The Bizarreness Effect and Visual Imagery: No Impact of Concurrent Visuo-Spatial Distractor Tasks Indicates Little Role for Visual Imagery

Miri Besken[1] and Neil W. Mulligan[2]

[1] Faculty of Economics, Administrative and Social Sciences, Department of Psychology, Bilkent University
[2] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

Ancient as well as modern writers have promoted the idea that bizarre images enhance memory. Research has documented bizarreness effects, with one standard technique finding that sentences describing unusual, implausible, or bizarre scenarios are better remembered than sentences describing plausible, every day, or common scenarios. Not surprisingly, this effect is often attributed to visual imagery, and the effect often referred to as the bizarre *imagery* effect. But the role of imagery has been disputed even as research has found it difficult to clearly distinguish the effects of imagery from other possible bases for the bizarreness advantage. The current experiments assessed the visual-imagery hypothesis by disrupting visual imagery processes during encoding, which should reduce the bizarreness effect if it is indeed due to imagery. Specifically, one group carried out a concurrent task that selectively disrupted visual working memory (and visual imagery) during the encoding of sentences; a control group encoded the sentences without distraction. Across four experiments, the distractor task was dynamic visual noise, the spatial tapping task, and a visual span task. Each experiment found a robust bizarreness effect that was never reduced by visuospatial distraction. Combined, meta-analytic, and Bayesian analyses concurred with the results of the individual experiments. The results indicate little role for visual imagery in the bizarreness effect.

Keywords: bizarreness effect, visual imagery, visual working memory, episodic memory, vividness

The idea that bizarre images facilitate memory has a long lineage, appearing in the *Ad Herennium*, the oldest surviving Latin book (90 B.C.E.), which emphasized the use of ridiculous or fantastic images as a method to maintain memories (*Ad Herennium*, cited in Yates, 1966, pp. 9–10). Modern self-help books for memory enhancement echo this ancient advice (Buzan, 1991; Lorayne & Lucas, 1974; Lorayne, 2010), as does recent research on memory therapeutics and memory training (Dalgleish et al., 2013; Dalgleish & Werner-Seidler, 2014; Matzen et al., 2016).

Modern experimental research has shown that, at least under some conditions, bizarreness can facilitate memory (Worthen, 2006). For example, in a classic experiment, McDaniel and Einstein (1986) presented participants with mixed list of bizarre (e.g., *The dog rode the bicycle down the street*) and common sentences (e.g., *The dog chased the bicycle down the street*), and asked them to generate images for each sentence, rating each for vividness.

Results revealed superior free recall for bizarre sentences over common sentences, a result replicated in numerous studies (e.g., Geraci et al., 2013; Hirshman et al., 1989; McDaniel et al., 1995; Waddill & McDaniel, 1998). Similarly, other studies have shown that bizarre pictures (e.g., black-and-white hand-drawn images) produce higher free recall than mundane pictures (Gounden et al., 2017; Gounden & Nicolas, 2012; Marchal & Nicolas, 2000).

The superiority of recall for bizarre over common items has generally been attributed to visual imagery, and the effect often labeled the *bizarre imagery* effect (e.g., Black et al., 2012; Campos et al., 2008, 2009; Howe et al., 2000; Kroll et al., 1986; McDaniel & Einstein, 1986; Thomas & Loftus, 2002). Furthermore, some researchers have argued that the bizarre imagery effect can only be produced through instructions that induce visual imagery (Burns, 1996; Cornoldi et al., 1988; McDaniel & Einstein, 1986). For example, McDaniel and Einstein (1986) obtained the bizarreness effect when participants were given incidental visual imagery instructions (generate an image for the sentence and rate it for vividness) but not when they were given semantic elaboration instructions (judge the degree to which the relation among the three underlined words in each sentence is unusual and then rate it on a 5-point scale). There is some research that supports the idea that the bizarreness effect is eliminated when participants are asked to rely on nonvisual encoding strategies (Burns, 1996). It has been argued that when participants read bizarre sentences, they generate an image of the sentence. Because participants have to put more effort

Miri Besken ⬛ https://orcid.org/0000-0002-8024-4173
Neil W. Mulligan ⬛ https://orcid.org/0000-0002-8307-0594
Correspondence concerning this article should be addressed to Miri Besken, Faculty of Economics, Administrative and Social Sciences, Department of Psychology, Bilkent University, Ankara, Çankaya 06800, Turkey. Email: mbesken@bilkent.edu.tr

into generating a bizarre compared with an ordinary image, they have better memory for those items (Wollen & Margres, 1987). For instance, Marshall et al. (1979) asked participants to form common and bizarre visual images from word pairs and write down the visual transformations required to produce the image. They found that participants usually made more than one visual transformation to form bizarre images (e.g., changing size or function of objects, giving human qualities to inanimate objects or animals), whereas common images usually only required one visual transformation (simple interaction), confirming the notion that bizarre images require more effort and elaboration (i.e., more mental transformations) than common images. In this version of the bizarre imagery account, one would expect more imagery for bizarre than common items. Another variant of the visual-imagery theories proposes that bizarre images are more distinctive in memory compared with mundane images and produce more unique images at the time of encoding that also stand out during retrieval, despite similar levels of elaboration for bizarre and mundane images. It is the distinctiveness of the images, not the amount of elaboration, producing the higher memory performance for bizarre than mundane items (Black et al., 2012; Campos et al., 2009; McDaniel & Einstein, 1986).

However, the role of visual imagery in the bizarreness effect has been a source of controversy. Some researchers have suggested that visual imagery is not essential for obtaining the bizarreness effect because the effect can sometimes be obtained even in the absence of overt imagery instructions (Hirshman et al., 1989; Weir & Richman, 1996; Worthen, 1997, 2006). Weir and Richman (1996), for example, gave participants one of three kinds of instructions to generate common and bizarre items from noun pairs: (a) create a common or bizarre image using the nouns and focus on the image; (b) write a common or bizarre sentence using the nouns and reread the sentence until the presentation of the next word pair; or (c) create a common or bizarre image, write a sentence describing the image, and then focus on the image for the remaining time. The bizarreness effect occurred in all groups, regardless of the instructions. Similarly, Worthen (1997) presented participants with bizarre and common word pairs accompanied either by participant-generated or experimenter-provided images under conditions that either emphasized imagery or nonimagery encoding. The results yielded a main effect for bizarreness, regardless of condition, giving support to the claim that visual imagery may not be essential for superior recall of bizarre items.

Another hint for doubting the primacy of visual imagery in the bizarreness effect comes from imagery research. If forming mental images is the fundamental factor in obtaining the bizarreness effect, then it might be expected that participants with better imaging skills would show an increased bizarreness effect whereas people with poor imaging ability should show a reduced effect. Anderson and Buyer (1994) tested this possibility by assessing participants' memory for bizarre and common study items, as well as assessing their visual imagery abilities with a battery of imagery tests. The results revealed no correlation between imagery ability and recall, nor did participants with higher imagery capabilities show an increased superiority for bizarre items. In fact, those who rated images as less vivid showed a larger bizarreness effect.

On the basis of the foregoing results, some researchers have suggested that the bizarreness effect might be driven by verbal elaboration rather than visual imagery (Anderson & Buyer, 1994; Hirshman et al., 1989; Kroll & Tu, 1988; Weir & Richman, 1996;

Worthen, 1997; Worthen & Deschamps, 2008). However, investigating the role of visual imagery, and clearly contrasting these two possibilities, has been challenging because most extant research used sentences, a form of material that permits both imagery and verbal/semantic encoding. Even when participants were asked to generate common and bizarre images from nouns or noun pairs, participants were typically asked to describe the image to the experimenter as a manipulation check, leading to the use of sentences and potential verbal elaboration.

More specifically, the finding that nonimagery encoding instructions may still give rise to a bizarreness effect (e.g., Hirshman et al., 1989; Kroll & Tu, 1988; McDaniel & Einstein, 1989; McDaniel et al., 1995) does not conclusively demonstrate that imagery is not required for the effect. Participants may still spontaneously generate images even under encoding instructions that do not explicitly require them (see Worthen, 1997, 2006, for discussion). For example, a participant reading the sentence *The dog rode the bicycle down the street* might spontaneously generate the image of a dog on a bicycle. However, it is reasonable to suggest that nonimagery instructions should reduce the use of imagery during encoding compared with imagery instructions. From the perspective of the visual-imagery hypothesis, nonimagery instructions should reduce the bizarreness effect even if it does not eliminate it, which is a generally found result. There are a few studies that found no significant reduction in the bizarreness effect when imagery and nonimagery encoding instructions were compared (e.g., Weir & Richman, 1996; Worthen, 1997), which would appear to be stronger evidence against the imagery hypothesis. However, some of the studies raise their own ambiguities. For example, in Worthen (1997), the nonimagery condition (the sentence condition) produced a bizarreness effect that was numerically less than half the size of the bizarreness effects from the imagery-based conditions, despite not producing a significant interaction. Consideration of this numerical trend renders the results ambiguous in their ability to clearly refute the imagery hypothesis.

Perhaps a better way to investigate the role of imagery in the bizarreness effect would be to implement manipulations that impair visual-imagery processes.[1] If the bizarreness effect relies critically on visual imagery then disruption to visual imagery during encoding should reduce the size of the effect. Alternatively, if the bizarreness effect is unmodified by such manipulations, it appears that visual imagery does not meaningfully contribute to

---

[1] The secondary tasks used in the current experiments originate from research on the Baddeley and Hitch (1974; Baddeley, 2003) working memory model that differentiates between two limited-capacity storage systems, the phonological loop and the visuo-spatial sketchpad, along with a central executive control mechanism. Numerous studies motivated by this model have purported to show that these two subsystems can be disrupted independently of one another. For example, tasks purported to rely on the visuo-spatial sketchpad are greatly impaired by secondary tasks that require visual or spatial processing but much less affected by verbal or phonological secondary tasks (Logie, 1996; Logie et al., 1990). More recent research argues that the retention of visual information is not mediated by a separate visual store (see Morey, 2018, for review). However, for purposes of the current research, it is important to note that distractor tasks have been identified which selectively impair visuo-spatial information processing and visual imagery; the current studies do not hinge on the question of separate retention devices.

the effect. One way to assess this is to introduce a secondary task during encoding that selectively disrupts visual imagery and visuospatial information processing. A number of such tasks have been identified that do not also produce broad disruption to verbal processing or verbal aspects of working memory (such as the phonological loop).

