction	Response	Correction	Response	Correction
barrell	barrel	girafe	giraffe	piccalo	piccolo
barrele	barrel	giraff	giraffe	pilers	pliers
barrow	barrel	giraffee	giraffe	piars	pliers
beatle	beetle	girafffe	giraffe	plier	pliers
beer	bear	girrafe	giraffe	pliyers	pliers
bellpepp34	bellpepper	girraffe	giraffe	plugin	plug
bettle	beetle	gitaur	guitar	plyers	pliers
bicucle	bicycle	gitter	guitar	potatoe	potato
bicyle	bicycle	glases	glasses	pottato	potato
bittle	bottle	glass	glasses	pumkin	pumpkin
bittle	beetle	glassesbottle	bottle	pumpkim	pumpkin
blueberrys	blueberries	gloves	glove	pumpkin	pumpkin
bolw	bowl	grape	grapes	rabit	rabbit
bootle	bottle	gutair	guitar	racoon	raccoon
broon	broom	haircomb	comb	rubarb	rhubarb
broon	broom	hamp	harp	rule	ruler
brum	broom	hand5	hand	seel	seal
busket	basket	harper	hamper	showel	shovel
bycycle	bicycle	hemmar	hammer	snakw	snake
camal	camel	hose	house	soak	socks
canddle	candle	idk	no	specs	spectacles
canle	candle	kacket	jacket	spon	spoon
canon	cannon	kangroo	kangaroo	ssnowman	snowman
carott	carrot	ladders	ladder	steplader	stepladder
carrots	carrot	lader	ladder	sterss	step
carrott	carrot	latter	ladder	sweetcirl	sweetcorn
celary	celery	leafe	leaf	tabaccopipe	tobaccopipe

(continued)

Response	Correction	Response	Correction	Response	Correction
celeary	celery	leamon	lemon	teakettle	kettle
cellary	celery	leema	lemur	thinbell	thimble
cerlery	celery	lettace	lettuce	thmble	thimble
chain2	chain	lip	lips	thunb	thumb
chestofdrawerserss	chest of drawers	longdress	dress	timbil	thimble
chestofdraws	chest of drawers	maledear	maledeer	timble	thimble
chisle	chisel	meercat	meerkat	toitouse	tortoise
chissel	chisel	mercat	meerkat	tomatoe	tomato
chizel	chisel	mice	mouse	tomoato	tomato
claranet	clarinet	mit	mitten	tortise	tortoise
clouds	cloud	mittens	mitten	tortiste	tortoise
cochroach	cockroach	monkeybut	monkeynut	tortus	tortoise
cock	cockerel	mortle	mortar	usplug	plug
cockaroach	cockroach	mousse	moose	vulture	vulture
cockrel	cockerel	muscat	muskrat	vicescripts	vicegrips
combe	comb	nectarin	nectarine	violin	violin
cycle	bicycle	nectarinee	nectarine	violen	violin
dear	deer	nectrine	nectarine	volion	violin
deere	deer	neddle	needle	waistcoast	waistcoat
dock	duck	noise	nose	wale	well
dolly	doll	none	no	wastecost	waistcoat
doormouse	dormouse	oencil	pencil	whisell	whistle
				whistel	whistle
				whitle	whistle
				whsitle	whistle
				windown	window

Appendix B: Normative data for all photograph items.

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
anchor				
anchor1-photo-colour	3.55 (1.32)	3 (1.11)	2.77 (1.19)	4.05 (1.05)
anchor1-photo-grey	3.03 (1.47)	2.81 (0.93)		3.95 (0.89)
anchor2-photo-colour	3.38 (1.56)	3.45 (1.23)	3.5 (1.19)	3.75 (1.25)
anchor2-photo-grey	2.82 (1.53)	3 (0.97)		3.52 (1.33)
anchor3-photo-colour	3.5 (1.3)	2.48 (0.93)	2.76 (1.3)	3.86 (1.01)
anchor3-photo-grey	3.04 (1.43)	2.46 (1.22)		4.27 (0.77)
apple				
apple1-photo-colour	4.8 (0.48)	3.1 (1.14)	3.43 (1.08)	4.15 (0.88)
apple1-photo-grey	4.85 (0.49)	2.32 (1.21)		3.05 (1.23)
apple2-photo-colour	4.91 (0.29)	3 (0.97)	3.5 (1.32)	3.76 (1.04)
apple2-photo-grey	4.24 (1.14)	2.4 (0.99)		2.75 (1.45)
apple3-photo-colour	4.77 (0.53)	3.16 (1.37)	3.68 (1.25)	4.27 (0.98)
apple3-photo-grey	4.73 (0.63)	2.38 (1.07)		3.38 (1.28)
ashtray				
ashtray1-photo-colour	3.77 (1.43)	3.05 (1.16)	2.71 (1.1)	3.9 (1.07)
ashtray1-photo-grey	4.15 (1.14)	2.95 (1.13)		3.5 (1.28)
ashtray2-photo-colour	3.86 (1.46)	3 (1.08)	2.25 (1.29)	4.1 (1.3)
ashtray2-photo-grey	3.52 (1.47)	3 (1.12)		3.5 (1.28)
ashtray3-photo-colour	3.36 (1.47)	3.75 (1.15)	3.21 (1.32)	3.5 (1.19)
ashtray3-photo-grey	4.14 (0.99)	3.09 (1.11)		3.76 (1)
balloon				
balloon1-photo-colour	4.4 (1.1)	1.63 (1)	2.35 (1.69)	4.62 (0.92)
balloon1-photo-grey	4.15 (1.18)	1.8 (1.11)		3.41 (1.1)
balloon2-photo-colour	4.45 (1)	2.14 (1.08)	1.82 (1.3)	4.5 (1)
balloon2-photo-grey	4.35 (0.81)	1.9 (1)		4.05 (0.94)
balloon3-photo-colour	4.09 (1.02)	1.86 (1.04)	1.86 (1.49)	4.46 (0.88)
balloon3-photo-grey	4.24 (0.94)	1.68 (0.99)		3.57 (1.4)
banana				
banana1-photo-colour	4.65 (0.99)	2.55 (1.5)	4.55 (0.76)	4.45 (0.86)
banana1-photo-grey	4.8 (0.7)	2.23 (1.1)	2.36 (1.5)	3.76 (0.94)
banana2-photo-colour	4.8 (0.41)	2.05 (1.24)	4 (1.3)	4.85 (0.37)
banana2-photo-grey	4.9 (0.45)	2.36 (1.36)		3.3 (1.42)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
banana3-photo-colour	4.33 (1.02)	2.05 (1.09)	4.59 (0.96)	4.67 (0.66)
banana3-photo-grey	4.86 (0.47)	2 (0.87)		3.58 (1.1)
barrel				
barrel1-photo-colour	3.53 (1.25)	3.57 (0.93)	3.95 (0.97)	4.9 (0.45)
barrel1-photo-grey	3.9 (1.29)	3 (1.02)		4.05 (1.05)
barrel2-photo-colour	4 (1.02)	3 (1.08)	3.5 (1.15)	4.43 (0.81)
barrel2-photo-grey	3.81 (1.54)	2.95 (1.1)		4.15 (0.93)
barrel3-photo-colour	3.5 (1.26)	3.38 (1.17)	2.96 (1.3)	4.22 (0.95)
barrel3-photo-grey	3.45 (1.5)	2.68 (1.13)		4 (0.95)
basket				
basket1-photo-colour	4.13 (1.04)	3.86 (0.96)	3.62 (1.24)	4.3 (1.13)
basket1-photo-grey	4.5 (0.83)	3.36 (1.09)		3.85 (1.14)
basket2-photo-colour	4.14 (1.08)	3.05 (1)	2.8 (1.36)	3.95 (0.97)
basket2-photo-grey	4.48 (0.93)	2.9 (1.45)		3.55 (1.23)
basket3-photo-colour	4.27 (0.83)	3.8 (1.08)	3.48 (1.53)	3.64 (1.14)
basket3-photo-grey	4.45 (0.86)	3.18 (1.1)		3.95 (0.92)
bear				
bear1-photo-colour	3.36 (1.53)	3.9 (0.79)	3.9 (0.72)	4.15 (1.09)
bear1-photo-grey	3.9 (1.34)	3.5 (1.15)		3.93 (1.08)
bear2-photo-colour	3.85 (1.53)	3.25 (0.91)	4.3 (0.73)	4.24 (1.22)
bear2-photo-grey	3.25 (1.45)	3.25 (1.16)		4.27 (1.12)
bear3-photo-colour	3.9 (1.37)	3.81 (1.25)	3.71 (1.1)	4.09 (0.87)
bear3-photo-grey	3.75 (1.48)	3 (1.02)		4 (1.23)
beetle				
beetle1-photo-colour	3.1 (1.59)	4 (0.79)	3.15 (1.5)	2.95 (1.07)
beetle1-photo-grey	3.1 (1.45)	3.1 (1.21)		2.95 (1.21)
beetle2-photo-colour	3.2 (1.4)	3.73 (1.16)	3 (1.38)	2.95 (1.19)
beetle2-photo-grey	3 (1.34)	2.95 (1.24)		2.7 (1.13)
beetle3-photo-colour	2.68 (1.25)	3.41 (0.96)	3.05 (1.13)	3.17 (1.05)
beetle3-photo-grey	3 (1.26)	3.18 (1.05)		3.24 (1.04)
bell				
bell1-photo-colour	4 (1.21)	2.47 (1.41)	3.55 (1.39)	4 (1.05)
bell1-photo-grey	3.8 (1.28)	2.45 (1.05)		3.64 (0.95)
bell2-photo-colour	3.9 (1.14)	3.32 (0.95)	3.41 (1.3)	3.95 (1.1)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
bell2-photo-grey	3.5 (1.15)	2.24 (0.62)		3.3 (0.86)
bell3-photo-colour	3.43 (1.41)	3.09 (1.02)	2.91 (1.34)	3.5 (1.06)
bell3-photo-grey	3.62 (1.47)	2.82 (0.96)		3.05 (1.36)
belt				
belt1-photo-colour	4.8 (0.52)	2.86 (1.21)	2.68 (1.46)	3.95 (0.83)
belt1-photo-grey	4.6 (0.89)	3.05 (0.97)		4 (0.92)
belt2-photo-colour	4.71 (0.72)	3 (1.38)	2.85 (1.46)	4.05 (1.05)
belt2-photo-grey	4.64 (0.79)	2.4 (1.1)		4.48 (0.87)
belt3-photo-colour	4.86 (0.64)	2.43 (1.29)	3.05 (1.43)	4.24 (0.77)
belt3-photo-grey	4.35 (1.03)	2.21 (1.28)		3.95 (1.09)
bicycle				
bicycle1-photo-colour	4.3 (1.13)	3.55 (1.1)	2.45 (1.47)	3.14 (1.17)
bicycle1-photo-grey	4.5 (1)	3.6 (1.04)	1.73 (1.19)	3.33 (1.2)
