# VISUAL VS. VERBAL PROCESSING: CONCURRENT MAINTENANCE, DISTRACTION, AND INDIVIDUAL DIFFERENCES

**by**
**BELGİN DERYALAR**

**Submitted to the Graduate School of Social Sciences**
**in partial fulfilment of**
**the requirements for the degree of Master of Science**

**Sabancı University**
**December 2022**

**VISUAL VS. VERBAL PROCESSING: CONCURRENT MAINTENANCE, DISTRACTION, AND INDIVIDUAL DIFFERENCES**

Approved by:

Asst. Prof. Olesya Blazhenkova ..........................................
(Thesis Supervisor)

Asst. Prof. Eren Günseli ..................................................

Assoc. Prof. Tilbe Göksun Yörük ........................................

Date of Approval: December 27, 2022

## ABSTRACT

VISUAL VS. VERBAL PROCESSING:CONCURRENT MAINTENANCE, DISTRACTION, AND INDIVIDUAL DIFFERENCES

BELGIN DERYALAR

PSYCHOLOGY M.S. THESIS, DECEMBER 2022

Thesis Supervisor: Asst. Prof. OLESYA BLAZHENKOVA

Keywords: visual memory, verbal memory, distraction, individual differences

This study aimed to investigate the recognition of concurrently encoded visual and verbal information presented under visual or verbal distraction. In four studies, we manipulated the meaning (same vs. different) and modality (visual vs. verbal) of the memory items and examined recognition performance. Specifically, we focused on congruency effects for meaning and modality, and their interactions. Additionally, we investigated the relationship between memory performance and individual differences in visual and verbal processing. We hypothesized that visual processing is more powerful than verbal, and found evidence for superior visual over verbal memory, and a greater visual disruptive effect. Next,we found congruency effects in meaning, that is meaning matching between the memory items improved their recognition, and in modality, that is memory decreased when the processed items shared the same modality. Further, we found partial support for the combined effects of meaning and modality congruency. Meaning incongruency reduced the recognition accuracy most when the distractor and maintained items' modalities were matching. We found that modality congruency impeded memory more when meaning was also incongruent. However, this combined effect was observed mostly for visual modality and accuracy measures. We did not find consistent and strong associations between memory performance and existing assessments of individual differences. However, for our memory task, we found correlations between efficiency in processing visual and verbal distractors and relevant recognition performance.

# ÖZET

GÖRSEL VE SÖZEL İŞLEMLEME: EŞZAMANLI AKILDA TUTMA, DİKKAT DAĞITMA VE BİREYSEL FARKLILIKLAR

BELGİN DERYALAR

PSİKOLOJİ YÜKSEK LİSANS TEZİ, ARALIK 2022

Tez Danışmanı: Asst. Prof. OLESYA BLAZHENKOVA

Anahtar Kelimeler: görsel bellek, sözel bellek, dikkat dağıtma, bireysel farklılıklar

Bu çalışma, görsel veya sözel dikkat dağıtma altında sunulan eşzamanlı olarak kodlanmış görsel ve sözel bilgilerin belleğe geri çağrılmasını araştırmayı amaçlamıştır. Yapılan dört deneyde, eşzamanlı olarak kodlanmış ve dikkat dağıtıcı öğelerin anlamını (aynı veya farklı) ve modalitesini (görsel veya sözel) manipüle ettik ve hatırlama performanslarını inceledik. Spesifik olarak, anlam ve modalite için uygunluk etkilerine ve bunların etkileşimlerine odaklandık. Ek olarak, hafıza performansı ile görsel ve sözel işlemlemedeki bireysel farklılıklar arasındaki ilişkiyi araştırdık. Görsel işlemlemenin sözelden daha güçlü olduğunu ve bellek üzerinde sözelden daha büyük bir bozucu etkisi olduğuna dair bazı kanıtlar tespit ettik. Beklentilerimiz ile tutarlı olarak, anlam açısından uygunluk etkisi tespit ettik, yani bellek öğeleri arasındaki anlam eşleşmesi belleği iyileştirdi ve modalite açısından, yani işlenmiş öğeler aynı modaliteyi, görsel veya sözlü, paylaştığında performans düştü. Ayrıca, anlam ve modalite uyumunun birleşik etkileri için kısmi destek bulduk. Anlam uyuşmazlığı, bellek performansının doğruluğunu en çok dikkat dağıtıcı modalite ve bellekte tutulan öğeler eşleştiğinde azalttı. Anlam da uyumsuz olduğunda modalite uyumunun belleği daha fazla engellediğini bulduk. Ancak, bu birleşik etki çoğunlukla görsel modalite ve doğruluk ölçümlerinde gözlendi. Bellek performansı ile bireysel farklılıklar arasında tutarlı ve güçlü ilişkiler bulamadık. Bununla birlikte, bellek görevi açısından, görsel veya sözel dikkat dağıtıcıları işlemlemedeki bireysel verimlilik ile görsel veya sözel bellek performansı arasında korelasyonlar bulduk.