The current study investigates the role of visual imagery in the bizarreness effect through a systematic investigation using secondary tasks that are known to disrupt visual imagery and visual working memory. These tasks, described in detail later, included dynamic visual noise, the spatial tapping task, and the visual span task. Earlier studies have tried to assess the visual imagery hypothesis by encouraging one type of processing (e.g., semantic elaboration) more than the other (e.g., visual imagery) through different instructions or materials. However, as described previously, these studies are limited in their ability to rule out visual imagery. Other research has examined correlations between the bizarreness effect and other factors, such as imaging ability (e.g., Anderson & Buyer, 1994). These studies are likewise limited because they do not experimentally assess the causal relationship proposed by the visual-imagery hypothesis.

Two earlier studies examined the effect of distraction on the bizarreness effect but neither used secondary tasks that selectively impair visual imagery or visual working memory. Worthen et al. (2000) investigated how the reaction time (RT) to a secondary task (pressing any key after the onset of "press any key" instruction during presentation of sentences) changes during the comprehension of common and bizarre sentences. However, this secondary task was not designed to selectively disrupt visual imagery. In addition, because the study was focused on the attentional requirements of encoding bizarre and common sentences, it did not include a condition with no secondary task; consequently, the influence of the secondary task on the size of the bizarreness effect could not be assessed. The second study required participants to view drawings of bizarre and common objects, accompanied by a divided attention task in which participants were asked to identify the occurrence of three consecutive odd numbers in a continuous auditory stream of digits (Gounden & Nicolas, 2012). This study is likewise not relevant as this secondary task does not influence visual working memory but primarily stresses the central executive and auditory processing in working memory.

The present experiments assess the visual imagery hypothesis in a series of four experiments. In each, participants were presented with a mixed list of bizarre and common sentences during the encoding phase of the experiment. Half the participants simultaneously carried out a secondary task that selectively impairs visual working memory and visual imagery. Subsequently, participants were given a recall test for the study sentences. Across the experiments, several different secondary tasks were used (as described later in the text). If visual imagery is the basis of the bizarreness effect, the secondary task condition should reduce or eliminate the superior memory performance for the bizarre compared with common items.

Experiment 1

Experiment 1 used dynamic visual noise (DVN) to investigate the role of visual imagery in the bizarreness effect. DVN entails a rectangular display consisting of small, flickering black and white squares on a computer screen, appearing very similar to visual static on an out-of-tune TV monitor (Quinn & McConnell, 1996). DVN is an ideal candidate as a secondary task and has been shown to disrupt visual imagery and visuospatial information processing in numerous studies. For example, DVN disrupts memory for words encoded under imagery instructions but not under rote rehearsal instructions (McConnell & Quinn, 2000; Quinn & McConnell, 1996, 1999). Particularly relevant for present purposes, the disruptive effects of DVN appear most pronounced on the formation and encoding of visual images (as opposed to their short-term retention; Quinn & McConnell, 2006). Likewise, when DVN is implemented during encoding, the concreteness effect (greater long-term memory for concrete than abstract words, assumed to be mediated by greater imagery in the concrete condition) is reduced (indeed, eliminated; Parker & Dagnall, 2009; see also Chubala et al., 2018). Similarly, Dean et al. (2005) found that DVN disrupted the generation of visual images in a size comparison task. DVN impairs visual information processing selectively, leaving verbal processing intact (McConnell & Quinn, 2000; Quinn & McConnell, 1996, 1999; Valenti & Galera, 2020). DVN likewise disrupts visual imagery processes (e.g., Baddeley & Andrade, 2000; Dean et al., 2005; Kemps, & Andrade, 2012; Smyth, & Waller, 1998; Valenti & Galera, 2020). Moreover, the dynamic visual display is meaningless and no response or verbalization is required; thus, the task requires minimal use of the central executive or verbal aspects of working memory (Quinn & McConnell, 1996). As shown in the work of Quinn and McConnell (1996, 1999, 2006) and Parker and Dagnall (2009), DVN has been successfully used in long-term memory experiments and has proven successful in disrupting imaginal processes for complex stimuli (Dean et al., 2008; Parker & Dagnall, 2019).

In the study phase of this experiment, participants were presented with a mixed list of bizarre and common sentences. Participants were asked to imagine each sentence and rated the vividness of the image. Mixed lists along with imagining and vividness ratings were used because they maximize the bizarreness effect (e.g., McDaniel & Einstein, 1986; Worthen, 1997, 2006) and provide the best opportunity for observing any potential interaction. The experimental group engaged in the DVN task while listening to the sentences over headphones. Two control groups were used in this experiment, in which the study sentences were either presented aurally or visually. The DVN task required aural presentation of the sentences because the distraction takes place in the visual modality. However, virtually all of the extant research on the bizarreness effect used visual presentation of the sentences. Consequently, the auditory control group, who listened to sentences over headphones, was implemented to provide an exact counterpart to the DVN group, and a visual control group, who read the sentences on the screen, was included to verify that the bizarreness effect was obtained with the procedures prevailing in the bulk of the prior research.

Method

Participants

Thirty-six undergraduates from the University of North Carolina (UNC) participated in exchange for course credit. This and all

subsequent experiments presented in this article were approved by The UNC Office of Human Research Ethics/Institutional Review Board (OHRE/IRB), Protocol 10–1938.[2]

Design and Materials

The design was a 2 (Sentence Type: Bizarre and Common) × 3 (Experimental Group: Visual Distraction vs. Auditory Control vs. Visual Control) mixed design, with sentence type manipulated within subjects and experimental group manipulated between subjects. The stimulus set consisted of 16 noun triplets, each used to create one bizarre and one common sentence (taken from McDaniel & Einstein, 1986), producing a total of 32 sentences. The bizarreness of the material was manipulated by presenting sentences that depict scenes that are either possible and unsurprising or impossible and surprising. For example, for the word triplet *banker*, *newspaper*, *puddle*, the common sentence was *The banker dropped the newspaper in the puddle*, whereas the bizarre sentence was *The banker floated across the puddle on a newspaper*. The 32 sentences, consisting of 16 bizarre and 16 common sentences, were further randomly divided into two mixed lists of eight common and eight bizarre sentences. These two lists were counterbalanced across subjects so that each noun triplet was presented in bizarre and common contexts equally often. For each list, two common sentences from the original McDaniel and Einstein (1986) materials were used as practice sentences.

The dynamic visual noise was the same as the technique used in Quinn and McConnell (1996) and consisted of a large rectangular display covering 80% of the computer screen and composed of small black and white squares (each of 4 × 4 pixels). Each square either remained the same or changed color (from black to white or vice versa) every .25 s with a probability of .5, giving the sense of a flickering rectangular display, with the ratio of the black and white squares kept equal throughout the presentation.

Procedure

The experiment consisted of three phases: a study phase, a distractor phase, and a testing phase. During the study phase, participants were told that the experiment was about individual differences in visual imagery in the face of distraction. They were told that they would listen to or read sentences while they tried to ignore a secondary stimulus that might be presented to them. Furthermore, they were instructed to create visual images of the sentences and rate the vividness of each image on a scale of 1 (*not at all vivid*) to 5 (*very vivid*). During the study phase, participants were presented with common and bizarre sentences in a random order, with the restriction that no more than two sentences of the same kind were presented consecutively. In line with McDaniel and Einstein (1986), participants were not informed about the nature of the sentences nor about the upcoming memory test. Participants were given two practice trials with common sentences.

In the visual distraction condition, the sentences were presented over headphones. Each trial started with a short beep for half a second, immediately followed by the DVN display, which was presented for a total of 10 s. Two seconds after the DVN was initiated, the sentence was presented over the headphones. Participants were given 8 s to encode each sentence and create a mental image, followed by a screen asking them to rate the vividness of the image within 4 s. Once they entered the rating, the program proceeded onto the next trial. The whole trial lasted up to14.5 s. The participants were instructed to watch the dynamic visual noise presented on the screen for the duration of encoding. Participants were told to look at the screen without looking in any other direction. The experimenter monitored the participants' gaze to ensure compliance with the instructions (all participants complied).

In the auditory control condition, each trial started with a beep and a blank screen. Two seconds later, the sentence was presented over the headphones. Eight seconds later, participants entered the vividness rating and then proceeded onto the next sentence. As in visual distraction condition, each trial lasted up to 14.5 s. In the visual control condition, the procedure was the same as the auditory control condition except that the sentence was presented on the screen rather than over the headphones.

After the study phase, participants were asked to solve math problems for 3 min to minimize recency effects (e.g., 67 + 46 =____). The distractor task was followed by the testing phase. Participants were asked to recall and write down as many sentences as they could remember from the encoding phase. They were also told that they should write any fragments, nouns, or verbs that they might recall, even if they did not remember the whole sentence.

Results

The vividness ratings and their RTs are reported in Table 1. Because preliminary analyses indicated no significant differences between the two control conditions in terms of vividness ratings or RTs ($Fs < 1.1$), the control conditions were combined to compare with the distraction condition. The ratings and the RTs were submitted to separate 2 × 2 analysis of variances (ANOVAs), with sentence type as a within-subjects factor and group (combined control group vs. visual distraction) as a between-subjects factor. For the ratings, the results yielded a main effect of sentence type with higher vividness ratings for common than bizarre sentences, $F(1, 34) = 10.65$, $MS_e = 1.06$, $p = .003$, $\eta_p^2 = .24$, and nonsignificant effects for group and the interaction ($Fs < 1$). For RTs, a significant effect of sentence type, $F(1, 34) = 5.46$, $MS_e = 1.958$, $p = .025$, $\eta_p^2 = .14$, indicates that vividness ratings were slower for bizarre than common sentences. The effect of group was also significant, $F(1, 34) = 4.92$, $MS_e = 22.84$, $p = .033$, $\eta_p^2 = .13$, indicating slower ratings in the visual distraction than control groups. The interaction was not significant, $F(1, 34) = .71$, $MS_e = 1.958$, $p = .407$.