bicycle2-photo-colour	4.65 (0.67)	2.95 (1.12)	1.57 (0.98)	4.05 (0.94)
bicycle2-photo-grey	4.76 (0.62)	3.41 (1.05)		3.65 (0.93)
bicycle3-photo-colour	3.86 (1.15)	3.27 (0.98)	1.91 (1.27)	3.38 (1.16)
bicycle3-photo-grey	4.05 (1.05)	3.48 (0.99)		3.33 (1.46)
book				
book1-photo-colour	4.85 (0.37)	3.15 (1.23)	2.4 (1.27)	3.59 (1.14)
book1-photo-grey	4.9 (0.45)	2.7 (0.99)	1.45 (0.93)	3.62 (0.8)
book2-photo-colour	4.75 (0.72)	3.1 (1.09)	1.71 (1.06)	3.65 (1.35)
book2-photo-grey	4.75 (0.55)	2.86 (1.08)		3.2 (1.15)
book3-photo-colour	4.33 (0.91)	3.05 (1.05)	2.27 (1.24)	3 (1.14)
book3-photo-grey	4.04 (1.36)	2.45 (0.51)		3 (1.18)
boot				
boot1-photo-colour	4.15 (1.18)	2.95 (1.05)	2.5 (1.47)	3.27 (1.2)
boot1-photo-grey	4.7 (0.47)	2.93 (1.36)	2.82 (1.6)	2.76 (0.94)
boot2-photo-colour	4.6 (0.6)	2.95 (1.2)	2.1 (0.94)	4.25 (0.85)
boot2-photo-grey	4.7 (0.66)	3.41 (1.01)		3.75 (1.02)
boot3-photo-colour	4.29 (0.96)	2.95 (1.09)	2.45 (1.44)	4.1 (1)
boot3-photo-grey	4.52 (0.67)	2.95 (1)		3.79 (1.22)
bottle				
bottle1-photo-colour	4.57 (0.73)	3.33 (1.35)	3.14 (1.2)	3.7 (0.98)
bottle1-photo-grey	4.85 (0.37)	2.45 (1.1)		3.15 (0.99)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
bottle2-photo-colour	4.82 (0.5)	2.25 (1.07)	2.25 (1.21)	2.81 (1.5)
bottle2-photo-grey	4.48 (1.12)	1.85 (0.93)		3.2 (1.11)
bottle3-photo-colour	4.45 (0.91)	1.92 (1.28)	2.04 (1.23)	2.55 (1.18)
bottle3-photo-grey	4.5 (0.96)	1.5 (1.01)		3.33 (1.43)
bowl				
bowl1-photo-colour	4.68 (0.78)	2.25 (1.16)	1.65 (0.99)	3.7 (1.13)
bowl1-photo-grey	4.81 (0.51)	2 (0.86)		3.03 (1)
bowl2-photo-colour	4.8 (0.41)	2.5 (1.1)	1.9 (1.55)	3.43 (1.16)
bowl2-photo-grey	4.75 (0.44)	1.9 (1.45)		4.05 (1.09)
bowl3-photo-colour	4.62 (0.67)	2.19 (1.25)	1.9 (1.45)	3.27 (1.28)
bowl3-photo-grey	4.52 (0.92)	1.73 (0.98)		3.09 (1.06)
bread				
bread1-photo-colour	4.82 (0.66)	3.45 (0.94)	3.85 (1.04)	3.8 (1.15)
bread1-photo-grey	4.52 (0.87)	2.8 (1.06)		3.07 (1.01)
bread2-photo-colour	5 (0)	3.2 (1.24)	3.75 (1.02)	3.81 (1.21)
bread2-photo-grey	4.9 (0.31)	2.35 (1.31)		3.36 (1.18)
bread3-photo-colour	4.71 (0.64)	3.48 (1.36)	3.29 (1.27)	3.41 (1.37)
bread3-photo-grey	4.67 (0.64)	2.59 (0.96)		3.26 (1.18)
broom				
broom1-photo-colour	4.09 (1.19)	2.3 (1.08)	2.55 (1.19)	3.35 (1.35)
broom1-photo-grey	4.29 (1.01)	2.55 (1.05)		3.27 (1.14)
broom2-photo-colour	4.2 (1.15)	2.85 (1.09)	3.3 (1.3)	3.14 (1.46)
broom2-photo-grey	4.05 (0.76)	2.1 (1.02)		3.41 (1.37)
broom3-photo-colour	4 (1)	2.38 (1.02)	2.67 (1.28)	3.73 (1.45)
broom3-photo-grey	4.12 (1.12)	2.27 (1.24)		3.32 (1.36)
brush				
brush1-photo-colour	4.33 (0.91)	3 (1.17)	2.15 (1.42)	2.27 (1.36)
brush1-photo-grey	4.41 (0.8)	2.8 (0.83)		2.85 (1.39)
brush2-photo-colour	4.1 (0.97)	3.5 (1.24)	3.65 (1.23)	2.18 (1.22)
brush2-photo-grey	4 (0.97)	3.2 (1.11)		2.24 (1.3)
brush3-photo-colour	4.38 (0.97)	3.13 (1.18)	2.87 (1.42)	2.64 (1.29)
brush3-photo-grey	4 (0.95)	3.14 (1.06)		2.55 (1.22)
button				
button1-photo-colour	4.45 (1)	2.77 (1.27)	1.91 (1.23)	3.05 (1.54)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
button1-photo-grey	4.57 (0.77)	2.33 (1.02)		4.05 (1.23)
button2-photo-colour	4.71 (0.78)	1.55 (0.89)	1.45 (0.83)	4.45 (1.1)
button2-photo-grey	4.82 (0.5)	1.5 (0.83)		4.24 (1.34)
button3-photo-colour	4.68 (0.57)	2.14 (1.46)	1.77 (1.15)	3.67 (1.28)
button3-photo-grey	4.55 (0.74)	2 (1.5)		3.55 (1.14)
cake				
cake1-photo-colour	4.62 (0.59)	3.35 (1.04)	2.5 (1.57)	3.27 (1.01)
cake1-photo-grey	4.73 (0.63)	3.9 (0.91)		3.35 (1.27)
cake2-photo-colour	4.5 (0.69)	4.4 (0.99)	2.6 (1.93)	3.68 (1.21)
cake2-photo-grey	4.7 (0.57)	4 (0.92)		2.62 (1.2)
cake3-photo-colour	4.79 (0.41)	3.7 (0.97)	2.91 (1.2)	3.14 (1.32)
cake3-photo-grey	4.19 (1.03)	2.9 (1)		2.55 (1.1)
camel				
camel1-photo-colour	3.4 (1.57)	3.6 (1.19)	3.75 (1.21)	4.18 (1.14)
camel1-photo-grey	3.65 (1.6)	3.8 (1.03)	2.36 (1.5)	3.95 (1.07)
camel2-photo-colour	3.9 (1.17)	3.52 (1.44)	3.95 (1.16)	3.85 (0.93)
camel2-photo-grey	3.3 (1.72)	3.82 (1.01)		3.35 (0.99)
camel3-photo-colour	3.29 (1.42)	3.5 (1.01)	4.41 (0.96)	4.29 (0.9)
camel3-photo-grey	2.87 (1.55)	3.18 (1.01)		3.92 (1.14)
candle				
candle1-photo-colour	4.6 (0.68)	2.33 (0.92)	3 (1.21)	3.67 (1.06)
candle1-photo-grey	4.45 (0.89)	1.8 (0.89)		2.95 (1.09)
candle2-photo-colour	4.71 (0.64)	2.59 (1.01)	2.45 (1.5)	3.65 (1.14)
candle2-photo-grey	4.35 (0.93)	1.76 (1)		3.8 (1.15)
candle3-photo-colour	3.5 (1.06)	2.95 (1)	2.55 (1.14)	3 (0.98)
candle3-photo-grey	4.14 (0.79)	2.41 (0.85)		2.62 (1.28)
cannon				
cannon1-photo-colour	3.2 (1.51)	3.32 (1.13)	3.32 (1.21)	3.7 (0.92)
cannon1-photo-grey	3.1 (1.49)	3.43 (0.87)		3.8 (1.36)
cannon2-photo-colour	3.9 (1.34)	3.3 (1.03)	3.75 (1.02)	3.95 (0.69)
cannon2-photo-grey	3.45 (1.44)	2.8 (0.95)		3.71 (1.19)
cannon3-photo-colour	3.32 (1.59)	3.55 (1.1)	2.91 (1.34)	3.1 (1.26)
cannon3-photo-grey	3 (1.38)	3.24 (1.13)		2.59 (1.5)
carrot				

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
carrot1-photo-colour	4.91 (0.29)	3 (1.12)	4.15 (1.27)	4.25 (0.85)
carrot1-photo-grey	4.57 (0.6)	2.25 (0.97)		2.87 (1.04)
carrot2-photo-colour	4.9 (0.31)	3.1 (1.17)	4.6 (0.68)	4.1 (1.26)
carrot2-photo-grey	4.75 (0.55)	2.35 (0.99)		3.55 (1.1)
carrot3-photo-colour	4.71 (0.56)	3.52 (1.21)	4.19 (0.98)	3.91 (1.11)
carrot3-photo-grey	4.6 (0.58)	3.23 (1.07)		3.64 (1.09)
celery				
celery1-photo-colour	4.4 (0.99)	2.33 (1.35)	4.5 (0.89)	4.1 (1)
celery1-photo-grey	4 (1.38)	1.95 (1.23)		2.91 (1.23)
celery2-photo-colour	4.15 (1.27)	3.32 (0.99)	4.55 (0.91)	3.9 (1.12)
celery2-photo-grey	3.9 (1.12)	3 (1.26)		2.75 (1.21)
celery3-photo-colour	3.22 (1.54)	3 (1.02)	4.27 (1.28)	4.04 (1.16)
celery3-photo-grey	3.71 (1.38)	3.14 (1.08)		2.86 (1.35)
chain				
chain1-photo-colour	4.1 (1.25)	2.3 (1.12)	3.1 (1.37)	3.67 (1.15)
chain1-photo-grey	3.55 (1.39)	2.1 (1.17)		3.27 (1.55)
chain2-photo-colour	4.2 (0.89)	2.55 (1.1)	2.82 (1.26)	3.6 (1.35)
chain2-photo-grey	3.75 (1.16)	1.9 (0.89)		4.55 (0.69)
chain3-photo-colour	3.77 (1.41)	2.7 (0.88)	3.39 (1.27)	4.08 (1.21)
chain3-photo-grey	3.67 (1.39)	2.32 (1.09)		3.86 (1.11)
chair				
chair1-photo-colour	4.76 (0.77)	3.15 (1.14)	2.9 (1.45)	3.53 (1.28)
chair1-photo-grey	4.64 (0.66)	3.1 (0.72)		3.4 (1.31)
chair2-photo-colour	4.6 (0.75)	3.1 (1.25)	3 (1.56)	3.77 (1.23)
chair2-photo-grey	4.95 (0.22)	2.7 (1.08)		3.38 (1.36)
chair3-photo-colour	5 (0)	2.13 (1.18)	3.04 (1.33)	4.32 (0.89)
chair3-photo-grey	4.76 (0.54)	2.57 (0.98)		3.41 (1.22)
cherry				
cherry1-photo-colour	4.4 (0.94)	2.55 (1.32)	4.3 (0.8)	4.18 (1.18)
cherry1-photo-grey	4.2 (1.11)	2.27 (1.11)	1.73 (1.1)	3.48 (1.03)
cherry2-photo-colour	4.3 (0.8)	2.19 (1.36)	4.05 (1.2)	4.7 (0.57)
cherry2-photo-grey	4.45 (0.83)	2.36 (1)		3.25 (1.25)
cherry3-photo-colour	4.14 (1.06)	2.59 (1.18)	4.41 (0.96)	4.48 (0.75)
cherry3-photo-grey	3.87 (1.18)	2.23 (0.97)		3.62 (1.06)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
chicken				
chicken1-photo-colour	4.33 (0.96)	4.14 (1.06)	3.81 (1.12)	3.85 (0.93)