## ACKNOWLEDGEMENTS

When I look back at my master's degree period, I am surprised at how much happiness and sadness I was able to accumulate. Seeing that the big pitfalls I went through during my life are now just small obstacles and giving me hope that the current pitfalls will also diminish. This period of my life is not only an academic development, but also a period where valuable life perspectives mature.

I would like to start conveying my thanks by mentioning Asst. Prof. Asuman Büyükcan Tetik, the person without whom I may not have started this journey. Not only during this period of my life, but since my undergraduate studies, she has never stopped supporting me. Knowing that there is an open door always gives me confidence. Academically, she has always been inspiring with her honesty, discipline and never-ending childlike enthusiasm in her research.

I am very grateful to all my friends who I worked with in the Self and Relationships Laboratory. I never felt like actually working since we were such a great team. I would like to thank İlker and Duygu, who never let their support during my painful experimental programming process, and to the members of the MACC Lab and Elif, who contributed with their energy and kindness. Special thanks to Duygu for her emotional support. Without her, I would have lost a big part of the fun in this process. Also, I thank my colleagues who welcomed me as one of them even though I had just started working, and who listened and supported me throughout this process.

I would like to express my gratitude to my jury members Asst. Prof. Eren Günseli and Assoc. Prof. Tilbe Göksun Yörük, who never hesitated to give constructive advice and always motivated me.

I would like to convey one of my most important thanks to my thesis supervisor Asst. Prof. Olesya Blazhenkova. She was very understanding and flexible throughout the process.

My biggest and most important thanks, of course, go to my family, especially my brother Ali. They were always by my side during all the difficult times, sometimes without even realizing it. Without them, I would have given up at the first obstacle. I am very lucky to have them. Alhamdulillahi Rabbil Alamin.

*To my beloved..*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATONS

# 1. INTRODUCTION

While performing daily tasks and activities, we often process visual and verbal information simultaneously. For example, while listening to someone, we are not only processing speech, but also paying attention to the facial expressions, gestures, or clothing of this person. When reading, we may not merely process the words and sentences but also look at illustrations in a book or may generate internal mental visual images, visualizing the narrative in our mind's eye. While checking the items that we see in a shopping bag, we may verbally count them to make sure we did not forget anything.

Processing visual and verbal representations simultaneously may aid or disrupt each other. This may depend on different factors, such as the possible conflict in the content of visual and verbal information or our individual traits. The current research focused on understanding how we memorize visual and verbal representations that were presented simultaneously, how these representations may influence each other during the concurrent processing and how they may be influenced by the new visual or verbal distracting information. Previous studies on visual and verbal processing have provided evidence of both differences and interactions between visual and verbal processing. They comprise a variety of approaches including neuroscience research examining neural underpinnings of visual vs. verbal processing, cognitive psychology investigating visual vs. verbal memory and perceptual representations, as well as psychometric assessment of individual differences in visual and verbal abilities.

## 1.1 Differences Between Visual and Verbal Processing

Neuropsychological studies demonstrated that verbal and visual information processing is underpinned by different neural circuits (Gazzaniga 2004; Thierry 2006). Visual information processing recruits visual cortical regions within a hierarchy. Simple features such as orientation (Hubel and Wiesel 1962), color (Johnson, Hawken, and Shapley 2001), and binocular disparities (Barlow, Blakemore, and Pettigrew 1967) are processed by the neurons in the primary visual cortex (i.e., V1). V1 projects visual information to higher-order cortical areas through the ventral (from the primary visual cortex to inferior temporal lobe) and the dorsal (from the primary visual cortex to inferior parietal lobe) streams respectively that are involved in the processing of physical properties of visual information such as color, texture, and form, and processing of spatial location and orientation of visual information, respectively (Baars and Gage 2010; Cabeza and Nyberg 2000; Gazzaniga 2009; Mishkin, Ungerleider, and Macko 1983).