---

[2] These experiments were conducted several years before submission of this paper, before a-priori power calculations were routinely reported. The sample size was chosen to be similar to sample sizes used in similar studies. Post-hoc power computations were based on effect sizes from McDaniel and Einstein (1986), the study on which these experiments were based. For the relevant conditions (mixed lists with imagery instructions), the average effect size for the bizarreness effect for noun recall and sentence access (the two primary dependent variables) were $d_z = .86$ and .93, respectively. The power to detect bizarreness effects of these sizes exceeds .95 for this and all subsequent experiments. McDaniel and Einstein also reported interactions in which the bizarreness effect was reduced by other conditions (e.g., semantic encoding instructions)—the average effect size for these interactions for noun recall and sentence access were $d = 1.18$ and 1.05, respectively. The power to detect interactions of this size exceeds .90 for noun recall and .80 for sentence access for this and all subsequent experiments.

Table 1
Study Phase: Mean Ratings and RTs (Standard Deviation)

	Vividness rating		Vividness RT (ms)	
Experiment 1	Bizarre	Common	Bizarre	Common
Visual control	2.99 (0.84)	3.58 (0.98)	839 (363)	752 (377)
Visual distraction	3.01 (0.86)	3.84 (0.67)	1,106 (251)	1,053 (315)
Auditory control	2.91 (1.04)	4.01 (0.81)	901 (362)	766 (295)

	Vividness rating		Vividness RT (ms)	
Experiment 2	Bizarre	Common	Bizarre	Common
Spatial distraction	3.36 (0.78)	4.17 (0.83)	1,201 (367)	1,169 (374)
Control	2.82 (0.89)	4.06 (0.70)	878 (343)	800 (348)

	Vividness rating		Vividness RT (ms)	
Experiment 3	Bizarre	Common	Bizarre	Common
Visual distraction	3.10 (1.02)	3.96 (0.78)	1,278 (443)	1,102 (345)
Control	2.92 (0.93)	4.03 (0.78)	968 (403)	879 (320)

	Plausibility rating		Plausibility RT (ms)	
Experiment 4	Bizarre	Common	Bizarre	Common
Visual distraction	1.64 (0.94)	4.32 (1.01)	766 (309)	752 (304)
Control	1.28 (0.26)	4.71 (0.25)	734 (328)	752 (330)

Previous research has used a few different measures of recall, although the measures generally yield very similar results. We used the two most common measures: sentence access and noun recall. The sentence access measure counts a sentence as accessed if at least one noun from the sentence has been recalled. Noun recall refers to the proportion of all of the nouns recalled. Both measures were computed separately for bizarre and common sentences. The descriptive statistics are presented in Table 2.

There were two planned analyses, one of the visual control condition to determine if the basic bizarreness effect was replicated under the standard presentation conditions, and a second comparing the visual-distraction and auditory control conditions that addresses the primary question of interest: is the bizarreness effect diminished by a concurrent task that disrupts visual imagery. However, we begin with an integrated analysis of the data in which the sentence access measure was submitted to 2 (Sentence Type) $\times$ 3 (Experimental Group) ANOVA. The analysis revealed a main effect for the sentence type, $F(1, 33) = 33.04$, $MS_e = .020$, $p < .001$, $\eta_p^2 = .50$, revealing that bizarre sentences were accessed more often than common sentences. There was no significant effect of group, $F < 1$, but the interaction was significant, $F(2, 33) = 4.14$, $MS_e = .020$, $p = .025$, $\eta_p^2 = .20$), indicating that the bizarreness effect differed in size across groups (although not in the direction predicted by the imagery hypothesis, as explicated subsequently). Noun recall was submitted to the same ANOVA and produced the same results: a significant main effect of sentence type, $F(1, 33) = 19.10$, $MS_e = .015$, $p < .001$, $\eta_p^2 = .37$, no effect of experimental group, $F < 1$, and a significant interaction, $F(2, 33) = 4.41$, $MS_e = .015$, $p = .020$, $\eta_p^2 = .21$, indicating that more nouns were recalled from bizarre than common sentences (the bizarreness effect), but that this effect differed across groups.

The first of the focused analyses demonstrated that the visual control condition produced a bizarreness effect in replication

of prior research. Specifically, both the sentence access measure, $t(11) = 4.08$, $p = .002$, $d = 1.18$, and noun recall, $t(11) = 2.79$, $p = .018$, $d = .81$, were significantly greater for bizarre than common sentences.

The second analyses compared the visual distraction and auditory control conditions by submitting both sentence access and noun recall to separate 2 (Sentence Type) $\times$ 2 (Experimental Group: Visual Distraction vs. Auditory Control) ANOVAs. The access measure revealed a main effect of sentence type, $F(1, 22) = 17.76$, $MS_e = .021$, $p < .001$, $\eta_p^2 = .45$ (the bizarreness effect), no effect of experimental group, $F < 1$, and a significant interaction, $F(1, 22) = 7.44$, $MS_e = .021$, $p = .012$, $\eta_p^2 = .25$. The analysis of noun recall produced the identical pattern, with a significant effect of sentence type, $F(1, 22) = 11.63$, $MS_e = .017$, $p = .003$, $\eta_p^2 = .35$, no effect of experimental group, $F < 1$, and a significant interaction, $F(1, 22) = 8.23$, $MS_e = .017$, $p = .009$, $\eta_p^2 = .27$. By both measures of recall, the bizarreness effect was significantly *larger* in the distraction than auditory-control condition. Additionally, in the distraction condition, tests of the simple effect of sentence type demonstrated a significant bizarreness effect for both sentence access, $t(11) = 5.90$, $p < .001$, $d = 1.70$, and noun recall, $t(11) = 5.65$, $p < .001$, $d = 1.63$, whereas the auditory control condition did not produce a significant bizarreness effect for either measure (both $ts < 1$). Finally, for completeness, the bizarreness effect was not significantly different between the visual-distraction condition and the visual-control condition. When the analyses were restricted to these two groups, the interaction between sentence type and experimental condition was nonsignificant for both sentence access, $F(1, 22) = 1.00$, $p = .33$ and noun recall, $F(1, 22) = 2.85$, $p = .11$.

Discussion

Although the vividness ratings are of secondary importance, it is important to note that the current results replicate the typical findings both that vividness ratings are lower for bizarre than common items (e.g., Burns, 1996; McDaniel & Einstein, 1986; Robinson-Riegler & McDaniel, 1994) and vividness ratings typically take longer for bizarre than common sentences (e.g., Campos et al., 2008; Einstein et al., 1989; Kroll & Tu, 1988). Vividness ratings were not affected by visual distraction but the speed of the ratings was slowed by distraction. We discuss both of these results in detail in the general discussion after examining their generality across our experiments.

More critical are the recall results, which were completely consistent whether measured with sentence access or noun recall.

Table 2
Experiment 1 Recall Results: Mean (Standard Deviation) Sentence Access and Noun Recall

	Sentence access		Noun recall	
Experimental group	Bizarre	Common	Bizarre	Common
Visual control	.54 (.12)	.32 (.15)	.39 (.09)	.26 (.14)
Visual distraction	.59 (.19)	.30 (.15)	.49 (.15)	.25 (.14)
Auditory control	.45 (.15)	.39 (.16)	.34 (.14)	.31 (.16)

Note. The visual distraction and auditory control conditions entailed auditory presentation of the study sentences; the visual control condition entailed visual presentation of study sentences.

First, the visual control condition entails the closest match to prior research using visual presentation of the study sentences. This condition faithfully replicated prior research in demonstrating a robust bizarreness effect.

Most important is the comparison between the visual distraction condition (which required auditory presentation of the study sentences with a visual secondary task) and its matched (auditory) control condition. If the bizarreness effect depends on imaginal processes, there should be an interaction between these two factors such that the advantage for the bizarre sentences decreased or was eliminated in the visual distraction condition. When the DVN condition is compared with its control group, the findings do not show a reduction in the size of the bizarreness effect. On the contrary, the concurrent presentation of DVN along with the sentences produced a robust bizarreness effect, significantly larger than in its matched (auditory) control condition (and numerically, if not significantly, larger than in the visual control condition). This experiment indicates that visual distraction at the very least does not decrease the bizarreness effect.

The finding that the bizarreness effect is not obtained in the auditory control condition is potentially interesting. Most bizarreness studies have used visual presentation of sentences, and obtained the effect often with 12–16 participants in a group. To our knowledge, there is only one published study that used auditory presentation of the sentences, and this study demonstrated a significant bizarre effect (Anderson & Buyer, 1994). However, Anderson and Buyer's (1994) study used a group of 80 participants. The bizarreness effect that they obtained in the auditory modality might be associated with the number of participants that they used. Speculatively, the fact that there are no other published studies of bizarreness conducted in the auditory modality might indicate that the effect is not generally obtained in the auditory modality especially with smaller sample sizes. It is unclear why this might be, but because the effect was not obtained in the auditory control condition, but was robust in the visual control condition, the subsequent experiments used visual presentation of the study sentences.

Experiment 2

Experiment 2 further investigated the visual imagery hypothesis in light of the results of Experiment 1. First, the outcome of Experiment 1 argues for the use of visual presentation of the sentences in the subsequent experiments to ensure robust bizarreness effects in the control condition. Critically, the other common distractor tasks for visual working memory can be used with visual presentation of the sentences. Second, although the secondary task used in Experiment 1 was an appropriate starting point for the present investigation, a secondary task which requires more active involvement from the participants, instead of the passive viewing of DVN, might provide a more rigorous test of the visual-imagery hypothesis. The current experiment used the spatial tapping task, in which participants are asked to repeatedly tap a certain sequence of keys in the shape of a figure. Spatial tapping is an ideal candidate because it has been shown to selectively disrupt visuospatial information processing but not verbal processing (e.g., Salway & Logie, 1995) and it likewise disrupts visual imagery (e.g., Baddeley & Andrade, 2000; Borst et al., 2012; Meneghetti et al., 2017; Smyth & Waller, 1998). For example, Baddeley and Andrade (2000) presented participants with visual shapes, which they later

had to imagine while performing a secondary task that selectively disrupts visual (spatial tapping) or verbal processing (counting). Participants' memory performance for the imagined shapes was disrupted by spatial tapping but not counting. Likewise, the ability to generate mental images for routes and locations is impaired by spatial tapping (Gyselinck et al., 2007; Meneghetti et al., 2017). Spatial tapping may also reduce visual working memory performance more than does DVN, and it requires repeated response from the participants (Andrade et al., 2002). Finally, the sample size of Experiment 2 was increased to provide an even more powerful test of the visual-imagery hypothesis.[3]

Method

Participants

Sixty participants from the University of North Carolina participated in the experiment in exchange for course credit or monetary compensation.