chicken1-photo-grey	4.25 (1.12)	3.59 (1.14)		2.9 (1.17)
chicken2-photo-colour	4.36 (1.05)	4 (0.92)	3.35 (1.04)	4.43 (0.98)
chicken2-photo-grey	4.38 (0.86)	3.4 (1.47)		3.7 (1.13)
chicken3-photo-colour	4.14 (0.94)	4.29 (0.86)	3.79 (1.35)	3.59 (1.1)
chicken3-photo-grey	4 (1.11)	3.59 (1.26)		3.29 (1.15)
chisel				
chisel1-photo-colour	3.91 (1.27)	2.8 (1.15)	2.4 (1.31)	3.7 (1.17)
chisel1-photo-grey	3.86 (1.2)	2.45 (1.15)		3.33 (1.35)
chisel2-photo-colour	3.45 (1.36)	3.6 (1.1)	2.7 (1.45)	3.62 (1.4)
chisel2-photo-grey	2.6 (1.5)	2.5 (1.4)		3.91 (0.75)
chisel3-photo-colour	3.1 (1.58)	3.14 (1.11)	2.62 (1.2)	3.73 (1.16)
chisel3-photo-grey	2.62 (1.31)	2.27 (1.12)		3.59 (1.33)
clock				
clock1-photo-colour	5 (0)	3.18 (0.96)	2.82 (1.22)	3.55 (1.19)
clock1-photo-grey	4.73 (0.74)	3.24 (1)		4.2 (0.83)
clock2-photo-colour	4.71 (0.64)	3.15 (1.14)	3.2 (1.51)	2.9 (1.12)
clock2-photo-grey	4.77 (0.43)	2.6 (0.88)		2.9 (1.3)
clock3-photo-colour	4.82 (0.39)	3.24 (1.04)	2.38 (0.97)	3.24 (1.34)
clock3-photo-grey	4.36 (0.9)	3.42 (1.38)		2.59 (1.18)
cloud				
cloud1-photo-colour	4.2 (1.15)	2.9 (1.25)	3.15 (1.04)	4.05 (0.9)
cloud1-photo-grey	4.75 (0.91)	2.7 (1.26)	3.64 (1.29)	4 (1.14)
cloud2-photo-colour	4.55 (0.94)	2.86 (1.2)	2.9 (1.34)	3.15 (1.35)
cloud2-photo-grey	4.2 (1.2)	2.23 (1.15)		2.55 (1.36)
cloud3-photo-colour	4.48 (0.81)	1.77 (0.92)	4.32 (0.95)	3.67 (1.32)
cloud3-photo-grey	4.14 (1.17)	2.18 (1.1)		3.71 (1.04)
comb				
comb1-photo-colour	4.55 (0.74)	2.45 (1.23)	2.25 (1.48)	3.45 (1.23)
comb1-photo-grey	4.67 (0.73)	2.05 (1)		3.6 (1.25)
comb2-photo-colour	4.5 (0.89)	2.85 (1.27)	1.8 (1.32)	4 (0.89)
comb2-photo-grey	4.6 (0.75)	1.85 (1.04)		3.91 (0.87)
comb3-photo-colour	4.57 (0.6)	2.33 (1.06)	1.81 (1.33)	3.82 (1.22)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
comb3-photo-grey	4.54 (0.78)	1.87 (0.97)		4.23 (0.87)
corn				
corn1-photo-colour	4.27 (0.88)	3.8 (0.95)	4.6 (0.75)	4.25 (1.07)
corn1-photo-grey	4.71 (0.56)	2.9 (1.29)		3.4 (1.13)
corn2-photo-colour	4.55 (0.69)	3.5 (1.19)	4.55 (0.89)	4.29 (1.27)
corn2-photo-grey	4.45 (0.69)	3 (1.34)		4.18 (0.73)
corn3-photo-colour	4.43 (0.81)	3.57 (1.16)	4.62 (0.59)	4 (1.2)
corn3-photo-grey	4.42 (0.83)	3.05 (1.09)		4.09 (0.92)
crown				
crown1-photo-colour	3.77 (1.41)	4.57 (0.6)	4.38 (0.67)	4.2 (0.95)
crown1-photo-grey	4.2 (1.2)	4.27 (1.08)		3.4 (1.19)
crown2-photo-colour	3.91 (1.48)	4 (1.08)	3.1 (1.41)	4.19 (1.03)
crown2-photo-grey	3.95 (1.47)	3.5 (0.95)		3.55 (1.23)
crown3-photo-colour	3.27 (1.45)	4.56 (0.71)	4 (1.32)	3.14 (1.21)
crown3-photo-grey	3.45 (1.57)	3.68 (1.21)		3 (1.05)
deer				
deer1-photo-colour	3.55 (1.54)	3.55 (1.06)	3.64 (1.18)	3.15 (1.31)
deer1-photo-grey	3.47 (1.5)	3.29 (0.9)		3.65 (1.18)
deer2-photo-colour	3.43 (1.4)	3.8 (1.15)	4.4 (0.94)	3.5 (1)
deer2-photo-grey	3.36 (1.26)	3.3 (1.22)		3.1 (1.51)
deer3-photo-colour	3.45 (1.34)	3.64 (1.05)	3.64 (1.26)	3.86 (0.91)
deer3-photo-grey	3.45 (1.3)	3.42 (1.1)		3.09 (1.16)
doll				
doll1-photo-colour	4.4 (1.23)	2.91 (1.06)	2.73 (1.35)	2.95 (1)
doll1-photo-grey	3.8 (1.37)	3.05 (0.74)		3.2 (1.11)
doll2-photo-colour	4.29 (1.06)	4.05 (1.05)	2.8 (1.28)	3.35 (1.14)
doll2-photo-grey	4.09 (1.19)	3.4 (1.14)		3.43 (1.33)
doll3-photo-colour	4 (1.2)	3.57 (1.25)	3.05 (1.12)	2.57 (1.16)
doll3-photo-grey	3.55 (1.3)	3.85 (1.35)		2.68 (1.21)
donkey				
donkey1-photo-colour	3.73 (1.2)	3.55 (1.1)	4.05 (0.94)	4.45 (0.76)
donkey1-photo-grey	4.14 (1.01)	3.55 (1.19)		3.93 (0.94)
donkey2-photo-colour	4.15 (1.09)	3.7 (1.13)	3.6 (1.1)	4.43 (0.98)
donkey2-photo-grey	3.65 (1.27)	3.05 (1.23)		4.23 (0.97)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
donkey3-photo-colour	3.95 (1.12)	3.62 (1.16)	3.19 (1.21)	4.32 (0.72)
donkey3-photo-grey	3.46 (1.53)	3.14 (1.17)		4.09 (0.97)
door				
door1-photo-colour	4.95 (0.22)	2.33 (1.12)	2.8 (1.51)	3.43 (1.12)
door1-photo-grey	4.8 (0.52)	2.2 (1.15)		3.27 (1.16)
door2-photo-colour	4.8 (0.52)	3.68 (1.09)	1.91 (1.44)	2.9 (1.17)
door2-photo-grey	4.5 (0.89)	2.71 (1.06)		2.9 (0.97)
door3-photo-colour	4.82 (0.39)	2.32 (0.95)	1.82 (1.05)	3.96 (1.08)
door3-photo-grey	4.81 (0.51)	1.82 (0.8)		3.57 (0.98)
dress				
dress1-photo-colour	4.09 (1.11)	3.45 (0.83)	1.6 (1.1)	2.8 (1.44)
dress1-photo-grey	4.38 (0.74)	2.7 (0.98)		2.37 (1.03)
dress2-photo-colour	4.7 (0.8)	3.9 (0.97)	2.2 (1.64)	2.19 (1.21)
dress2-photo-grey	4.15 (1.27)	3.4 (1.31)		2.82 (1.3)
dress3-photo-colour	4 (1.05)	2.71 (1.01)	1.62 (0.97)	2.77 (0.97)
dress3-photo-grey	4.42 (0.97)	2.05 (0.95)		3.27 (0.88)
dresser				
dresser1-photo-colour	4.25 (1.25)	2.4 (1.27)	2.15 (1.14)	3.14 (1.32)
dresser1-photo-grey	4.8 (0.52)	2.27 (1.08)	2.27 (1.42)	3.1 (1.34)
dresser2-photo-colour	4.5 (0.69)	2.76 (0.89)	1.62 (1.12)	2.65 (1.31)
dresser2-photo-grey	4.6 (0.6)	2.77 (0.92)		2.9 (1.45)
dresser3-photo-colour	4.33 (0.8)	3 (1.02)	2.68 (1.17)	3 (1.26)
dresser3-photo-grey	4.55 (0.6)	2.95 (1.05)		3.52 (1.29)
drum				
drum1-photo-colour	3.9 (1.41)	3.59 (1.05)	2.91 (1.31)	3.6 (1.19)
drum1-photo-grey	4 (1.02)	3.48 (0.87)		3.95 (0.89)
drum2-photo-colour	3.81 (1.36)	3.5 (1.1)	2.45 (1.23)	3.55 (1.15)
drum2-photo-grey	4.05 (1.25)	3.45 (1.05)		4.1 (1.18)
drum3-photo-colour	3.86 (1.46)	3.24 (1.18)	2.76 (1.26)	3.9 (0.89)
drum3-photo-grey	3.96 (1.22)	3.25 (1.03)		3.23 (1.27)
duck				
duck1-photo-colour	4.71 (0.64)	3.75 (1.25)	3.85 (1.18)	4.03 (0.96)
duck1-photo-grey	4.09 (1.23)	3.9 (1.07)		3.55 (0.94)
duck2-photo-colour	4.35 (0.99)	4.65 (0.75)	3.75 (1.41)	4.23 (0.97)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
duck2-photo-grey	4.4 (0.82)	4.1 (0.91)		3.67 (1.32)
duck3-photo-colour	4.04 (1.12)	3.23 (1.23)	3.59 (1.01)	3.59 (1.18)
duck3-photo-grey	4.62 (0.59)	3.38 (1.07)		3.18 (1.14)
eagle				
eagle1-photo-colour	3.65 (1.5)	3.87 (1.14)	4.3 (1.17)	4 (1.05)
eagle1-photo-grey	3.15 (1.63)	2.95 (1.19)		3.23 (1.19)
eagle2-photo-colour	3.4 (1.5)	4.05 (1.33)	4.23 (1.07)	3.95 (1.15)
eagle2-photo-grey	3.4 (1.19)	3.71 (1.01)		3.75 (1.07)
eagle3-photo-colour	3.13 (1.46)	3.5 (1.01)	4.36 (0.95)	4.46 (0.88)
eagle3-photo-grey	3.14 (1.35)	3.73 (1.03)		3.48 (1.03)
fence				
fence1-photo-colour	4.5 (0.74)	2.25 (1.07)	2.6 (1.23)	3.5 (1.05)
fence1-photo-grey	4.57 (0.68)	1.95 (1)		3.23 (0.94)
fence2-photo-colour	4.75 (0.55)	3.1 (1.02)	2.75 (1.48)	3.33 (1.2)
fence2-photo-grey	4.5 (0.61)	2.9 (1.41)		2.86 (1.28)
fence3-photo-colour	4.57 (0.87)	3.05 (1.43)	2.52 (1.36)	3.86 (1.21)
fence3-photo-grey	4.46 (0.83)	2.41 (1.26)		3.36 (1.29)
fish				
fish1-photo-colour	4.62 (0.59)	3.6 (1.14)	3 (1.41)	3.53 (1.17)
fish1-photo-grey	4 (1.11)	3.05 (0.94)		3.3 (1.49)
fish2-photo-colour	4.2 (0.95)	4.05 (0.94)	2.9 (1.37)	3.23 (1.41)
fish2-photo-grey	4.4 (0.75)	4 (0.92)		3.05 (1.32)
fish3-photo-colour	4.38 (1.1)	3.32 (1.25)	2.91 (1.38)	3.5 (1.34)
fish3-photo-grey	4.1 (0.89)	3.29 (1.15)		3.95 (0.95)
flag				
flag1-photo-colour	4.62 (0.8)	2.4 (1.1)	2.1 (1.33)	2.9 (1.42)
flag1-photo-grey	3.73 (1.39)	2.15 (0.88)		3.3 (1.38)