On the other hand, processing core components of language architecture (sound, syntax, and meaning/semantic; (Jackendoff 2000) recruits a different network in the brain. In particular, language processing is underpinned by such regions as the supramarginal gyrus (Dehaene-Lambertz et al. 2005; Jacquemot et al. 2003) and the region anterolateral to Heschl's gyrus in the superior temporal gyrus (Obleser et al. 2007) for sound, the lateral anterior temporal lobe (Humphries et al. 2007) and the anterior superior temporal gyrus (Friederici and Kotz 2003; Friederici and Frisch 2000) for syntax and left middle temporal gyrus (MTG), the supramarginal gyrus (SMG), and left inferior temporal gyrus (Kotz et al. 2002; Poldrack et al. 1999; Vigneau et al. 2006; Visser, Jefferies, and Lambon Ralph 2010) for meaning (Friederici 2011; Kemmerer 2014, for a review).

An influential Dual Coding Theory, proposed by Paivio 1971, suggested that visual and verbal information processing occurs through separate channels and each channel creates distinct types of representational units: imagens (for visual information) and logogens (for verbal information). According to this theory, the mental codes of visual and verbal representational units organize the information that can be stored, manipulated, and retrieved for subsequent use. There are structural and functional distinctions between the processing of visual and verbal information. The imagery system deals with a representation of visual information, which organization into higher-order units is spatial-based, and it involves parallel processing. The verbal system is specialized for processing information in a sequential manner. In a similar vein, the multi-component memory model (Baddeley, Hitch, and Allen 2021; Baddeley and Hitch 1974) regards working memory as consisting of subsystems that are responsible for processing and maintenance of verbal (phonological loop) and visual (Visio-spatial sketchpad) information. A phonological loop is considered an 'inner ear' that holds spoken and written information, while the visio-spatial sketchpad processes and maintains visual and spatial information around the environment to help to locate ourselves in relation to other objects that surround us. These two subsystems work independently but are monitored and coordinated by the same executive component, i.e., executive control.

## 1.2 Interactions Between Visual and Verbal Processing

Despite the differences, there are mutual connections and interactions between visual and verbal processing. There is evidence of overlapping neural underpinnings and shared cognitive resources, e.g., during simultaneous processing in working memory. WM has a limited capacity and not all items are efficiently maintained in WM (Logie, Camos, and Cowan 2020, for a review), thus when the resources are shared within the same modality, the conflict may occur. Besides, research suggests that visual and verbal representations may transform the information, and therefore the meanings may influence each other.

Neuroimaging studies have suggested that there are not only distinct but also overlapping regions (e.g., visual cortex, V1, V2, posterior temporal gyri, the inferior frontal gyrus) that are responsible for visual and verbal processing (Bonner and Epstein 2021; Heilbron et al. 2020; Shinkareva et al. 2011; Thierry and Price 2006). A meta-analysis study conducted by Wang et al. 2010 revealed that the left precuneus, parahippocampal gyrus, posterior cingulate, fusiform gyrus, and culmen are activated stronger during the processing of concrete word/sentence compared to abstract concepts. The activation in such regions is linked with the imageability of concrete concepts since regions such as parietal and occipital lobes (Kosslyn, Ganis, and Thompson 2001; Sack et al. 2005), posterior cingulate (Johnson et al. 2006; Kilts et al. 2004), left fusiform gyrus and culmen (D'Esposito et al. 1997; Ganis, Thompson, and Kosslyn 2004; Mestres-Missé et al. 2008) are also activated during mental image generation.

Mazoyer et al. 2002 reported different activations in response to concrete (a left predominant ventral pathway and a left prefrontal activation) and abstract (bilateral superior temporal language area activation) concepts. The neural underpinning of visual processing corresponds to verbal information, so, visual areas are activated in response to both visual and verbal information, though language-related brain areas prioritize verbal information.

In Paivio's Dual Coding Theory, the interaction between distinct types of representations was defined as a referential connection that describes the communication between the units of verbal and nonverbal subsystems. This communication between the subsystems enables memory to generate the verbal equivalent of visual information and the name of the visual information (1971). Research showed that reading evokes visual representations (Zwaan et al. 1998), and generation of visual images in reading enhances reading accuracy and comprehension (Commodari et al. 2020).

Interaction between language and imagery makes transformation of the information from one to the other possible (i.e., visually representing verbal information and vice versa), even sometimes inevitable (Mazoyer et al. 2002; Postle and Hamidi 2007; Wickens 1973), and one coding supports the other (Lewis-Peacock, Drysdale, and Postle 2015). The transformation between visual and verbal codes is explained by the 'concreteness effect', suggesting that recognizing and remembering concrete information is faster than for abstract concepts (Kroll and Merves 1986; Schwanenflugel 2013) because the imageability of concrete concepts is greater than the abstract ones, e.g., visualizing "hammer" is easier than visualizing "freedom" (Wang et al. 2010). Emprical evidence showed that individuals remember concrete pictures (Bellhouse-King and Standing 2007) and concrete words that evoke visual representations better than abstract words (Mazoyer et al. 2002; Yui, Ng, and Perera-WA 2017). Dual coding theory asserts that individuals perform better in such tasks because in case one code is forgotten, accessibility of the other code (i.e., dual coding) might be remained (Paivio 1965).