Design and Materials

Sentence type (bizarre vs. common) was manipulated within subjects and experimental group (spatial distraction and control) was manipulated between subjects. The sentence materials of Experiment 1 were used. The spatial tapping board consisted of four wooden cubes (2.5 cm on a side), arranged 14 cm apart in a diamond shape on a square wooden board. The board and spatial tapping task were modeled after the materials and procedure of Smyth et al. (1988). To make sure that the participants tapped the positions rhythmically, an audio-file with a metronome beat of 120 times a minute (two beats per second) was recorded.

Procedure

For the spatial distraction condition, the study phase was the same as Experiment 1 with two exceptions. First, the sentences were presented on the screen rather than over the headphones. Second, the dynamic visual noise was replaced with spatial tapping. In the distraction condition, each study trial started with a beep for half a second. Following this, the metronome beat began and persisted throughout the trial. The participants were instructed to tap the four pegs on the board repetitively, following a diamond shape pattern, with each tap corresponding to a single beat of the metronome. To ensure that the participants did not look at the board, a separate board was placed between the computer screen and the tapping board to block the participants' view of the tapping board and their own hand. The participant could reach around the occluding board and comfortably carry out the tapping task. Two seconds after the beep sound, the sentence was presented on the screen for the duration of 8 s (note that the tapping task began 1.5 s before the sentence was presented and continued throughout the entire trial). The

---

[3] The sample sizes, more than twice the group sizes of Experiment 1, were chosen to be substantially larger than typical experiments on the bizarreness effect. Post hoc power computations indicated power exceeding .99 both to detect a bizarreness effect and to detect an interaction with bizarreness, based on the effect sizes reported in McDaniel and Einstein (1986; see Footnote 1 for details). For effect sizes one-third smaller, the power to detect the bizarreness effect likewise exceeded .99, and the power to detect the interaction was .92 for noun recall and .85 for sentence access.

participant was instructed to read the sentence and create a mental image. Next, the vividness scale was displayed on the screen, requiring a rating within 4 s, yielding a trial lasting up to 14.5 s.

In the spatial distraction condition, before the main set of study trials, the participants first practiced the distractor task by itself for 30 s and then had two practice trials with common sentences to assure that the participant fully understood the procedure.

In the control condition, the study trial started with a beep and a blank screen, followed by the metronome beat that played throughout the trial (as in the distraction condition). Two seconds later, the sentence was presented visually. Eight seconds later, the vividness prompt was displayed and participants had up to 4 s to enter the vividness rating. As in the distraction condition, each trial lasted up to 14.5 s. The distractor and testing phases were the same as Experiment 1.

Results and Discussion

The vividness ratings and RTs from the study phase (see Table 1) were submitted to separate 2×2 ANOVAs, with sentence type as a within-subjects factor and experimental group as a between-subjects factor. As in Experiment 1, the analysis of ratings revealed a main effect of sentence type, with higher vividness ratings for common than bizarre sentences, $F(1, 58) = 86.46$, $MS_e = .36$, $p < .001$, $\eta_p^2 = .60$. Neither the effect of group, $F(1, 58) = 3.53$, $p = .07$ nor the interaction, $F(1, 58) = 3.84$, $p = .06$, were significant. The vividness RTs likewise comported with Experiment 1, demonstrating a significant effect of sentence type, $F(1, 58) = 9.22$, $MS_e = 10,009$, $p = .004$, $\eta_p^2 = .14$, and experimental group, $F(1, 58) = 14.44$, $MS_e = 246,758$, $p < .001$, $\eta_p^2 = .20$, and no interaction, $F = 1.67$, $p = .20$. That is, vividness RTs were significantly slower for bizarre than common sentences, and for the spatial distraction than control condition.

The recall results are presented in Table 3. The sentence access and noun recall measures were submitted to separate 2 (Sentence Type) $\times$ 2 (Experimental Group) ANOVAs. The access measure revealed a main effect for the sentence type, $F(1, 58) = 47.19$, $MS_e = .019$, $p < .001$, $\eta_p^2 = .45$, with the bizarre sentences accessed more often than common sentences. The main effect of experimental group was not significant, $F(1, 58) = 2.98$, $p = .090$ (though a nonsignificant numerically lower sentence access is observed in the distraction condition). The interaction was not significant, $F(1, 58) = .035$, $p = .85$. The same pattern was obtained for noun recall. The proportion recalled for was significantly higher for bizarre than common sentences, $F(1, 58) = 39.32$, $MS_e = .014$, $p < .001$, $\eta_p^2 = .40$. Neither experimental group nor its

Table 3
Experiment 2 Recall Results: Mean (Standard Deviation) Sentence Access and Noun Recall

Experimental group	Sentence access		Noun recall	
	Bizarre	Common	Bizarre	Common
Spatial distraction	.43 (.20)	.26 (.13)	.36 (.19)	.21 (.11)
Control	.48 (.13)	.32 (.15)	.38 (.12)	.26 (.14)

Note. The visual distraction and auditory control conditions entailed auditory presentation of the study sentences; the visual control condition entailed visual presentation of study sentences.

interaction with the sentence type was significant: $F(1, 58) = 1.42$, $p = .24$, for experimental group and, $F(1, 58) = .67$, $p = .42$, for the interaction.

First, there was a significant advantage for the bizarre sentences compared with common sentences. Thus, the bizarreness effect was replicated in the control condition. It is unclear why the auditory sentences of Experiment 1 failed to produce a robust bizarreness effect but it is clear that using visual sentences in the present control condition, and the visual-presentation control condition of Experiment 1, produces a substantial bizarreness effect, a prerequisite for our ability to determine if visual-spatial disruption reduces the effect. Second, despite the robust effect in the control condition, there is no evidence that the secondary task reduces the size of the bizarreness effect. Spatial tapping has been documented to selectively disrupt visuospatial processing and cause decrements in imaginal processing (e.g., Baddeley & Andrade, 2000; Salway & Logie, 1995). Given that there was no reduction in the size of the bizarreness effect, the role of visuospatial processes in the emergence of the effect should be questioned. Taken together, Experiments 1 and 2 indicate that concurrent tasks that are known to selectively impair visual working memory and visual imagery produce no measurable reduction in the bizarreness effect. In Experiment 1, the comparison between the distraction and auditory-control conditions revealed a significantly *larger* bizarreness effect under distraction. This is quite compelling evidence that the bizarreness effect is at least not smaller under visual distraction, as proposed by the visual-imagery hypothesis. Experiment 2 provides converging evidence that the bizarreness effect is not reduced by visuospatial distraction, but in this case, the evidence is in the form of a null interaction (rather than a significant interaction in the direction opposing the hypothesis). Of course, evidence from null interactions needs to be carefully scrutinized. We report additional (meta-analytic and Bayesian) analyses on this issue after reporting Experiments 3 and 4.

Experiment 3

The first two experiments used two commonly used methods for selectively impairing visual-spatial processing and in neither case did the secondary task decrease the bizarreness effect. However, despite a numerical difference in Experiment 2, these secondary tasks also did not produce an overall decrement in memory performance. This may be because visual imagery plays little role in the encoding of sentences bizarre or common. Alternatively, these secondary tasks may not have impaired overall encoding because they were either too passive or too repetitive. Perhaps the bizarreness effect only reduces in size with concurrent tasks having an active maintenance component throughout the processing of the sentences. Experiment 3 and 4 used a concurrent task that requires just such maintenance of visuospatial information.

The secondary task in Experiment 3 is the visual span task modeled after Della Sala et al. (1999; see also, Logie et al., 1990). It was implemented as follows. Participants were presented with a first visual image which had to be maintained while they encoded the sentence. After processing the sentence, a second image was presented and participants judged whether it was the same or different from the first image. The participants in the control group were exposed to the same sequence of image-sentence-image but were not asked to maintain or compare the images. The task

images comprised of checkerboard patterns filled with randomly placed black and white squares. The grid was 5 × 6 to maximize the employment of visual working memory (Della Sala et al., 1999). Consistent with the task's engagement of visual working memory, the task is known to impair visual imagery processes (e.g., Logie et al., 1990). For example, performance on the visual span task is selectively impaired by a visual imagery secondary task compared with a verbal secondary task, and in turn disproportionately impairs a visual imagery task relative to a verbal task when the span task itself is used as a distractor (Logie et al., 1990; see Della Sala et al., 1999, for similar results). Similarly, for math problems requiring visual imagery, the visual span task impairs performance more than does a phonological span task, whereas for math problems with little reliance on visual imagery, the pattern of interference reverses (Trbovich & LeFevre, 2003).

The design of Experiment 3 differed from Experiments 1 and 2 in a few ways. To begin with, this experiment had an active maintenance component which changed on each trial. Thus, this task both required participant response (in contrast to DVN in Experiment 1) and those responses could not become automatized throughout the study (as might happen with spatial tapping in Experiment 2). Moreover, because participants were tested on each trial for their maintenance of the visual pattern, it was possible to determine whether the successful maintenance of the visual image actually reduced memory performance for bizarre and common sentences on a trial-by-trial basis.

Method

Participants

Thirty-two undergraduates participated in exchange for course credit or monetary compensation.

Materials and Design

Sentence type (bizarre vs. common) was manipulated within subjects and experimental group (visual distraction vs. control) was manipulated between subjects. The sentence stimuli were the same as in the previous experiments. The material for the visual distractor task consisted of checkerboard patterns adapted from Della Sala et al. (1999). A black and white visual pattern was created by randomly filling in half of the squares in a grid of 5 × 6. Eighteen patterns were prepared (two for practice trials and 16 for experimental trials), along with a modified version of each in which the location of one of the black squares was switched with an adjacent white square, making the pattern slightly different than the original one.

Procedure

Each study trial began with a half-second beep followed by a visual matrix for 4 s. This was followed by a sentence for 8 s, the vividness rating for up to 4 s, and another visual pattern for 4 s. On a random half of trials, the second pattern was the same as the original and on the other half it was the altered version. Each trial lasted up to 20.5 s. Participants in the visual distraction group were told to keep the matrix in mind while reading the sentence and creating a mental image of the sentence. They were told that they would see another visual matrix after the vividness rating task and this matrix might be exactly the same as the first matrix or it

might be slightly changed (with a black and white square swapped). Participants were told to press "s" for same or "d" for different. Participants in the control group were told that they would see visual matrices before and after the sentence but were not otherwise told to process the matrices. These participants were not told anything about the similarity or difference between the patterns. In both the control and visual-distraction groups, participants were given two practice trials (with common sentences) to get used to the procedure. The distractor and test phases were the same as Experiments 1 and 2.