flag2-photo-colour	4.05 (1.1)	2.85 (1.35)	4.25 (1.25)	2.95 (1.46)
flag2-photo-grey	4.6 (0.68)	2.25 (0.91)		2.48 (1.33)
flag3-photo-colour	4.28 (1.02)	2.5 (1.1)	2.09 (1.31)	3.09 (1.19)
flag3-photo-grey	3.81 (1.17)	2.29 (0.96)		2.73 (1.52)
flower				
flower1-photo-colour	4.45 (1)	3.8 (1.36)	3.1 (1.41)	3.5 (1.34)
flower1-photo-grey	4.65 (0.67)	3.37 (1.03)	1.82 (1.08)	3.14 (1.31)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
flower2-photo-colour	4.25 (1.07)	3.62 (1.16)	1.81 (0.98)	3.2 (1.2)
flower2-photo-grey	3.9 (1.21)	3.95 (1)		2.4 (0.99)
flower3-photo-colour	4.14 (0.85)	3.82 (0.85)	3.05 (1.33)	3.14 (1.42)
flower3-photo-grey	4.22 (0.8)	3.59 (1.01)		2.92 (1.1)
flute				
flute1-photo-colour	3.35 (1.23)	3.8 (1.03)	4.1 (1.07)	3.95 (1.24)
flute1-photo-grey	3.3 (1.45)	3.25 (1.16)		3.77 (1.11)
flute2-photo-colour	2.55 (1.39)	2.86 (1.08)	2.55 (1.26)	2.8 (1.36)
flute2-photo-grey	2.95 (1.5)	2.48 (1.17)		3.05 (1.43)
flute3-photo-colour	3.04 (1.72)	3.32 (0.95)	3.59 (1.14)	3.92 (1.35)
flute3-photo-grey	3.05 (1.75)	2.95 (1.05)		4.05 (1.07)
foot				
foot1-photo-colour	4.95 (0.22)	2.8 (1.24)	3.3 (1.42)	4.27 (0.98)
foot1-photo-grey	5 (0)	2.97 (0.93)	1.82 (0.98)	4.05 (0.8)
foot2-photo-colour	4.85 (0.49)	2.62 (1.36)	2.29 (1.27)	3.55 (1.32)
foot2-photo-grey	4.65 (0.81)	3.36 (1.14)		3.2 (1.2)
foot3-photo-colour	4.81 (0.51)	2.18 (1.01)	2.95 (1.36)	3.9 (1.37)
foot3-photo-grey	4.91 (0.43)	2.59 (1.14)		3.96 (1.12)
frog				
frog1-photo-colour	4.2 (1.32)	3.86 (1.28)	3.77 (1.07)	3.95 (1.05)
frog1-photo-grey	4.27 (0.87)	3.9 (0.94)		3.75 (0.97)
frog2-photo-colour	4.05 (1.16)	3.85 (1.09)	3.95 (1.1)	4 (1.03)
frog2-photo-grey	4 (1.31)	3.55 (1.19)		3.81 (1.25)
frog3-photo-colour	4 (1.07)	3.86 (1.15)	3.71 (1.38)	4.14 (0.96)
frog3-photo-grey	4.05 (1.13)	3.88 (1.2)		3.32 (0.99)
giraffe				
giraffe1-photo-colour	4.43 (1.03)	4.1 (1.07)	4.85 (0.37)	4.5 (0.97)
giraffe1-photo-grey	3.41 (1.59)	3.65 (1.09)		4.15 (1.18)
giraffe2-photo-colour	3.5 (1.36)	4.25 (0.91)	4.75 (0.55)	4.59 (0.96)
giraffe2-photo-grey	4 (1.41)	3.8 (0.77)		4.24 (1.09)
giraffe3-photo-colour	3.83 (1.43)	3.78 (1.09)	4.39 (0.94)	4.86 (0.47)
giraffe3-photo-grey	4.14 (1.2)	3.71 (1.19)		4.18 (1.05)
glasses				
glasses1-photo-colour	4.57 (0.6)	2.6 (1.19)	1.95 (1.32)	3.87 (1.04)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
glasses1-photo-grey	4.64 (0.85)	2.1 (0.79)		4 (0.97)
glasses2-photo-colour	4.8 (0.41)	2.35 (1.23)	2.35 (1.63)	4.09 (0.92)
glasses2-photo-grey	4.65 (0.81)	2.15 (1.04)		3.71 (1.45)
glasses3-photo-colour	4.88 (0.45)	2.68 (0.84)	2.18 (1.14)	3.05 (1.09)
glasses3-photo-grey	4.33 (0.91)	2.43 (1.03)		3.32 (1.21)
goat				
goat1-photo-colour	4.19 (1.03)	3.9 (1.07)	3.8 (0.95)	4 (1.05)
goat1-photo-grey	3.23 (1.31)	3.75 (0.55)		3.7 (1.08)
goat2-photo-colour	3.6 (1.1)	4 (0.97)	3.25 (1.45)	4.09 (0.97)
goat2-photo-grey	3.95 (1)	3.7 (1.03)		3.9 (1.14)
goat3-photo-colour	3.38 (1.53)	3.91 (1.04)	3.83 (1.15)	3.32 (1.13)
goat3-photo-grey	3.43 (1.21)	3.81 (1.25)		2.86 (1.21)
grapes				
grapes1-photo-colour	4.65 (0.75)	3.53 (1.14)	4.2 (0.89)	3.76 (1)
grapes1-photo-grey	4.7 (0.57)	2.85 (1.31)		3 (1.31)
grapes2-photo-colour	4.45 (0.76)	3.82 (1.18)	3.64 (1.14)	3.25 (1.16)
grapes2-photo-grey	4.2 (1.01)	3.48 (0.98)		3 (0.97)
grapes3-photo-colour	4.61 (0.66)	3.32 (0.99)	3.68 (0.95)	3.96 (1.04)
grapes3-photo-grey	4.43 (0.68)	3.45 (0.91)		3.14 (1.01)
guitar				
guitar1-photo-colour	4.15 (1.27)	3.35 (1.09)	3.05 (1.23)	4.45 (0.86)
guitar1-photo-grey	4.35 (0.99)	3.3 (1.02)	2 (1.1)	3.9 (0.7)
guitar2-photo-colour	4.45 (0.89)	2.86 (1.2)	2.1 (1)	4.3 (1.03)
guitar2-photo-grey	4.25 (1.12)	3.5 (1.1)		3.85 (1.09)
guitar3-photo-colour	3.67 (1.46)	2.86 (1.04)	3.32 (1.25)	3.95 (1.16)
guitar3-photo-grey	4.3 (0.93)	3.05 (0.72)		4.17 (1.05)
hammer				
hammer1-photo-colour	4.5 (0.89)	2.4 (1.22)	3.7 (1.22)	4.62 (0.5)
hammer1-photo-grey	4.5 (0.76)	2.1 (1.02)		3.5 (1.19)
hammer2-photo-colour	4.75 (0.55)	2.95 (1.05)	3.23 (1.19)	4.4 (0.88)
hammer2-photo-grey	4.55 (0.83)	2.43 (1.12)		4.55 (1)
hammer3-photo-colour	4.32 (0.95)	2.77 (0.97)	2.68 (1.52)	4.25 (1.03)
hammer3-photo-grey	4.29 (1.01)	2.36 (0.85)		3.71 (1.19)
hand				

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
hand1-photo-colour	4.9 (0.4)	3.33 (0.91)	3.14 (1.56)	4.6 (0.75)
hand1-photo-grey	5 (0)	3.14 (1.28)		4.2 (0.89)
hand2-photo-colour	4.73 (0.88)	3.4 (1.23)	3.15 (1.39)	4.62 (0.67)
hand2-photo-grey	5 (0)	3.05 (1.32)		3.9 (0.91)
hand3-photo-colour	4.86 (0.64)	3.62 (1.44)	2.92 (1.28)	2.73 (1.24)
hand3-photo-grey	4.86 (0.47)	3.24 (1.04)		3.43 (1.25)
harp				
harp1-photo-colour	3.25 (1.77)	3.7 (1.02)	3.55 (1.32)	3.81 (0.75)
harp1-photo-grey	2.75 (1.48)	3.15 (1.31)		3.68 (1.09)
harp2-photo-colour	3.15 (1.57)	4.09 (1.06)	2.91 (1.27)	4.15 (0.88)
harp2-photo-grey	3.4 (1.14)	3.14 (1.06)		4.25 (0.85)
harp3-photo-colour	2.64 (1.47)	3.41 (0.96)	2.91 (1.31)	4.46 (0.78)
harp3-photo-grey	3 (1.58)	3.09 (0.92)		3.76 (1.22)
horse				
horse1-photo-colour	4.27 (0.94)	3.45 (1.23)	2.45 (1.15)	4.3 (0.86)
horse1-photo-grey	4.48 (0.93)	3.35 (0.99)		3.73 (0.94)
horse2-photo-colour	4.4 (0.88)	3.75 (0.97)	3.55 (1)	3.86 (1.06)
horse2-photo-grey	3.9 (1.17)	3.55 (1.32)		3.68 (1.17)
horse3-photo-colour	4.14 (1.06)	3.62 (1.2)	3.19 (1.33)	4.14 (1.13)
horse3-photo-grey	3.96 (1.37)	3.23 (1.19)		3.73 (0.98)
house				
house1-photo-colour	4.57 (0.98)	2.95 (1.05)	2.15 (1.23)	2.57 (1.17)
house1-photo-grey	4.5 (1.06)	2.45 (0.89)		2.8 (1.4)
house2-photo-colour	4.75 (0.44)	3.55 (1.19)	2.6 (1.6)	2.86 (0.83)
house2-photo-grey	4.75 (0.55)	3.35 (0.93)		2.33 (1.11)
house3-photo-colour	4.67 (0.87)	4.09 (1.2)	3.04 (1.52)	2.64 (1.26)
house3-photo-grey	4.29 (1.01)	3.95 (1.32)		2 (0.93)
iron				
iron1-photo-colour	4.37 (1.07)	3.71 (1.06)	2.9 (1.45)	3.95 (1.39)
iron1-photo-grey	4.9 (0.45)	2.95 (1.25)		4.3 (0.98)
iron2-photo-colour	4.59 (0.73)	3.45 (1.28)	2.65 (1.57)	3.9 (1.45)
iron2-photo-grey	4.33 (1.32)	3.75 (1.37)		4.2 (0.83)
iron3-photo-colour	4.14 (0.99)	2.72 (1.17)	2.48 (1.42)	2.73 (1.39)
iron3-photo-grey	4.27 (1.12)	1.95 (1.05)		3.9 (0.83)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
jacket				
jacket1-photo-colour	4.5 (0.89)	3.15 (1.14)	2.95 (1.36)	2.41 (1.14)
jacket1-photo-grey	4.8 (0.41)	3.33 (1.06)	2.27 (1.56)	2.52 (1.12)
jacket2-photo-colour	4.55 (0.51)	2.76 (0.89)	2.71 (1.27)	3 (1.21)
jacket2-photo-grey	4.7 (0.47)	3.14 (0.94)		3.1 (1.17)
jacket3-photo-colour	4.43 (0.75)	2.55 (0.91)	1.73 (0.94)	2.71 (1.19)
jacket3-photo-grey	4.32 (0.84)	2.91 (0.9)		3.29 (1.16)
kettle				
kettle1-photo-colour	4.37 (1.22)	3.05 (0.8)	2.81 (1.54)	3.3 (1.03)
kettle1-photo-grey	4.75 (0.64)	2.68 (1.25)		2.7 (1.17)
kettle2-photo-colour	4.5 (0.8)	2.85 (1.23)	2.6 (1.05)	3.1 (1.41)
kettle2-photo-grey	4.29 (1.06)	2.75 (1.29)		2.6 (1.23)
kettle3-photo-colour	4.09 (1.19)	2.64 (1.11)	2.2 (1.08)	2.14 (1.13)
kettle3-photo-grey	4.5 (0.67)	2.32 (1.04)		2.05 (0.97)
kite				