Results

The vividness ratings and RTs from the study phase (see Table 1) were submitted to separate 2 (Sentence Type: Bizarre vs. Common) × 2 (Experimental Group: Visual Distraction vs. Control) ANOVAs. The vividness ratings were again greater for common sentences than bizarre sentences, $F(1, 30) = 26.86$, $MS_e = .58$, $p < .001$, $\eta_p^2 = .47$ (the effect of experimental group and the interaction were nonsignificant, $Fs < 1$). The vividness RTs were again significantly slower for bizarre than common sentences, $F(1, 30) = 6.26$, $MS_e = 44,676$, $p = .018$, $\eta_p^2 = .17$, and for the visual distraction than control condition, $F(1, 30) = 4.64$, $MS_e = 245,893$, $p = .038$, $\eta_p^2 = .13$ (the interaction was nonsignificant, $F < 1$).

The mean proportion correct for the visual span task was .75, substantially above the chance level of .50, $t(15) = 8.08$, $p < .001$. Moreover, the successful completion of the distractor task was not different for bizarre ($M = .76$, $SD = .18$) and common ($M = .74$, $SD = .14$) trials, $t < 1$.

The recall measures of noun recall and sentence access (see Table 4) were submitted to separate 2 (Sentence Type) × 2 (Experimental Group) ANOVAs. The analysis of sentence access indicated a main effect of sentence type, with greater access for bizarre than common sentences, $F(1, 30) = 17.52$, $MS_e = .020$, $p < .001$, $\eta_p^2 = .37$. Furthermore, there was a main effect of experimental group, $F(1, 30) = 5.64$, $MS_e = .029$, $p = .024$, $\eta_p^2 = .16$, indicating that the visual distraction task reduced sentence access compared with the control condition. Critically, the interaction was not significant, $F(1, 30) = .18$, $p = .68$. The same pattern was found for noun recall: A significant main effect for sentence type, $F(1, 30) = 10.80$, $MS_e = .017$, $p = .003$, $\eta_p^2 = .27$, with greater recall for bizarre than common sentences; a significant main effect of experimental group, $F(1, 30) = 8.54$, $MS_e = .028$, $p = .007$, $\eta_p^2 = .22$), with worse recall in the visual distraction than control condition, and no interaction, $F(1, 30) = .65$, $p = .43$.

The foregoing analyses were based on the entire (unconditionalized) recall data set. Additional analyses were conducted on recall scores restricted to study trials with correct answers on the visual span task. These conditionalized recall scores produced the same pattern of results as the unconditionalized analyses (in the current experiment as well as Experiment 4). Given that the results are the same, only the unconditionalized analyses are reported (although the conditionalized means are presented in Table 4).

Discussion

Experiment 3 investigated the effect of a concurrent visual working memory task with an active maintenance component on

Table 4

Experiment 3 Recall Results: Mean (Standard Deviation) Sentence Access and Noun Recall

Experimental group	Unconditionalized scores				Conditionalized scores			
	Sentence access		Noun recall		Sentence access		Noun recall	
	Bizarre	Common	Bizarre	Common	Bizarre	Common	Bizarre	Common
Visual distraction	.33 (.18)	.20 (.11)	.22 (.14)	.14 (.10)	.29 (.18)	.18 (.13)	.20 (.14)	.15 (.11)
Control	.45 (.18)	.28 (.15)	.37 (.19)	.24 (.16)				

Note. The conditionalized scoring only applies to the visual distraction condition.

the bizarreness effect. The experiment yielded higher memory performance for bizarre compared with common sentences—the typical bizarreness effect. Moreover, experimental group had an effect on overall memory performance, with lower performance under visual distraction. The encoding of the sentences was generally impaired with the visual span task as a concurrent task. However, the interaction was not significant, indicating that the size of the bizarreness effect was not moderated by this visual working memory task. The visual span task reliably occupies visual working memory, correlates with many other measures of visual working memory, and impairs visual imagery (Della Sala et al., 1999; Logie et al., 1990). Even though this visual distractor task with active maintenance component disrupted overall memory performance, it did so proportionally for common and bizarre items. Along with the results of Experiments 1 and 2, the present results provide no evidence that visual working memory and visual imagery contribute to the bizarreness effect.

Experiment 4

Experiment 3 showed that the size of the bizarreness effect was not reduced by the concurrent visual span task even though it reduced overall memory performance. However, the use of overt imagery instructions in this and the earlier experiments may have induced participants to generate images despite the distractor task. If the bizarreness effect is caused (even partially) by visual imagery, one might still expect that the distractor tasks would reduce the efficacy of imagery construction and reduce the bizarreness effect. But, erring on the side of caution, we should consider that perhaps the directive to use imagery may have worked at cross-purposes with the visual distraction task such that participants successfully used imagery despite the distractor task. If so, perhaps removing the overt instructions to construct an image would render the concurrent distraction task more effective, leading to a reduced bizarreness effect. Alternatively, the prior experiments may have correctly reflected a lack of visual imagery input into the bizarreness effect, in which case removing the imagery instructions would not change the results observed in the earlier experiments.

In Experiment 4, the visual imagery instructions were replaced with plausibility ratings in which participants judge whether the sentences describe events that can happen in real life. Even though many bizarreness studies use vividness ratings, previous research has successfully uncovered bizarreness effects using other types of ratings. Some studies have used "bizarreness" ratings (e.g., Kroll & Tu, 1988; McDaniel et al., 1995, McDaniel & Einstein, 1989) and others have used pleasantness ratings (e.g., Worthen, 1997). The distractor task was the visual span task from Experiment 3,

which successfully reduced overall memory performance. The visual-imagery hypothesis again predicts that the bizarreness effect should be reduced under this distraction.

Method

Participants

Thirty-two participants participated in this experiment in exchange for course credit or monetary compensation.

Materials and Procedure

Sentence type (bizarre vs. common) was manipulated within subjects and experimental group (visual distraction vs. control) was manipulated between subjects. The methods were identical to Experiment 3 with one modification. The rating screen asked for "Plausibility (1–5)" instead of "Vividness (1–5)." Participants were asked to think about the plausibility of the sentences that they read and rate each with a number from 1 to 5. A rating of 1 indicated that the event mentioned in the sentence could never happen in real life and a rating of 5 indicated that it could definitely happen in real life.

Results and Discussion

The plausibility ratings and RTs from the study phase (see Table 1) were submitted to separate 2 (Sentence Type) × 2 (Experimental Group) ANOVAs. The analysis of the ratings yielded a main effect of sentence type with higher plausibility ratings for common than bizarre sentences, $F(1, 30) = 185.96$, $MS_e = .80$, $p < .001$, $\eta_p^2 = .86$, and nonsignificant effects for group ($F < 1$) and the interaction, $F(1, 30) = 2.75$, $p = .11$. The RTs for the plausibility ratings demonstrated no significant effects of sentence type, group, or interactions (all Fs < 1).

The mean proportion correct for the visual span task was .75, significantly above chance, $t(15) = 9.78$, $p < .001$, $d = .24$. Moreover, participants' successful completion of the distractor task was not different for bizarre ($M = .75$, $SD = .12$) and common ($M = .75$, $SD = .14$) sentences, $t < 1$. Sentence access and noun recall (see Table 5) were submitted to ANOVAs with the sentence type and experimental group as factors. The access measure revealed the same results as Experiment 3, with a significant main effect for the sentence type, $F(1, 30) = 15.58$, $MS_e = .017$, $p < .001$, $\eta_p^2 = .34$, a significant main effect for the experimental group, $F(1, 30) = 9.54$, $MS_e = .028$, $p = .004$, $\eta_p^2 = .24$, and no interaction $F(1, 30) = .68$, $p = .42$. The analysis of noun recall revealed one difference from the access measure: Despite a numerical difference in the predicted direction, the main effect of sentence type was not significant,

Table 5
Experiment 4 Recall Results: Mean (Standard Deviation) Sentence Access and Noun Recall

Experimental group	Unconditionalized scores				Conditionalized scores			
	Sentence access		Noun recall		Sentence access		Noun recall	
	Bizarre	Common	Bizarre	Common	Bizarre	Common	Bizarre	Common
Visual distraction	.35 (.12)	.19 (.18)	.26 (.10)	.18 (.17)	.37 (.18)	.22 (.17)	.27 (.12)	.20 (.16)
Control	.45 (.14)	.35 (.15)	.36 (.14)	.33 (.15)				

Note. The conditionalized scoring only applies to the visual distraction condition.

$F(1, 30) = 3.41$, $MS_e = .012$, $p = .075$, $\eta_p^2 = .10$. The main effect of experimental group was significant, $F(1, 30) = 8.69$, $MS_e = .029$, $p = .006$, $\eta_p^2 = .23$, with the control group performing better than the visual distraction group. The interaction was not significant, $F(1, 30) = .77$, $p = .39$. As with Experiment 3, the analyses were repeated with recall results conditionalized on correct performance on the visual span task, and the results were identical to the unconditionalized results (see Table 5).

In this experiment, we investigated the role of visual distractors paired with encoding instructions that did not explicitly require the formation of an image. Despite this change, the critical results were consistent with the findings of Experiment 3 in which the same visual distractor task was paired with explicit imagery instructions. First, there was a significant access advantage for bizarre sentences. In addition, the visual distraction condition reduced access to sentences in general. However, the size of the bizarreness effect was not reduced by visual distraction. Quite the opposite, numerically (if not significantly) the bizarreness effect was larger in the distraction than control condition. The results were similar for the unconditional and conditional access scores, indicating that even when we only take into account only those trials on which participants unambiguously maintained the visual distraction task, the access to bizarre sentences was higher than common sentences and the size of the bizarreness effect was unaffected by distraction.

The noun recall measure was somewhat less informative, failing to produce a significant bizarreness effect. Prior studies have also shown that ratings other than vividness are less likely to produce the bizarreness effect in the noun recall measure, or the effect is sometimes smaller (e.g., Kroll & Tu, 1988; Worthen, 1997; Worthen & Roark, 2002), so it is not too surprising that noun recall yielded only a numerical difference in the expected direction for the bizarreness effect. Of course, this means that the ability to detect an interaction between sentence type and group was reduced but only for the noun recall measure (and only in this one experiment). The results for sentence access are quite interpretable; the bizarreness effect found in Experiment 4 on sentence access is robust and of a comparable size to that found in Experiment 3. Consequently, the results of sentence access bear the same degree of interpretability as in earlier experiments.