kite1-photo-colour	4.25 (1.16)	3.1 (1.06)	2.45 (1.67)	4.05 (0.74)
kite1-photo-grey	4.2 (1.06)	2.4 (1.23)		3.32 (1.13)
kite2-photo-colour	4.2 (1.11)	3.05 (1.17)	2 (1.38)	3.85 (0.99)
kite2-photo-grey	3.6 (1.05)	2.33 (1.06)		3.45 (1.15)
kite3-photo-colour	3.23 (1.45)	3.48 (0.73)	2.09 (1.5)	3.08 (1.35)
kite3-photo-grey	3.29 (1.35)	3.14 (0.77)		2.14 (1.01)
knife				
knife1-photo-colour	4.57 (0.86)	2.67 (0.91)	3.43 (0.93)	3.5 (1.15)
knife1-photo-grey	4.95 (0.22)	2.32 (1.09)		3.6 (0.99)
knife2-photo-colour	4.91 (0.29)	2.6 (0.82)	3.15 (1.18)	3.67 (1.2)
knife2-photo-grey	4.67 (0.91)	2.2 (0.77)		3.7 (1.3)
knife3-photo-colour	4.23 (1.23)	3.21 (1.02)	3.08 (1.35)	2.14 (0.99)
knife3-photo-grey	4.27 (0.94)	2.81 (1.03)		2.62 (0.92)
ladder				
ladder1-photo-colour	4.75 (0.55)	2 (1.07)	2.95 (1.4)	3.1 (1.37)
ladder1-photo-grey	4.3 (1.06)	1.62 (0.67)		3.95 (1.05)
ladder2-photo-colour	4.52 (0.98)	1.95 (1.28)	1.8 (1.01)	4.1 (0.97)
ladder2-photo-grey	4.5 (0.74)	1.6 (0.94)		4.24 (1.09)
ladder3-photo-colour	4.5 (0.96)	2.41 (1.22)	2.91 (1.41)	3.19 (1.29)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
ladder3-photo-grey	4.27 (0.98)	1.71 (0.95)		3.5 (1.06)
lamp				
lamp1-photo-colour	4.5 (0.86)	2.76 (0.77)	2.9 (1.14)	3.55 (1.28)
lamp1-photo-grey	4.85 (0.49)	2.41 (0.85)		3.2 (1.2)
lamp2-photo-colour	4.64 (0.9)	3.35 (0.67)	2.55 (1.28)	4.05 (1.02)
lamp2-photo-grey	4.71 (0.56)	2.95 (1)		3.6 (1.05)
lamp3-photo-colour	4.45 (1.06)	3.92 (1.1)	2.62 (1.24)	2.91 (1.15)
lamp3-photo-grey	4.77 (0.53)	3.32 (0.95)		2.81 (1.03)
leaf				
leaf1-photo-colour	4.8 (0.52)	3.55 (1.1)	3.14 (1.28)	3.35 (1.35)
leaf1-photo-grey	4.63 (0.89)	3.43 (1.21)		3.4 (1.19)
leaf2-photo-colour	4.76 (0.54)	2.85 (1.31)	3.05 (1.15)	4.1 (1.07)
leaf2-photo-grey	4.86 (0.64)	2.9 (1.02)		3.48 (1.33)
leaf3-photo-colour	4.68 (0.57)	3.1 (1.26)	2.9 (0.94)	3.38 (1.32)
leaf3-photo-grey	4.55 (0.96)	2.62 (1.24)		2.77 (1.27)
lemon				
lemon1-photo-colour	4.76 (0.7)	3.05 (1.32)	4.5 (0.76)	3.8 (1.19)
lemon1-photo-grey	4.27 (0.98)	2.45 (0.94)		3.35 (1.35)
lemon2-photo-colour	4.65 (0.59)	3.05 (1.43)	4.7 (0.57)	4.91 (0.29)
lemon2-photo-grey	4.5 (0.69)	2.95 (1.23)		3.76 (1.14)
lemon3-photo-colour	4.75 (0.61)	2.26 (1.48)	4.26 (1.18)	5 (0)
lemon3-photo-grey	3.81 (1.17)	2.62 (1.02)		3.45 (1.34)
lion				
lion1-photo-colour	4.1 (1.21)	3.64 (1.05)	4.32 (0.99)	4.7 (0.57)
lion1-photo-grey	4.2 (1)	3.33 (1.24)		4.35 (0.81)
lion2-photo-colour	4.1 (1.37)	3.75 (0.91)	4.65 (0.67)	4.65 (0.93)
lion2-photo-grey	4.05 (1.21)	3.4 (1.1)		4.29 (0.78)
lion3-photo-colour	3.95 (1.29)	3.95 (1.09)	4.32 (1.09)	4.67 (0.48)
lion3-photo-grey	3.41 (1.37)	3.75 (1.39)		3.41 (1.18)
lips				
lips1-photo-colour	4.8 (0.62)	3.23 (1.07)	3.09 (0.97)	3.8 (0.95)
lips1-photo-grey	4.87 (0.51)	3.29 (1.01)		3.8 (0.95)
lips2-photo-colour	4.86 (0.65)	2.15 (0.81)	3.25 (1.29)	4.55 (0.6)
lips2-photo-grey	4.82 (0.85)	2.15 (0.99)		4.14 (1.28)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
lips3-photo-colour	4.82 (0.66)	2.41 (1.05)	3.27 (1.49)	4.05 (0.92)
lips3-photo-grey	4.77 (0.53)	1.96 (1.21)		3.32 (1.25)
lobster				
lobster1-photo-colour	3.45 (1.5)	4.17 (1.15)	4.15 (0.99)	4.43 (0.68)
lobster1-photo-grey	3.5 (1.5)	3.65 (1.27)		3.36 (1.09)
lobster2-photo-colour	4.24 (1.14)	4.05 (1)	4.18 (1.05)	4.15 (0.81)
lobster2-photo-grey	3.75 (1.33)	3 (1.3)		3.35 (1.23)
lobster3-photo-colour	2.95 (1.53)	3.5 (0.91)	4.27 (0.83)	4.44 (0.71)
lobster3-photo-grey	2.95 (1.4)	3.18 (0.91)		3.29 (1.31)
lock				
lock1-photo-colour	4.3 (0.99)	3.81 (1.21)	4.05 (0.92)	4.15 (1.18)
lock1-photo-grey	4.5 (0.69)	3.32 (1.25)		3.45 (1.39)
lock2-photo-colour	4.27 (1.08)	3.4 (1.19)	2.7 (1.03)	3.48 (1.4)
lock2-photo-grey	4.29 (1.1)	2.6 (1.35)		4.1 (0.97)
lock3-photo-colour	4.04 (1.02)	3.54 (1.22)	3.75 (1.26)	3.82 (1.14)
lock3-photo-grey	4.36 (0.85)	2.86 (0.94)		3.57 (1.25)
mitten				
mitten1-photo-colour	4.05 (1.25)	3.95 (0.94)	1.85 (1.14)	3.35 (1.18)
mitten1-photo-grey	4.48 (0.68)	3.2 (1.11)		3.67 (1.06)
mitten2-photo-colour	4.1 (1.17)	3.35 (0.88)	2 (1.59)	3.38 (1.32)
mitten2-photo-grey	3.75 (1.21)	2.3 (1.26)		3.82 (1.05)
mitten3-photo-colour	4.14 (0.96)	2.43 (1.33)	1.86 (1.49)	3.68 (1.36)
mitten3-photo-grey	4.33 (0.87)	1.91 (1.06)		3.77 (0.97)
monkey				
monkey1-photo-colour	3.62 (1.24)	3.95 (1.32)	3.85 (1.14)	3.47 (1.2)
monkey1-photo-grey	3.09 (1.6)	3.9 (0.97)		3.1 (1.25)
monkey2-photo-colour	3.1 (1.25)	4.15 (1.14)	4.15 (1.14)	3.59 (1.01)
monkey2-photo-grey	3.65 (1.42)	3.85 (0.99)		3.1 (1.18)
monkey3-photo-colour	3.79 (1.41)	3.36 (1.22)	3.59 (1.01)	3.55 (0.96)
monkey3-photo-grey	3.62 (1.24)	3.29 (1.15)		3.23 (1.23)
moon				
moon1-photo-colour	4.6 (0.68)	2.6 (1)	2.8 (1.15)	2.71 (1.27)
moon1-photo-grey	4.05 (1.47)	2.25 (1.02)		2.95 (1.4)
moon2-photo-colour	4.1 (1.17)	2.45 (1.14)	3.36 (1.05)	2.55 (1.28)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
moon2-photo-grey	3.55 (1.39)	1.9 (0.94)		2.6 (1.27)
moon3-photo-colour	4.26 (1.01)	2.73 (1.28)	3.32 (0.99)	2.46 (1.25)
moon3-photo-grey	4.14 (1.11)	2.09 (0.97)		2.24 (1.37)
mouse				
mouse1-photo-colour	3.68 (1.25)	3.75 (0.91)	3.6 (1.19)	3.75 (1.29)
mouse1-photo-grey	4.43 (0.87)	3.35 (1.09)		3.5 (1.17)
mouse2-photo-colour	4.15 (0.99)	3.9 (0.91)	3.7 (1.03)	3.76 (1.45)
mouse2-photo-grey	3.8 (1.24)	3.1 (1.37)		3.95 (1.05)
mouse3-photo-colour	3.76 (1.04)	3.9 (1.26)	3.33 (1.24)	3.41 (1.18)
mouse3-photo-grey	3.71 (1.43)	3.13 (1.18)		3 (1.38)
nail				
nail1-photo-colour	4.17 (1.21)	2.14 (1.31)	3.57 (1.43)	3.7 (1.75)
nail1-photo-grey	4.5 (0.95)	2.18 (1.14)		3.4 (1.7)
nail2-photo-colour	4.32 (0.95)	2.05 (1.05)	2.7 (1.22)	3.33 (1.77)
nail2-photo-grey	4.67 (0.58)	1.5 (0.76)		3.85 (1.27)
nail3-photo-colour	4.18 (1.14)	2.12 (1.17)	3.88 (1.51)	3.77 (1.69)
nail3-photo-grey	4.41 (1.05)	1.77 (1.07)		4.14 (1.39)
needle				
needle1-photo-colour	3.87 (1.22)	2.48 (1.4)	3.48 (1.08)	4.6 (0.82)
needle1-photo-grey	4.45 (1.05)	2.27 (1.24)		4.4 (0.99)
needle2-photo-colour	4.09 (1.38)	1.9 (0.85)	2.4 (1.23)	4 (1.34)
needle2-photo-grey	4.38 (0.97)	1.7 (1.03)		3.95 (1.15)
needle3-photo-colour	3.64 (1.18)	1.92 (1.32)	3.76 (1.54)	4.05 (1.05)
needle3-photo-grey	3.82 (1.22)	1.95 (0.74)		3.81 (1.29)
nose				
nose1-photo-colour	5 (0)	2.83 (1.42)	3.15 (1.76)	4 (0.84)
nose1-photo-grey	5 (0)	2.55 (1.19)		3.82 (1.26)
nose2-photo-colour	4.95 (0.22)	3.23 (1.11)	2.68 (1.52)	3.85 (1.09)
nose2-photo-grey	4.7 (0.66)	2.38 (1.02)		3.6 (0.99)
nose3-photo-colour	4.86 (0.47)	3.17 (0.94)	3.13 (1.36)	4.21 (0.98)
nose3-photo-grey	4.76 (0.7)	2.55 (1.26)		3.29 (1.19)
onion				
onion1-photo-colour	4.9 (0.31)	2.83 (1.29)	3.9 (1.33)	4.67 (0.73)
onion1-photo-grey	4.8 (0.52)	2.55 (1.15)		3.36 (1.22)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
onion2-photo-colour	4.8 (0.7)	3.45 (1.37)	3.82 (0.96)	4.25 (1.07)
onion2-photo-grey	4.5 (0.69)	2.95 (1.2)		3.65 (0.93)
onion3-photo-colour	4.48 (0.79)	2.59 (1.22)	3.86 (1.08)	4.75 (0.68)
onion3-photo-grey	4.38 (0.92)	2.23 (0.97)		3.24 (1.18)