Combined, Meta-Analytic, and Bayesian Analyses

With the exception of Experiment 1, in which the distraction condition produced a significantly larger bizarreness effect than its matched control condition, the evidence against the visual imagery account takes the form of null interactions between sentence type and experimental group (distraction vs. control). Although the individual experiments were reasonably powered to detect a bizarreness effect and interactions involving it (see Footnote 1), supplementary analyses provide even stronger evidence on this issue. First, pooled analyses provide quite high power. In the pooled analyses, all the distraction and control conditions of the four experiments were merged and submitted to a 2 (Sentence Type: Bizarre vs. Common) $\times$ 2 (Experimental Group: Distraction vs. Control) ANOVA. The combined analyses have power of .98 and .95 to detect an interaction effect one half the size of those found in McDaniel and Einstein (1986) for the sentence access and noun recall measures, respectively.[4] The pooled results for both the access measure and for noun recall yielded significant main effects for sentence type and experimental condition, reflecting the bizarreness effect and the generally negative overall impact of a distractor task during encoding (e.g., Mulligan, 2008). However, the critical interaction was nonsignificant for both sentence access, $F(1, 158) = 1.22$, $MS_e = .019$, $p = .27$, $\eta_p^2 = .008$, and noun recall, $F(1, 158) = 2.31$, $MS_e = .015$, $p = .13$, $\eta_p^2 = .014$. A very high-powered analysis produces no evidence for an interaction and, it should be noted, any appearance of a trend is in the opposite direction suggested by the visual-imagery hypothesis (the bizarreness effect is numerically larger in the distraction condition).

The foregoing analyses produced null effects for the critical interaction and conventional significance testing does not allow one to quantify evidence for the null hypothesis (Rouder et al., 2009). To further assess the state of the evidence in the preceding analyses, we computed Bayes Factors using the statistical software program JASP (JASP Team, 2018, jasp-stats.org; Wagenmakers et al., 2018). The Bayes Factor (BF_{01}) represents the probabilities of the null and alternative hypotheses conditionalized on observed data, that is, the posterior odds. Interpretation of the Bayes Factor is straightforward with low values ($\leq 1/3$) favoring the alternative hypothesis, high values (≥ 3) favoring the null hypothesis, and intermediate values indicating an ambiguous outcome (see Rouder et al., 2009; Table 1). In the present case, the H_A, based on the visual-imagery hypothesis, is that the bizarreness effect is larger in the control than distraction condition (note that this is a directional hypothesis), whereas H_O is the opposite (the bizarreness effect is either equal or larger in the distraction condition). For the combined analysis, $BF_{01} = 11.49$ and 13.86 for sentence access and noun recall, respectively. This constitutes strong evidence for the

---

[4] The power to detect an interaction effect just one-third that size is still appreciable at .80 and .71.

null hypothesis that the bizarreness effect is no smaller in the distraction than control condition.

Another way to combine results across experiments is with a single-paper meta-analysis (SPM), conducted here using the procedures of McShane and Böckenholt (2017).[5] The analysis produces results wholly consistent with the combined ANOVA. For the sentence access measure, the meta-analysis revealed a main effect of sentence type (SPM estimate = .31, 95% confidence interval, CI [.21, .42]), demonstrating a robust bizarreness effect; a main effect of experimental group (SPM estimate = .12, 95% CI [.02, .23]), demonstrating reduced memory in the distraction condition; and a nonsignificant interaction effect (SPM estimate = −.07, 95% CI [−.17, .04]). The results for noun recall were identical: a main effect of sentence type (SPM estimate = .20, 95% CI [.09, .31]), a main effect of experimental group (SPM estimate = .13, 95% CI [.02, .24]), and a nonsignificant interaction effect (SPM estimate = −.05, 95% CI [−.16, .06]). These meta-analyses indicate that the bizarreness effect was not reduced by visuospatial distraction. As noted, any trend is in the opposite direction. In summary, combined, meta-analytic, and Bayesian analyses all converge on the same conclusion: the bizarreness effect is not reduced by distraction impairing visual working memory and visual imagery.

General Discussion

The goal of the current project was to assess the role of visual imagery in the emergence of the bizarreness effect. There are different variants of the visual-imagery hypothesis, but the general idea is that when participants generate images for both common and bizarre items the images for the bizarre items are visually more distinctive (e.g., Black et al., 2012; Campos et al., 2009; Geraci et al., 2013; McDaniel & Einstein, 1986) or require more transformations in their creation that leads to better encoding for these items and, therefore, greater recall (e.g., Marshall et al., 1979; Wollen & Margres, 1987; see Worthen, 2006, for review). More generally, both ancient and modern writers propose that bizarre imagery enhances memory, and that the result is a bizarre *imagery* effect (e.g., Black et al., 2012; Burns, 1996; Buzan, 1991; Campos et al., 2008, 2009; Cornoldi et al., 1988; Dalgleish et al., 2013; Lorayne, 2010; Matzen et al., 2016; McDaniel & Einstein, 1986; Yates, 1966).

In the present experiments, the visual-imagery hypothesis was investigated by selectively disrupting visual working memory and visual imagery during encoding and determining if this reduced the bizarreness effect. Experiment 1 used dynamic visual noise, Experiment 2 used spatial tapping, and Experiments 3 and 4 used a visual span task. None of the distraction conditions reduced the size of the bizarreness effect, implying that the bizarreness effect has little reliance on visual working memory or visual imagery during encoding. It is important to emphasize that all these tasks are known to selectively disrupt visuospatial processing and visual imagery. DVN disrupts visual processing but leaves verbal rehearsal intact (e.g., Quinn & McConnell, 1996, 1999). Similarly, spatial tapping leads to decrements in visuospatial processing but does not interfere with free recall of words (Salway & Logie, 1995). The visual span task leads to selective decrements in processing of other visual and imagistic stimuli (Della Sala et al., 1999; Logie et al., 1990). One might be concerned that the

distractor tasks in Experiments 1 and 2 were not strong enough to disrupt long-term memory at all. This concern is partly belied by the finding that DVN during encoding reduced the size of the concreteness effect in long-term memory, attributed to DVN's disruption of imagery processing (Chubala et al., 2018; Parker & Dagnall, 2009; see Dean et al., 2008; Parker & Dagnall, 2019, for related results).[6] Regardless, even when the distractor task did disrupt overall memory performance (in Experiments 3 and 4), the size of the bizarreness effects remained undiminished.

Although all of the distractor tasks used in these experiments are known to disrupt visual working memory and visual imagery, there are differences. The DVN task is visually disruptive but does not require active response on the part of the participant. The spatial tapping task impairs visuospatial processing and requires active response, but does not place strong demands on the maintenance of information. The visual span task, in turn, requires maintenance of visual information. Critically, regardless which variant of distraction was implemented, the bizarreness effect was not reduced.

Experiments 1–3 used explicit imagery instructions (vividness ratings) to make sure that we obtained the greatest difference between bizarre and common items in the control condition to maximize the chance of observing any interactions with the experimental conditions. Experiment 4 used encoding instructions that do not explicitly focus on visual imagery to see if the sensitivity of the bizarreness effect to distraction changed. The reasoning behind this was to determine if the use of imagery instructions renders the bizarreness effect resistant to any type of distraction. The use of plausibility ratings required participants to refer to their world knowledge and long-term memory structures to indicate the level of plausibility. This manipulation does not necessarily rule out imagery processing; however, using these instructions might lessen the reliance on imagery processing during encoding and focus more on semantic processes. First, the results indicated that the bizarreness effect is obtained in the control group with instructions that do not explicitly require imagery. However, the superior performance for bizarre sentences is not as pronounced with plausibility ratings as it is with vividness ratings: the sentence access measure revealed a robust bizarreness effect, but noun recall did not produce a significant effect. More centrally, the distraction condition did not reduce the size of the bizarreness effect in this experiment. Even when participants were not instructed to create mental images, which might have been thought to override visuospatial distraction, the results did not indicate any role for visual imagery in producing the effect.

One aspect of the results deserves additional analysis: In Experiments 1 through 3, the distractor tasks did not significantly affect vividness ratings, as might have been expected for tasks that impair visual imagery. There are several points to make regarding this result. First, as reviewed earlier, the distractor tasks used in

---

[5] Specifically, sentence access and noun recall were both submitted to separate meta-analyses based on a 2 (Sentence Type: Bizarre vs. Common) × 2 (Experimental Group: Distraction vs. Control) design, with contrasts to examine the main effect of sentence type, the main effect of experimental group, and the interaction effect: (1 1 −1 −1), (1 −1 1 −1), and (1 −1 −1 1), respectively.

[6] As an aside, this implies that the concreteness effect and the bizarreness effect have different bases, the former influenced by visual imagery but not the latter.

the present experiments have been repeatedly shown to impair objective measures of visuospatial information processing in prior research; the present experiments implemented these tasks in the same ways as that prior research. Second, the time to make vividness ratings is usually taken to reflect the speed or ease of forming visual images (e.g., D'Angiulli et al., 2013; Kroll & Tu, 1988; McDaniel & Einstein, 1986; see also Hertzog et al., 2003). When vividness ratings take longer for bizarre than common items, this is taken as evidence that bizarre items are more difficult to image than common items (Campos et al., 2008; Einstein et al., 1989; Kroll & Tu, 1988; McDaniel & Einstein, 1986). This result was replicated in the present experiments. In this regard, it is noteworthy that vividness ratings took longer in the distractor than control condition in each of Experiments 1 through 3. Given the usual interpretation of vividness RTs, this is consistent with the idea that the distractor tasks impaired imagistic processing.

Third, the lack of an effect of distractor task on vividness ratings may be a consequence of the between-subjects manipulation of experimental group. The bizarreness manipulation is required to be a mixed-list within-subjects manipulation to ensure a robust bizarreness effect in recall (e.g., McDaniel & Einstein, 1986; McDaniel et al., 1995). Given that requirement, the manipulation of the distraction condition needed to be between subjects—a fully within-subject experiment would have produced too few study sentences in each cell of the design to allow reliable assessment of the bizarreness effect. The need for a between-subjects manipulation of experimental group means that each participant experiences visual imagery in only the control condition or only the distractor condition. To the extent that vividness ratings are influenced by within-subject comparisons, participants in the distractor condition, for example, may not rate their visual images as less vivid because they lack a direct comparison with visual images developed without distraction. This possibility is consistent with research in metamemory that demonstrates that some effects on metamemorial judgments are observed in within-subject designs, in which comparisons across conditions can influence ratings, but not in between-subjects designs, in which a comparison is not possible (e.g., Susser et al., 2013). Also relevant are the effects of bizarreness on vividness ratings. As noted, bizarre items usually produce lower vividness ratings that common items, a result readily observed with mixed-list within-subject manipulations of item type (as in the present experiments; e.g., Burns, 1996; McDaniel & Einstein, 1986; Robinson-Riegler & McDaniel, 1994) but that can be reduced or eliminated in pure-list and between-subjects designs (e.g., Campos et al., 2008; Kroll & Tu, 1988). It is possible that the between-subjects manipulation of distractor task in the present experiments obscured any effect of distraction on vividness ratings.

Returning to the primary result, we note that we have evaluated the imagery account by examining whether the bizarreness effect interacts with visual-imagery distraction. The standard imagery accounts in the literature lead to this predictions (McDaniel & Einstein, 1986; Worthen, 2006). However, it is important to consider a version of the imagery account that may not make this prediction. In particular, in one version of the bizarre imagery account, both bizarre and common sentences elicit imagery, but because bizarre items are elaborated more, this could give them an advantage in subsequent memory performance (Hirshman et al., 1989; Marshall et al., 1979). This account of bizarre imagery could

potentially predict no interaction if visual distractor tasks are used. In this case, imagery elaboration might be reduced to a similar degree for both bizarre and common sentences, potentially leading to equivalent forgetting across encoding conditions. However, looking at vividness ratings and RTs in Experiments 1–3 should give us a pause for this interpretation, because if vividness ratings and RTs reflect the strength of imagery, bizarre sentences constitute the weaker imagery condition rather than the *more elaborated*, *stronger* imagery condition. Participants consider bizarre images less vivid than common images, and they take a longer time to visualize them, rendering this version of bizarre imagery account unlikely. The other version of the bizarre imagery account is the *distinctiveness account* (e.g., Black et al., 2012; Campos et al., 2009; Geraci et al., 2013; McDaniel & Einstein, 1986). In this version of the bizarre imagery account, bizarre sentences are not necessarily elaborated more. Despite similar or less elaboration, bizarre sentences produce more distinctive imagery than common sentences. Visual distractor tasks are expected to disrupt the creation of these unique distinctive images, predicting a decrement in the size of the bizarreness effect when visual distractors are used. This distinctiveness account would predict an interaction. In this current set of studies, this version of the bizarre imagery account seems more defensible, as the vividness ratings and RTs do not directly refute this account.

If the bizarreness effect, at least as elicited with sentences, is not due to visual imagery, then what is the cause? The current experiments were designed to assess the visual-imagery hypothesis and so do not directly assess other views. Nevertheless, the results are consistent with the view that the effect is due to greater verbal or semantic processing in the bizarre condition. We have emphasized that the distractor tasks used are selective in disrupting visual imagery and visual working memory processes. This selectively largely leaves verbal and semantic processing unimpaired and so, according to these views, would leave the verbal and semantic processes producing the bizarreness effect largely unaffected—with the consequent expectation that the bizarreness effect would be undiminished by these forms of distraction.

Of course, the visual-imagery hypothesis has been supported by evidence other than the mere existence of the bizarreness effect. One prominent source of support is the finding mentioned earlier that imagery instructions often produce larger bizarreness effects than other encoding instructions. Indeed, there is some hint of this in the present results in which the plausibility ratings required in Experiment 4 produced only a nonsignificant numerical difference in the predicted direction on noun recall whereas the imagery instructions of the earlier experiments produced clear and robust bizarreness effects on this measure (although the bizarreness effect on sentence access in Experiment 4 was quite similar to that found in earlier experiments). If visual imagery is not essential for the bizarreness effect, why might imagery instructions enhance the bizarreness effect? We can speculate, from research on reading and sentence comprehension (e.g., Britt et al., 2014; Swets et al., 2008; Van den Broek et al., 2001), that different reading goals can influence the type and degree of lexical and semantic analysis (and syntactic analysis, for that matter) that a sentence receives. If the goal is to form a mental image, the type and degree of analysis of the sentence might be different and more extensive than if the goal is to assess the plausibility of the sentence, and, furthermore, these differences may be more pronounced for bizarre sentences. To

create a visual image of a bizarre sentence, the participant must fully understand the components of the sentence and their relations to one another to successfully create an integrated picture of the (odd) situation represented by the sentence. If the goal is simply to assess plausibility (or other attribute, like pleasantness), the reader may only need to get to the point of comprehending that the sentence represents an unusual, unlikely or unpleasant scenario, without needing to do the additional and differentiating semantic analyses required to develop an exact "picture" of the unlikely situation. From the perspective of the verbal elaboration view, imagery instructions may simply induce more extensive verbal and semantic analyses of the written sentence than does a plausibility judgment. In fact, the comparison of the reaction times for vividness versus plausibility between Experiment 3 and 4 reveal shorter reaction times for plausibility ratings than vividness ratings, even though the tasks are designed exactly in the same manner. Moreover, RT for plausibility ratings in Experiment 4 are not significantly different across sentence types, differently from previous experiments where vividness RTs are longer for bizarre than common items. From this perspective, the end product of the goal (the image itself) may be less critical to the memorial consequences than the processes used to reach that goal.

What of bizarreness effects that are produced by images rather than sentences (e.g., Gounden et al., 2017; Gounden & Nicolas, 2012)? It may be that bizarreness effects induced by actual images (rather than by sentences that might elicit mental images) differ from the present bizarreness effect and are due directly to imagery information in memory. However, for the sake of parsimony and because the two forms of bizarreness effect are already known to share at least some similarities (Gounden & Nicolas, 2012), it is important to consider the possibility of a common bizarreness effect whether induced by external images or by sentences. In the case of bizarre images (e.g., such as those used by Gounden et al., 2017), it is plausible that the bizarre image requires additional consideration on the part of the participant as they attempt to make sense of what they are seeing. This "additional consideration" may by quite similar to the additional semantic analysis (and potentially, overt verbal processing) that the elaboration view assumes operates in the case of bizarre sentences. Alternatively, this additional consideration can be a form of higher level visual elaboration that are not interfered by the type of visual distractor tasks used in the current study. As noted, these are speculations and should be subject to future investigation.

A final word relates to the long-running imagery debate regarding the necessity of positing multiple forms of mental representation (including imagistic, or depictive, representations) versus the more parsimonious proposal of a single representational format (e.g., propositional representations; e.g., Kosslyn, 1994; Pylyshyn, 1981). The present results may be situated in terms of this long debate—in the sense that an effect often attributed to imagery, the bizarreness effect, may actually not be due to imagery but rather to verbal and semantic representations (consistent with a propositional basis). However, we should be clear that although it is of interest to relate the current conclusions to long-standing issues in cognitive science, we do not draw broad conclusions against imagery-based representations from the present results. Indeed, the present research is predicated on the idea that visual imagery is a distinct representational format that can be impaired by certain distractor tasks, and that such distraction has been shown to reduce or eliminate effects based on imagery (e.g., the concreteness effect; Parker & Dagnall, 2009). In this more general sense, the current research assumes varied mental representations (including imagistic; Pearson & Kosslyn, 2015) but concludes that one particular effect often attributed to visual imagery may instead have other bases.

References

Anderson, D. C., & Buyer, L. S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *The American Journal of Psychology*, *107*(2), 207–222. https://doi.org/10.2307/1423037

Andrade, J., Kemps, E., Werniers, Y., May, J., & Szmalec, A. (2002). Insensitivity of visual short-term memory to irrelevant visual information. *The Quarterly Journal of Experimental Psychology Section A*, *55*(3), 753–774. https://doi.org/10.1080/02724980143000541

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*(1), 126–145. https://doi.org/10.1037/0096-3445.129.1.126

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Black, S. R., McCown, S., Lookadoo, R. L., Leonard, R. C., Kelley, M., DeCoster, J., Wayde, E., & Spence, S. A. (2012). Aging, imagery, and the bizarreness effect. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *19*(5), 566–591. https://doi.org/10.1080/13825585.2011.639868

Borst, G., Niven, E., & Logie, R. H. (2012). Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. *Memory & Cognition*, *40*(3), 360–372. https://doi.org/10.3758/s13421-011-0151-7

Britt, M. A., Richter, T., & Rouet, J. F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *49*(2), 104–122. https://doi.org/10.1080/00461520.2014.916217

Burns, D. J. (1996). The bizarre imagery effect and intention to learn. *Psychonomic Bulletin & Review*, *3*(2), 254–257. https://doi.org/10.3758/BF03212428

Buzan, T. (1991). *Use your perfect memory: Dramatic new techniques for improving your memory*. Penguin Publishing. https://www.penguinrandomhouse.com/books/351277/use-your-perfect-memory-by-tony-buzan/

Campos, A., Gómez-Juncal, R., & Pérez-Fabello, M. J. (2008). Experience in imagery and imagery vividness. *Imagination, Cognition and Personality*, *27*(4), 337–348. https://doi.org/10.2190/IC.27.4.d

Campos, A., Pérez-Fabello, M. J., & Gómez-Juncal, R. (2009). The importance of image formation time in image-aided methods for word memorization. *Imagination, Cognition and Personality*, *28*(4), 321–329. https://doi.org/10.2190/IC.28.4.c

Chubala, C., Surprenant, A. M., Neath, I., & Quinlan, P. T. (2018). Does dynamic visual noise eliminate the concreteness effect in working memory? *Journal of Memory and Language*, *102*, 97–114. https://doi.org/10.1016/j.jml.2018.05.009

Cornoldi, C., Cavedon, A., De Beni, R., & Baldi, A. P. (1988). The influence of the nature of material and of mental operations on the occurrence of the bizarreness effect. *The Quarterly Journal of Experimental Psychology Section A*, *40*(1), 73–85. https://doi.org/10.1080/14640748808402283

Dalgleish, T., Navrady, L., Bird, E., Hill, E., Dunn, B. D., & Golden, A.-M. (2013). Method-of-loci as a mnemonic device to facilitate access to self-affirming personal memories for individuals with depression.