orange				
orange1-photo-colour	4.68 (0.65)	2.3 (1.45)	4.7 (0.73)	4.75 (0.55)
orange1-photo-grey	4.48 (0.87)	2.35 (1.04)		2.83 (1.29)
orange2-photo-colour	4.9 (0.31)	3.25 (1.37)	4.85 (0.37)	4.81 (0.68)
orange2-photo-grey	4.25 (0.91)	1.9 (1.37)		3.68 (1.04)
orange3-photo-colour	4.86 (0.36)	3.57 (1.33)	4.67 (0.58)	4.45 (1.06)
orange3-photo-grey	4.17 (1.2)	2.26 (1.14)		3.23 (1.07)
ostrich				
ostrich1-photo-colour	3.47 (1.38)	3.81 (0.68)	4.19 (1.12)	4.7 (0.47)
ostrich1-photo-grey	3.4 (1.57)	3.18 (1.01)		4.3 (0.73)
ostrich2-photo-colour	3.27 (1.35)	3.85 (0.99)	3.55 (0.89)	4.52 (0.68)
ostrich2-photo-grey	3.24 (1.41)	3.5 (1.43)		3.9 (1.07)
ostrich3-photo-colour	3.18 (1.59)	3.8 (1)	4.12 (1.01)	3.77 (1.15)
ostrich3-photo-grey	2.95 (1.7)	3.36 (1.26)		3.43 (1.25)
peach				
peach1-photo-colour	4.45 (1)	3.14 (1.13)	3.91 (1.02)	3.95 (1.1)
peach1-photo-grey	4.23 (1.1)	2.95 (1.2)		2.95 (1.39)
peach2-photo-colour	4.19 (1.29)	3.2 (1.28)	4.2 (1.01)	4.45 (0.76)
peach2-photo-grey	4.18 (1.05)	2.2 (0.95)		2.9 (1.41)
peach3-photo-colour	4.14 (1.08)	2.29 (0.96)	3.95 (1.16)	4 (1.05)
peach3-photo-grey	3.68 (1.52)	1.83 (1.01)		1.73 (0.98)
peacock				
peacock1-photo-colour	3.87 (1.36)	4.43 (0.68)	4.76 (0.44)	4.4 (0.68)
peacock1-photo-grey	3.7 (1.56)	3.59 (1.3)		3.1 (1.12)
peacock2-photo-colour	3.73 (1.49)	4.55 (0.83)	4.2 (1.06)	4 (1.1)
peacock2-photo-grey	3.52 (1.47)	3.85 (1.31)		2.55 (1.05)
peacock3-photo-colour	3.45 (1.41)	4.44 (0.65)	4.56 (0.92)	4.05 (0.95)
peacock3-photo-grey	3.5 (1.47)	3.82 (1.33)		3.24 (1.14)
peanut				
peanut1-photo-colour	4.38 (0.86)	3.2 (1.2)	4.1 (1.02)	3.97 (1.38)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
peanut1-photo-grey	3.91 (1.11)	2.75 (1.12)		3.35 (1.31)
peanut2-photo-colour	4 (0.97)	3.45 (1.15)	4.4 (0.75)	3.95 (1.46)
peanut2-photo-grey	4.2 (0.89)	3.45 (0.94)		3.86 (1.28)
peanut3-photo-colour	4.25 (1.03)	2.83 (1.19)	4.13 (1.1)	4.05 (1.17)
peanut3-photo-grey	4.05 (1.12)	3.05 (0.86)		4.14 (1.04)
pear				
pear1-photo-colour	4.57 (0.86)	3.14 (1.24)	4.48 (0.6)	5 (0)
pear1-photo-grey	4.8 (0.41)	2.59 (1.22)		3.55 (1.23)
pear2-photo-colour	4.68 (0.78)	2.8 (1.2)	3.6 (1.1)	4.33 (0.73)
pear2-photo-grey	4.33 (1.06)	2.45 (1.05)		3.55 (1.05)
pear3-photo-colour	4.32 (0.99)	3.08 (1.19)	4.12 (1.13)	4.09 (0.92)
pear3-photo-grey	4.41 (1.05)	2.32 (1.13)		3.14 (0.96)
pencil				
pencil1-photo-colour	4.95 (0.22)	2.25 (1.55)	2.15 (1.6)	3.77 (1.19)
pencil1-photo-grey	4.5 (0.91)	1.75 (1.02)		3.85 (1.04)
pencil2-photo-colour	4.8 (0.7)	2.8 (1.54)	2.3 (1.49)	3.86 (1.13)
pencil2-photo-grey	4.9 (0.31)	2.6 (1.23)		3.71 (1.01)
pencil3-photo-colour	4.83 (0.48)	1.86 (1.04)	2.41 (1.14)	3.77 (0.97)
pencil3-photo-grey	4.52 (0.75)	1.86 (0.96)		2.64 (1.22)
penguin				
penguin1-photo-colour	4 (1.45)	3.4 (1.33)	4.5 (0.95)	4.62 (0.59)
penguin1-photo-grey	3.5 (1.4)	2.65 (1.18)		3.91 (0.87)
penguin2-photo-colour	3.85 (1.57)	3.82 (1.26)	4.55 (0.6)	4.65 (0.59)
penguin2-photo-grey	4.05 (1.15)	2.9 (1.26)		4.5 (0.69)
penguin3-photo-colour	3.59 (1.53)	3.45 (0.91)	4.18 (0.91)	4.5 (0.78)
penguin3-photo-grey	3.52 (1.44)	2.77 (0.75)		3.71 (1.1)
pepper				
pepper1-photo-colour	4.73 (0.46)	2.2 (1.36)	2.95 (1.1)	3.1 (1.48)
pepper1-photo-grey	4.71 (0.64)	2.4 (1.1)		2.77 (1.3)
pepper2-photo-colour	4.65 (0.67)	3.1 (1.29)	3.5 (1.36)	2.52 (1.66)
pepper2-photo-grey	4.55 (0.6)	2.15 (1.39)		3.05 (1.33)
pepper3-photo-colour	4.38 (0.67)	3.33 (1.32)	3 (1.18)	2.82 (1.22)
pepper3-photo-grey	4.29 (1)	2.35 (0.78)		2.95 (1.46)
piano				

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
piano1-photo-colour	4.48 (0.75)	3.9 (1.25)	3.55 (1.19)	3.87 (1.11)
piano1-photo-grey	3.95 (1.13)	3.95 (1.15)		4.2 (0.95)
piano2-photo-colour	4.15 (0.81)	4.15 (1.09)	3.6 (1.27)	4.45 (0.74)
piano2-photo-grey	4.15 (1.23)	3.4 (0.94)		4.14 (1.15)
piano3-photo-colour	4.29 (1.04)	3.77 (0.97)	3.5 (1.1)	3.86 (0.89)
piano3-photo-grey	4.14 (1.01)	3.71 (1.15)		3.95 (0.9)
pipe				
pipe1-photo-colour	3.43 (1.25)	4.1 (1.33)	3.2 (1.11)	2.8 (1.71)
pipe1-photo-grey	2.82 (1.5)	4.35 (0.88)		2.8 (1.47)
pipe2-photo-colour	3 (1.38)	2.75 (1.62)	3.55 (1.23)	2.77 (1.6)
pipe2-photo-grey	3.65 (1.23)	2.6 (1.1)		2.86 (1.53)
pipe3-photo-colour	3 (1.53)	2.26 (1.01)	3.17 (1.3)	2.86 (1.49)
pipe3-photo-grey	3.29 (1.52)	2.38 (1.12)		2.68 (1.7)
pitcher				
pitcher1-photo-colour	4.25 (1.07)	3.3 (1.26)	3.45 (1.36)	3.18 (1.4)
pitcher1-photo-grey	4.5 (0.76)	2.5 (0.9)	2.27 (1.1)	2.9 (1.18)
pitcher2-photo-colour	4.45 (0.6)	2.81 (1.17)	1.57 (1.12)	2.3 (0.98)
pitcher2-photo-grey	3.85 (1.14)	2.64 (1)		3.1 (1.25)
pitcher3-photo-colour	3.67 (1.2)	2.18 (1.14)	1.55 (0.96)	2.29 (1.01)
pitcher3-photo-grey	3.39 (1.08)	2.09 (0.81)		2.67 (1.27)
pliers				
pliers1-photo-colour	4.09 (1.31)	3 (1.08)	2.4 (1.05)	3.9 (0.97)
pliers1-photo-grey	3.95 (1.16)	3.05 (1.05)		3.4 (1.16)
pliers2-photo-colour	4.25 (0.79)	2.75 (1.33)	2.9 (1.37)	4.19 (0.98)
pliers2-photo-grey	3.6 (1.31)	2.45 (1.23)		3.73 (1.2)
pliers3-photo-colour	4.05 (1.16)	2.67 (1.06)	2.33 (1.32)	3.82 (1.05)
pliers3-photo-grey	3.58 (1.18)	2.09 (0.97)		3.82 (1.1)
plug				
plug1-photo-colour	4.05 (1.36)	2.65 (1.14)	3 (1.17)	2.09 (1.27)
plug1-photo-grey	4.45 (0.83)	2.7 (1.15)	3.18 (1.4)	2.33 (1.39)
plug2-photo-colour	4.1 (1.29)	2.19 (0.93)	2.86 (1.31)	2 (1.49)
plug2-photo-grey	4.5 (1.15)	3.18 (1.1)		2.35 (1.5)
plug3-photo-colour	3.81 (1.4)	2.5 (0.86)	3.5 (1.19)	2.24 (1.3)
plug3-photo-grey	3 (1.45)	2.73 (1.12)		2.48 (1.42)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
potato				
potato1-photo-colour	4.67 (1.03)	2.71 (1.45)	4 (1.26)	5 (0)
potato1-photo-grey	4.85 (0.37)	2.45 (1.1)		4 (1.12)
potato2-photo-colour	4.82 (0.5)	2.75 (1.12)	3.7 (1.3)	4.52 (0.81)
potato2-photo-grey	4.38 (1.24)	2.45 (1.15)		3.5 (1.24)
potato3-photo-colour	5 (0)	2.6 (1.55)	4.08 (1.08)	4.55 (0.67)
potato3-photo-grey	4.82 (0.5)	2.48 (1.08)		3.76 (1)
pumpkin				
pumpkin1-photo-colour	4.23 (1.23)	3 (1.17)	4.2 (1.28)	4.6 (0.75)
pumpkin1-photo-grey	4.1 (1.04)	2.55 (1)		3.2 (1.27)
pumpkin2-photo-colour	4 (1.08)	2.95 (1.23)	4.45 (0.83)	4.14 (1.2)
pumpkin2-photo-grey	3.5 (1.19)	2.25 (1.02)		3.18 (1.05)
pumpkin3-photo-colour	4.05 (0.97)	3.24 (1.37)	4.52 (0.98)	4.55 (0.74)
pumpkin3-photo-grey	3.79 (1.25)	2.35 (1.15)		3.55 (1.06)
rabbit				
rabbit1-photo-colour	4.35 (1.18)	3.45 (1.37)	3.64 (0.95)	3.9 (0.85)
rabbit1-photo-grey	4.3 (0.99)	3.38 (0.97)		4.05 (1.05)
rabbit2-photo-colour	4.14 (1.28)	3.2 (1.06)	3.6 (1.1)	3.7 (1.22)
rabbit2-photo-grey	4 (1.23)	3.1 (1.17)		3.43 (1.12)
rabbit3-photo-colour	4.23 (0.97)	3.95 (1.2)	3.81 (1.25)	3.9 (1.09)
rabbit3-photo-grey	4.09 (1.11)	3.8 (1.04)		3.77 (0.81)
raccoon				
raccoon1-photo-colour	3.45 (1.73)	3.53 (1.14)	4.15 (0.88)	4.33 (0.86)
raccoon1-photo-grey	3.45 (1.36)	3.25 (1.33)		4 (1.02)
raccoon2-photo-colour	3.6 (1.47)	4.09 (1.06)	4.55 (0.67)	3.95 (1.19)
raccoon2-photo-grey	3.65 (1.27)	3.62 (1.4)		4.3 (0.73)