Clinical Psychological Science, 1(2), 156–162. https://doi.org/10.1177/2167702612468111

Dalgleish, T., & Werner-Seidler, A. (2014). Disruptions in autobiographical memory processing in depression and the emergence of memory therapeutics. *Trends in Cognitive Sciences*, 18(11), 596–604. https://doi.org/10.1016/j.tics.2014.06.010

D'Angiulli, A., Runge, M., Faulkner, A., Zakizadeh, J., Chan, A., & Morcos, S. (2013). Vividness of visual imagery and incidental recall of verbal cues, when phenomenological availability reflects long-term memory accessibility. *Frontiers in Psychology*, 4, 1. https://doi.org/10.3389/fpsyg.2013.00001

Dean, G. M., Dewhurst, S. A., Morris, P. E., & Whittaker, A. (2005). Selective interference with the use of visual images in the symbolic distance paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1043–1068. https://doi.org/10.1037/0278-7393.31.5.1043

Dean, G. M., Dewhurst, S. A., & Whittaker, A. (2008). Dynamic visual noise interferes with storage in visual working memory. *Experimental Psychology*, 55(4), 283–289. https://doi.org/10.1027/1618-3169.55.4.283

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, 37(10), 1189–1199. https://doi.org/10.1016/S0028-3932(98)00159-6

Einstein, G. O., McDaniel, M. A., & Lackey, S. (1989). Bizarre imagery, interference, and distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(1), 137–146. https://doi.org/10.1037/0278-7393.15.1.137

Geraci, L., McDaniel, M. A., Miller, T. M., & Hughes, M. L. (2013). The bizarreness effect: Evidence for the critical influence of retrieval processes. *Memory & Cognition*, 41(8), 1228–1237. https://doi.org/10.3758/s13421-013-0335-4

Gounden, Y., Cerroti, F., & Nicolas, S. (2017). Secondary distinctiveness effects: Orthographic distinctiveness and bizarreness effects make independent contributions to memory performance. *Scandinavian Journal of Psychology*, 58(1), 9–14. https://doi.org/10.1111/sjop.12338

Gounden, Y., & Nicolas, S. (2012). The impact of processing time on the bizarreness and orthographic distinctiveness effects. *Scandinavian Journal of Psychology*, 53(4), 287–294. https://doi.org/10.1111/j.1467-9450.2012.00945.x

Gyselinck, V., De Beni, R., Pazzaglia, F., Meneghetti, C., & Mondoloni, A. (2007). Working memory components and imagery instructions in the elaboration of a spatial mental model. *Psychological Research*, 71(3), 373–382. https://doi.org/10.1007/s00426-006-0091-1

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language*, 28(5), 594–609. https://doi.org/10.1016/0749-596X(89)90015-6

Howe, M. L., Courage, M. L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology*, 36(6), 778–792. https://doi.org/10.1037/0012-1649.36.6.778

JASP Team. (2018). JASP (Version 0.9) [Computer software]. https://jasp-stats.org/

Kemps, E., & Andrade, J. (2012). Dynamic visual noise reduces confidence in short-term memory for visual information. *Cognitive Processing*, 13(2), 183–188. https://doi.org/10.1007/s10339-011-0429-5

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. MIT Press. https://doi.org/10.7551/mitpress/3653.001.0001

Kroll, N. E., Schepeler, E. M., & Angin, K. T. (1986). Bizarre imagery: The misremembered mnemonic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1), 42–53. https://doi.org/10.1037/0278-7393.12.1.42

Kroll, N. E. A., & Tu, S.-F. (1988). The bizarre mnemonic. *Psychological Research*, 50(1), 28–37. https://doi.org/10.1007/BF00309407

Logie, R. H. (1996). The seven ages of working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 31–65). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195100990.003.0002

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, 75(1), 55–74. https://doi.org/10.1016/0001-6918(90)90066-O

Lorayne, H. (2010). *Ageless memory: The memory expert's prescription for a razor-sharp mind*. Black Dog & Leventhal Publishers. https://www.blackdogandleventhal.com/titles/harry-lorayne/ageless-memory/9781603761765/

Lorayne, H., & Lucas, J. (1974). *The memory book*. Ballantine Books. https://en.calameo.com/read/005368705282ea79d0f86

Marchal, A., & Nicolas, S. (2000). Is the picture bizarreness effect a generation effect? *Psychological Reports*, 87(1), 331–340. https://doi.org/10.2466/pr0.2000.87.1.331

Marshall, P. H., Nau, K. L., & Chandler, C. K. (1979). A structural analysis of common and bizarre visual mediators. *Bulletin of the Psychonomic Society*, 14(2), 103–105. https://doi.org/10.3758/BF03329412

Matzen, L. E., Trumbo, M. C., Haass, M. J., Hunter, M. A., Silva, A., Stevens-Adams, S. M., Bunting, M. F., & O'Rourke, P. (2016). Practice makes imperfect: Working memory training can harm recognition memory performance. *Memory & Cognition*, 44(8), 1168–1182. https://doi.org/10.3758/s13421-016-0629-4

McConnell, J., & Quinn, J. G. (2000). Interference in visual working memory. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 53(1), 53–67. https://doi.org/10.1080/713755873

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1), 54–65. https://doi.org/10.1037/0278-7393.12.1.54

McDaniel, M. A., & Einstein, G. O. (1989). Sentence complexity eliminates the mnemonic advantage of bizarre imagery. *Bulletin of the Psychonomic Society*, 27(2), 117–120. https://doi.org/10.3758/BF03329915

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 422–435. https://doi.org/10.1037/0278-7393.21.2.422

McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability *Journal of Consumer Research*, 43(6), 1048–1063. https://doi.org/10.1093/jcr/ucw085

Meneghetti, C., Labate, E., Pazzaglia, F., Hamilton, C., & Gyselinck, V. (2017). The role of visual and spatial working memory in forming mental models derived from survey and route descriptions. *British Journal of Psychology*, 108(2), 225–243. https://doi.org/10.1111/bjop.12193

Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, 144(8), 849–883. https://doi.org/10.1037/bul0000155

Mulligan, N. W. (2008). Attention and memory. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference* (pp. 7–22). Elsevier. https://doi.org/10.1016/B978-012370509-9.00134-0

Parker, A., & Dagnall, N. (2009). Concreteness effects revisited: The influence of dynamic visual noise on memory for concrete and abstract words. *Memory*, 17(4), 397–410. https://doi.org/10.1080/09658210902802967

Parker, A., & Dagnall, N. (2019). Reduced impact of imagery processing on false autobiographical recollection: The effects of dynamic visual noise. *Memory*, 27(2), 163–173. https://doi.org/10.1080/09658211.2018.1495236

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National*

Academy of Sciences of the United States of America, *112*(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, *88*(1), 16–45. https://doi.org/10.1037/0033-295X.88.1.16

Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *49*(1), 200–215. https://doi.org/10.1080/713755613

Quinn, J. G., & McConnell, J. (1999). Manipulation of Interference in the Passive Visual Store. *The European Journal of Cognitive Psychology*, *11*(3), 373–389. https://doi.org/10.1080/713752322

Quinn, J. G., & McConnell, J. (2006). The interval for interference in conscious visual imagery. *Memory*, *14*(2), 241–252. https://doi.org/10.1080/09658210500210019

Robinson-Riegler, B., & McDaniel, M. A. (1994). Further constraints on the bizarreness effect: Elaboration at encoding. *Memory & Cognition*, *22*(6), 702–712. https://doi.org/10.3758/BF03209255

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Salway, A. F. S., & Logie, R. H. (1995). Visuospatial working memory, movement control and executive demands. *British Journal of Psychology*, *86*(2), 253–269. https://doi.org/10.1111/j.2044-8295.1995.tb02560.x

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *The Quarterly Journal of Experimental Psychology Section A*, *40*(3), 497–514. https://doi.org/10.1080/02724988843000041

Smyth, M. M., & Waller, A. (1998). Movement imagery in rock climbing: Patterns of interference from visual, spatial and kinaesthetic secondary tasks. *Applied Cognitive Psychology*, *12*(2), 145–157.

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, *41*(7), 1000–1011. https://doi.org/10.3758/s13421-013-0323-8

Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory & Cognition*, *30*(3), 423–431. https://doi.org/10.3758/BF03194942

Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31*(5), 738–745. https://doi.org/10.3758/BF03196112

Valenti, L., & Galera, C. (2020). Dynamic visual noise has the same effect on visual memory and visual imagery tasks. *Psychology & Neuroscience*, *13*(1), 114–125. https://doi.org/10.1037/pne0000183

van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087. https://doi.org/10.3758/BF03206376

Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*, *26*(1), 108–120. https://doi.org/10.3758/BF03211374

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Weir, D., & Richman, C. L. (1996). Subject-generated bizarreness: Imagery or semantic processing. *The American Journal of Psychology*, *109*(2), 173–185. https://doi.org/10.2307/1423271

Wollen, K. A., & Margres, M. G. (1987). Bizarreness and the Imagery Multiprocess Model. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes* (pp. 103–127). Springer. https://doi.org/10.1007/978-1-4612-4676-3_5

Worthen, J. B. (1997). Resiliency of bizarreness effects under varying conditions of verbal and imaginal elaboration and list composition. *Journal of Mental Imagery-International Imagery Association*, *21*, 167–194.

Worthen, J. B. (2006). Resolution of discrepant memory strengths: An explanation of the effects of bizarreness on memory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 65–88). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.003.0007

Worthen, J. B., & Deschamps, J. D. (2008). Humour mediates the facilitative effect of bizarreness in delayed recall. *British Journal of Psychology*, *99*(Pt. 4), 461–471. https://doi.org/10.1348/000712608X298476

Worthen, J. B., Garcia-Rivas, G., Green, C. R., & Vidos, R. A. (2000). Tests of a cognitive-resource-allocation account of the bizarreness effect. *The Journal of General Psychology*, *127*(2), 117–144. https://doi.org/10.1080/00221300009598573

Worthen, J. B., & Roark, B. (2002). Free recall accuracy for common and bizarre verbal information. *The American Journal of Psychology*, *115*(3), 377–394. https://doi.org/10.2307/1423423

Yates, F. A. (1966). *The art of memory*. The University of Chicago Press. https://doi.org/10.1086/ahr/73.1.89