raccoon3-photo-colour	2.83 (1.56)	3.59 (1.01)	4.32 (1.04)	4.46 (0.72)
raccoon3-photo-grey	2.86 (1.35)	3.27 (1.12)		3.62 (1.12)
ring				
ring1-photo-colour	4.62 (0.67)	2.35 (0.75)	2.9 (1.17)	3.3 (1.29)
ring1-photo-grey	4.23 (1.07)	2.7 (0.98)		3.4 (1.39)
ring2-photo-colour	2.55 (1.61)	3.45 (1.19)	3.2 (1.36)	2.09 (0.87)
ring2-photo-grey	2.75 (1.71)	2.85 (1.04)		2 (1.18)
ring3-photo-colour	4.08 (1.15)	1.82 (0.85)	2.14 (1.13)	3.27 (1.24)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
ring3-photo-grey	4 (1.14)	1.9 (0.94)		2.59 (1.3)
ruler				
ruler1-photo-colour	4.76 (0.44)	2 (1.26)	2.15 (1.39)	3.33 (1.24)
ruler1-photo-grey	4.45 (1.06)	2.05 (1.05)		3.3 (1.17)
ruler2-photo-colour	4.4 (0.82)	3 (1.59)	3.2 (1.61)	3.86 (1.08)
ruler2-photo-grey	4.55 (0.69)	2.65 (1.04)		4.1 (0.89)
ruler3-photo-colour	4.4 (1.12)	2.55 (1.14)	2.45 (1.14)	3.27 (1.49)
ruler3-photo-grey	4.1 (1.04)	2.62 (1.02)		3.68 (0.99)
screw				
screw1-photo-colour	4.8 (0.62)	3.09 (1.15)	3.27 (1.39)	4.6 (0.6)
screw1-photo-grey	4.4 (1)	3.38 (1.16)		4.9 (0.31)
screw2-photo-colour	4.67 (0.66)	3.1 (1.41)	3.4 (1.43)	4.45 (0.69)
screw2-photo-grey	4.77 (0.53)	2.6 (1.05)		4.38 (0.92)
screw3-photo-colour	4.5 (0.74)	3.05 (1.13)	3.23 (1.45)	4.48 (0.68)
screw3-photo-grey	4.09 (1.35)	2.67 (1.34)		4.41 (0.8)
seal				
seal1-photo-colour	3.8 (1.44)	3.23 (1.23)	3.59 (1.33)	3.95 (1.1)
seal1-photo-grey	3.63 (1.43)	3.48 (1.08)		3.8 (1.15)
seal2-photo-colour	3.76 (1.41)	2.95 (1.15)	4.15 (1.04)	3.95 (1.15)
seal2-photo-grey	3.68 (1.36)	3.25 (1.07)		4.38 (0.97)
seal3-photo-colour	3.64 (1.47)	3.57 (1.16)	4.1 (1.3)	3.71 (1.01)
seal3-photo-grey	3.32 (1.43)	2.84 (1.07)		3.27 (1.16)
sheep				
sheep1-photo-colour	4.3 (1.03)	3.64 (1.29)	3.64 (1.14)	3.95 (1.15)
sheep1-photo-grey	4.1 (1.16)	3.43 (0.87)		4 (0.97)
sheep2-photo-colour	4 (1.22)	3.75 (1.25)	4.3 (0.66)	3.85 (1.09)
sheep2-photo-grey	4.18 (0.91)	3.6 (1.1)		4 (0.89)
sheep3-photo-colour	3.91 (1.02)	3.41 (1.26)	3.95 (1.29)	3.95 (0.97)
sheep3-photo-grey	3.95 (1.05)	3.32 (1.11)		3.68 (1.21)
shirt				
shirt1-photo-colour	4.6 (0.75)	3.35 (1.14)	2.25 (1.45)	3.36 (1.22)
shirt1-photo-grey	4.75 (0.64)	3.17 (0.87)	2.09 (1.22)	3.71 (1.01)
shirt2-photo-colour	4.45 (0.89)	2.52 (1.29)	2.33 (1.2)	3.1 (1.41)
shirt2-photo-grey	4.75 (0.55)	3.27 (1.12)		2.6 (1.39)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
shirt3-photo-colour	4.57 (0.51)	3.09 (1.23)	1.82 (1.18)	3.38 (1.12)
shirt3-photo-grey	4.59 (0.73)	3.18 (0.66)		3.76 (1.01)
shoe				
shoe1-photo-colour	4.8 (0.61)	3.19 (1.08)	2.67 (1.2)	4.05 (1.32)
shoe1-photo-grey	4.95 (0.22)	2.64 (0.9)		3.6 (1.14)
shoe2-photo-colour	4.77 (0.69)	3.55 (0.94)	2.95 (1.1)	3.52 (1.44)
shoe2-photo-grey	4.76 (0.89)	3.05 (1.36)		3.35 (1.35)
shoe3-photo-colour	4.64 (0.73)	3.67 (1.24)	2.79 (1.56)	3.41 (1.59)
shoe3-photo-grey	4.91 (0.29)	2.95 (1.21)		3.86 (0.91)
skirt				
skirt1-photo-colour	4.15 (1.09)	3.23 (1.3)	2.3 (1.63)	2.81 (1.08)
skirt1-photo-grey	3.95 (1.36)	3.05 (1.23)		2.77 (1.15)
skirt2-photo-colour	4.15 (1.18)	3.23 (1.19)	1.68 (1.39)	2.6 (1.27)
skirt2-photo-grey	4 (1.08)	3.05 (1.02)		2.4 (0.94)
skirt3-photo-colour	3.59 (1.26)	2.95 (0.79)	1.77 (1.02)	3.25 (1.19)
skirt3-photo-grey	4 (0.95)	2.73 (0.94)		2.67 (1.06)
skunk				
skunk1-photo-colour	3.19 (1.66)	3.65 (0.93)	4.4 (0.99)	4.17 (1.34)
skunk1-photo-grey	3.09 (1.66)	3.7 (1.08)		4.15 (0.88)
skunk2-photo-colour	2.85 (1.53)	3.65 (1.04)	4.35 (0.88)	3.91 (0.87)
skunk2-photo-grey	3 (1.45)	3.35 (1.14)		3.43 (1.25)
skunk2-photo-grey2	Nan NA	Nan NA		2 NA
skunk3-photo-colour	2.54 (1.35)	3.45 (1.1)	3.55 (1.44)	3.23 (1.15)
skunk3-photo-grey	2.76 (1.45)	3.27 (1.08)		2.82 (1.18)
snail				
snail1-photo-colour	4.14 (0.94)	4.1 (0.72)	3.25 (1.21)	4.15 (1.04)
snail1-photo-grey	4.57 (0.6)	3 (1.03)		3.63 (1.03)
snail2-photo-colour	4.55 (0.69)	3.95 (1.1)	4.25 (0.72)	4.24 (1)
snail2-photo-grey	4.2 (1.2)	3.5 (1.32)		3.95 (1.09)
snail3-photo-colour	4.14 (1.24)	4.05 (1.2)	2.95 (1.2)	3.5 (1.1)
snail3-photo-grey	4.29 (1.04)	3.7 (1.06)		4.05 (0.79)
snake				
snake1-photo-colour	4.1 (1.45)	3.05 (1.17)	2.82 (0.91)	3.65 (1.09)
snake1-photo-grey	4 (1.29)	3.1 (0.94)		3.55 (1)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
snake2-photo-colour	4.38 (1.02)	2.95 (1.1)	2.6 (1.23)	3.45 (0.94)
snake2-photo-grey	3.77 (1.48)	2.65 (0.99)		3.86 (1.11)
snake3-photo-colour	3.5 (1.41)	3.05 (0.97)	2.71 (1.15)	3.57 (1.03)
snake3-photo-grey	3.5 (1.41)	3.08 (1.18)		3.5 (0.96)
snowman				
snowman1-photo-colour	4.25 (1.07)	2.41 (1.18)	3.73 (1.52)	3.05 (1.19)
snowman1-photo-grey	4.17 (0.99)	2.33 (1.15)		3.6 (1.19)
snowman2-photo-colour	4.14 (1.28)	3.4 (1.14)	4.4 (1.05)	3.9 (0.91)
snowman2-photo-grey	4.05 (1.25)	2.6 (1.14)		4.1 (0.94)
snowman3-photo-colour	4.23 (1.02)	3.05 (1.12)	4.29 (1.23)	3.33 (0.97)
snowman3-photo-grey	3.74 (1.39)	2.88 (1.19)		2.68 (1.21)
socks				
socks1-photo-colour	4.9 (0.45)	2.27 (1.12)	2 (1.38)	2.9 (1.29)
socks1-photo-grey	4.8 (0.81)	1.81 (0.87)		4.05 (1.1)
socks2-photo-colour	4.81 (0.87)	2.05 (0.94)	1.65 (0.99)	3.55 (1.39)
socks2-photo-grey	4.91 (0.43)	2 (0.86)		3.29 (1.45)
socks3-photo-colour	5 (0)	2.67 (1.35)	1.9 (1.34)	3.14 (1.2)
socks3-photo-grey	4.73 (0.88)	2.25 (1.26)		3.04 (1.19)
spider				
spider1-photo-colour	4.37 (1.13)	4 (0.77)	3.38 (1.16)	3.3 (1.45)
spider1-photo-grey	4.35 (0.99)	3.68 (1.21)		3.15 (1.27)
spider2-photo-colour	4.59 (0.85)	4.25 (0.91)	3.2 (1.2)	3.33 (1.49)
spider2-photo-grey	4.48 (0.75)	3.95 (1.1)		2.85 (1.04)
spider3-photo-colour	4.05 (1.17)	4.54 (0.93)	3.5 (1.38)	2.27 (1.39)
spider3-photo-grey	4.18 (1.22)	4.05 (1.09)		3.14 (1.24)
spoon				
spoon1-photo-colour	4.95 (0.22)	2.65 (1.42)	3.85 (1.46)	4.37 (0.85)
spoon1-photo-grey	4.73 (0.88)	2.5 (1.24)		4.35 (0.88)
spoon2-photo-colour	4.6 (0.68)	2.7 (1.26)	2.95 (1.39)	3.82 (1.14)
spoon2-photo-grey	4.9 (0.31)	2.55 (1.1)		3.67 (1.28)
spoon3-photo-colour	4.92 (0.28)	2.14 (1.28)	3.64 (1.43)	4.3 (1.06)
spoon3-photo-grey	4.76 (0.54)	2.62 (1.24)		4.36 (1.09)
stool				
stool1-photo-colour	3.5 (1.3)	2.85 (1.04)	2.85 (1.18)	2.45 (1.36)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
stool1-photo-grey	3.52 (1.25)	2.35 (0.88)		1.93 (1.05)
stool2-photo-colour	4.3 (0.8)	3.1 (1.02)	2.8 (1.4)	3.71 (1.19)
stool2-photo-grey	4.3 (0.86)	2.55 (1)		3.23 (1.15)
stool3-photo-colour	4.19 (0.93)	2.86 (1.15)	2.24 (1.26)	3.14 (1.21)
stool3-photo-grey	4.42 (0.65)	2.32 (0.78)		3.41 (1.22)
swan				
swan1-photo-colour	4.62 (0.59)	3.45 (1.28)	4.4 (0.99)	4.27 (1.01)
swan1-photo-grey	3.91 (1.11)	3.2 (1.15)		3.65 (1.35)
swan2-photo-colour	3.9 (1.17)	4.15 (0.88)	4.45 (1.1)	4.64 (0.58)
swan2-photo-grey	4.3 (1.03)	3.9 (0.85)		3.81 (1.17)
swan3-photo-colour	4 (1.35)	2.59 (1.22)	4 (1.2)	4.45 (0.96)
swan3-photo-grey	4.33 (0.8)	2.62 (1.16)		4.05 (0.84)
swing				
swing1-photo-colour	4.03 (1.38)	2.33 (1.06)	3.05 (1.28)	3.65 (0.93)
swing1-photo-grey	4.25 (1.12)	2.09 (0.87)		2.95 (1.28)
swing2-photo-colour	4.05 (1.25)	2.35 (0.88)	2.8 (1.28)	3.14 (1.35)
swing2-photo-grey	4.05 (1.2)	1.9 (0.64)		2.95 (1.39)
swing3-photo-colour	3.45 (1.41)	2.72 (1.14)	3.28 (1.34)	2.86 (1.28)
swing3-photo-grey	3.82 (1.37)	2.14 (0.83)		3.29 (0.96)
table				
table1-photo-colour	4.91 (0.43)	2.15 (1.27)	2.8 (1.36)	3.95 (1)
table1-photo-grey	4.9 (0.44)	1.9 (0.91)		3.47 (1.14)
table2-photo-colour	4.95 (0.22)	1.95 (1.15)	1.8 (1.32)	3.52 (1.36)
table2-photo-grey	4.8 (0.62)	1.5 (0.95)		3.32 (1.13)
table3-photo-colour	4.14 (1.24)	3.48 (1.08)	2.19 (1.21)	2.27 (0.98)
table3-photo-grey	4.29 (0.91)	3.27 (1.12)		2.23 (0.97)
thimble				
thimble1-photo-colour	3.4 (1.64)	2.87 (1.04)	3.4 (1.31)	3.95 (1.28)
thimble1-photo-grey	3.45 (1.64)	2.4 (1.1)		3.91 (1.38)
thimble2-photo-colour	3.25 (1.62)	3.86 (1.13)	2.64 (1.47)	3.7 (1.03)
thimble2-photo-grey	3.15 (1.14)	3.33 (1.46)		4.55 (0.89)
thimble3-photo-colour	2.73 (1.39)	3.09 (1.34)	3.41 (1.4)	4.08 (1.35)
thimble3-photo-grey	2.86 (1.56)	2.59 (1.3)		3.48 (1.6)
thumb				

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
thumb1-photo-colour	4.9 (0.31)	2.95 (1.28)	3.2 (1.44)	3.77 (1.34)
thumb1-photo-grey	5 (0)	2.63 (1.22)	1.82 (0.87)	3.52 (1.29)
thumb2-photo-colour	4.55 (0.69)	2.57 (1.12)	2.81 (1.54)	4.15 (0.99)
thumb2-photo-grey	4.67 (0.91)	3.05 (0.9)		3.3 (1.45)
thumb3-photo-colour	4.67 (0.73)	3.32 (1.09)	2.77 (1.34)	3.52 (1.44)
thumb3-photo-grey	4.43 (1.08)	3.14 (1.32)		4.04 (1.12)
tiger				
tiger1-photo-colour	3.45 (1.71)	4.1 (0.91)	4.45 (0.76)	4.55 (0.83)
tiger1-photo-grey	4.1 (1.18)	3.55 (1.23)		3.53 (1.07)
tiger2-photo-colour	4.15 (1.18)	4.05 (0.83)	4.65 (0.59)	4.57 (0.81)
tiger2-photo-grey	3.6 (1.35)	3.2 (1.28)		4.18 (0.73)
tiger3-photo-colour	4 (1.26)	4.05 (1.2)	4.14 (1.06)	4.59 (0.67)
tiger3-photo-grey	3.75 (1.57)	3.45 (1.22)		4.18 (0.85)
toaster				
toaster1-photo-colour	4.7 (0.66)	2.9 (1.02)	2.6 (1.35)	3.32 (1.13)
toaster1-photo-grey	4.9 (0.31)	2.87 (1.07)	2.55 (1.69)	3.48 (0.93)
toaster2-photo-colour	4.75 (0.55)	3.05 (1.28)	1.9 (1.04)	3.95 (1.23)
toaster2-photo-grey	4.4 (0.75)	3.68 (1.13)		4.1 (0.91)
toaster3-photo-colour	4.57 (0.6)	3.36 (1.05)	1.73 (1.2)	3.19 (1.33)
toaster3-photo-grey	4.41 (0.59)	3.5 (0.74)		3.12 (1.3)
tomato				
tomato1-photo-colour	4.85 (0.49)	2.5 (1.36)	4.35 (1.14)	4.67 (0.58)
tomato1-photo-grey	4.5 (0.95)	2.2 (1.15)		3 (1.27)
tomato2-photo-colour	4.76 (0.54)	3.45 (1.3)	4.18 (0.91)	4.3 (0.73)
tomato2-photo-grey	4.3 (1.08)	2.43 (0.93)		3.05 (1.32)
tomato3-photo-colour	4.35 (0.93)	2.36 (1.22)	4.41 (0.85)	4.92 (0.28)
tomato3-photo-grey	4.38 (0.97)	2.27 (1.16)		3.14 (1.39)
train				
train1-photo-colour	4.36 (0.9)	4.55 (0.6)	2.6 (1.35)	3.3 (1.69)
train1-photo-grey	4.05 (1.02)	4.1 (1.33)		2.57 (1.1)
train2-photo-colour	4.15 (0.93)	4.3 (0.47)	2.3 (1.53)	2.81 (1.25)
train2-photo-grey	3.85 (1.14)	3.5 (1.5)		3.05 (1.25)
train3-photo-colour	4.52 (0.68)	4 (1.26)	2.57 (1.36)	2.91 (0.87)
train3-photo-grey	4.42 (1.02)	3.95 (1)		3.18 (1.22)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
tree				
tree1-photo-colour	4.9 (0.3)	4.05 (1.28)	3.7 (1.03)	3.8 (1.24)
tree1-photo-grey	4.64 (0.95)	4.2 (0.83)		3.55 (1.23)
tree2-photo-colour	4.9 (0.31)	4.1 (1.07)	4.05 (1.19)	4.59 (0.59)
tree2-photo-grey	4.9 (0.31)	3.4 (1.1)		3.67 (1.46)
tree3-photo-colour	4.88 (0.34)	3.64 (1.18)	4.14 (0.94)	4.05 (1.21)
tree3-photo-grey	4.38 (1.07)	3.29 (1.27)		3.27 (1.12)
trumpet				
trumpet1-photo-colour	3.3 (1.59)	3.6 (1.14)	3.5 (0.95)	3.77 (1.11)
trumpet1-photo-grey	3.55 (1.54)	3.43 (1.01)	2.82 (0.98)	3.81 (1.17)
trumpet2-photo-colour	3.8 (1.2)	2.9 (1.3)	4.33 (0.97)	4.9 (0.31)
trumpet2-photo-grey	3.5 (1.54)	3.5 (1.01)		4.05 (1)
trumpet3-photo-colour	3.38 (1.53)	3.36 (0.85)	3.95 (1.25)	4.19 (0.81)
trumpet3-photo-grey	2.77 (1.31)	3.27 (0.83)		4.12 (1.12)
turtle				
turtle1-photo-colour	3.45 (1.43)	3.25 (1.12)	3.65 (1.14)	3.41 (1.22)
turtle1-photo-grey	3.65 (1.39)	3.43 (0.97)	2 (1.1)	3.24 (1.18)
turtle2-photo-colour	3.95 (1.19)	3.24 (1.41)	4.29 (1.01)	3.1 (1.37)
turtle2-photo-grey	3.35 (1.39)	3.68 (1.21)		2.55 (1.39)
turtle3-photo-colour	3.52 (1.44)	4.09 (0.92)	4.18 (1.1)	3.81 (1.54)
turtle3-photo-grey	3.36 (1.43)	3.35 (1.11)		3.5 (1.5)
vest				
vest1-photo-colour	4.05 (1.28)	1.95 (1.13)	1.86 (1.21)	1.85 (1.09)
vest1-photo-grey	3.73 (1.26)	2 (1.22)		2.7 (1.38)
vest2-photo-colour	4.05 (1.16)	2.05 (0.94)	2.05 (1.15)	2.2 (1.28)
vest2-photo-grey	4 (1.15)	2.1 (0.91)		2.76 (1.48)
vest3-photo-colour	3.68 (1.25)	3.05 (1.05)	2.18 (1.1)	2.33 (1.2)
vest3-photo-grey	3.23 (1.38)	2.8 (1.15)		1.95 (1.17)
violin				
violin1-photo-colour	4 (0.89)	3.7 (1.17)	4.2 (0.95)	4.47 (1.07)
violin1-photo-grey	3.36 (1.47)	3.1 (0.91)		4 (1.12)
violin2-photo-colour	3.6 (1.54)	3.4 (1.14)	4.05 (1.1)	4.5 (0.91)
violin2-photo-grey	3.7 (1.53)	3.25 (0.85)		4.19 (0.87)
violin3-photo-colour	3.08 (1.56)	3.41 (1.22)	3.77 (1.27)	4.05 (1)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
violin3-photo-grey	3.57 (1.21)	3.52 (1.33)		3.32 (1.17)
watch				
watch1-photo-colour	4.6 (0.5)	3.05 (1.05)	2.85 (1.5)	3.59 (1.18)
watch1-photo-grey	4.8 (0.7)	2.87 (0.97)	2.45 (1.63)	3.19 (1.08)
watch2-photo-colour	4.85 (0.37)	3.33 (0.97)	1.86 (1.2)	3.8 (1.06)
watch2-photo-grey	4.7 (0.66)	3.36 (1.14)		3.15 (1.31)
watch3-photo-colour	4.43 (0.81)	3.09 (1.15)	2.05 (1.17)	3.43 (1.12)
watch3-photo-grey	4.41 (0.85)	3.23 (1.02)		3.83 (1.2)
well				
well1-photo-colour	3.71 (1.1)	3.9 (1.25)	3.45 (1.36)	4.23 (1.14)
well1-photo-grey	2.95 (1.43)	3.65 (0.75)		4.1 (1.02)
well2-photo-colour	3.25 (1.52)	3.65 (1.23)	3 (1.69)	4 (1.2)
well2-photo-grey	3.95 (1.1)	3.45 (0.94)		4.05 (1.07)
well3-photo-colour	2.88 (1.42)	3.91 (0.75)	3 (1.07)	3.86 (1.08)
well3-photo-grey	3.14 (1.31)	3.67 (1.11)		3.59 (1.1)
whistle				
whistle1-photo-colour	4.05 (1.15)	2.75 (1.33)	2.9 (1.21)	4.27 (0.83)
whistle1-photo-grey	4.25 (1.21)	2.2 (1.13)	3 (1)	4.24 (0.94)
whistle2-photo-colour	4.3 (1.13)	1.95 (0.8)	2.9 (1.18)	4.35 (0.99)
whistle2-photo-grey	4 (1.08)	2.95 (1)		4.25 (1.02)
whistle3-photo-colour	3.43 (1.29)	2.18 (1.01)	3.27 (1.35)	4.29 (1.06)
whistle3-photo-grey	3.48 (1.34)	2.55 (0.96)		4.46 (1.14)
window				
window1-photo-colour	4.7 (0.57)	2.75 (1.25)	2.3 (1.22)	2.82 (1.26)
window1-photo-grey	4.9 (0.45)	2.87 (1.11)	1.91 (1.3)	2.81 (1.17)
window2-photo-colour	4.75 (0.79)	3.62 (0.97)	2 (1.14)	4 (1.12)
window2-photo-grey	4.55 (0.69)	3.45 (1.3)		3.4 (1.19)
window3-photo-colour	4.57 (0.6)	2.64 (0.9)	2.27 (1.2)	3.05 (1.32)
window3-photo-grey	4.82 (0.39)	2.64 (0.9)		3.92 (0.72)
zebra				
zebra1-photo-colour	3.41 (1.56)	3.8 (1.11)	4.75 (0.44)	4.6 (0.6)
zebra1-photo-grey	4.29 (1.06)	3.25 (1.12)		4.13 (1.17)
zebra2-photo-colour	3.9 (1.37)	3.8 (1.01)	4.55 (0.89)	4.52 (0.81)
zebra2-photo-grey	3.45 (1.5)	3.9 (1.33)		4.73 (0.55)

(continued)

Photograph	Familiarity	Visual Complexity	Colour	Mental Imagery
			Diagnosticity	
zebra3-photo-colour	4.1 (1.26)	3.62 (1.07)	4.76 (0.44)	4.23 (1.11)
zebra3-photo-grey	3.5 (1.59)	3.23 (1.23)		4.68 (0.65)