While transformation between visual and verbal information could aid memory, at the same time, conflicting meaning representations might have detrimental effects on performance. The famous Stroop experiments (1935) might be one of the examples that provided evidence of interference between visual and verbal processing. People are slower reading the written color names that mismatch with the color of the ink (e.g., the word "purple" in blue ink) than when they match (e.g., the word "purple" in purple ink). Moreover, different variations of Stroop tasks showed that not only the color of the ink negatively affects reading the color name, but also the meaning of the written word negatively affects the identification of incongruent visual color. Thus, verbal meaning may influence visual perception and vice versa. Notably, the interference of conflicting words at the time for naming ink colors was greater than the interference of conflicting colors at the time for reading words.

The greater influence of word meaning on the ink color than vice versa could be explained by reading being more automatic than color naming (Luo 1999). Another explanation was related to an inability to ignore irrelevant information at different salience levels (Durgin 2000) since the salience of the stimuli is reduced by extra distraction, the Stroop interference is reduced (Kahneman and Chajczyk 1983). In this regard, literature suggested that perceptual encoding of visual information is richer (Stenberg 2006), whereas words are restricted to letters and orthographic conventions (Nelson, Cermak, and Craik 1979). This greater physical variability of the pictures compared to words makes pictures more distinctive and salient, and easier to be recalled or recognized (Van der Cruyssen et al. 2020).

The interference between the visual and verbal modalities may occur because attention has limited capacity (Cowan, Saults, and Blume 2014) and the simultaneous use of storage and processing functions of WM leads to interference (Doherty et al. 2019; Rhodes et al. 2019). It should be noted that Embedded-processes Approach to WM (Cowan 1988) proposing a unitary system that processes the information based on features such as colors, orientations and abstract ideas, suggests that the interference may not be always severe, because the control mechanisms help the maintained items to be moved to the activated LTM so that the load on the items that are in the focus of attention is reduced. Such a mechanism helps one to protect maintained information from external distraction that is currently the focus of attention.

Research showed that concurrent processing of visual and verbal information in working memory (as in Stroop) costs more when the retained items share more features, e.g., color, orientation, semantic content (Baddeley and Hitch 1974; Cocchini et al. 2002; Cowan and Morey 2007; Fougnie and Marois 2006). Thus, not only similarity/dissimilarity in semantic meaning but also format and featural similarity may affect memory for concurrently processed items. For instance, modality congruency (Bae and Luck 2019; Kim, Kim, and Chun 2005; Oberauer et al. 2018) was shown to have a detrimental effect on memory performance.

Previous literature showed that the congruency effect influences the speed and accuracy of the performance and is determined both by the modality type (Bae and Luck 2019). Both the semantic (meaning) and format (modality) of the processed representation may affect memory, as in the case of the concurrently processed items. Moreover, they may have a combined effect, as the cognitive load is increasing due to an increased conflict, thus reducing memory performance. Concurrent processing of visual (e.g., red rectangle) and verbal Stroop (e.g., the word “red” in blue ink) items was found to be affected more by the modality-similar distraction compared to -different distraction i.e., verbal memory distraction increased interference more than spatial distractor (Kim et al., 2005). That distraction was found to be more profound when the maintained items and the distractors highly match, i.e., modality congruency effect (Oberauer et al. 2018).

It should be noted that cross-modality interference is also possible, as verbal, or auditory tasks were found to reduce visual WM performance when they were given in a retention period (Bae and Luck 2019; Makovski, Shim, and Jiang 2006). As suggested by Baddeley and Hitch 1974, in a dual task when the tasks are from different modalities (i.e., one verbal one visual), interference is possible since two slave systems are coordinated by shared executive control. Moreover, the interference may be augmented when visual and verbal items are similar.

In addition to modality-based interference during concurrent processing, an outside distractor might also interfere with the processing of visual and verbal information. The effect of the distractor in encoding and maintenance phases was found to be reduced when the memory was highly loaded (Bollinger et al. 2009; Konstantinou et al. 2014; Roper and Vecera 2014; Rose et al. 2005). The reason behind this protection of maintained items from the distractor information explained by increased focal-task engagement (i.e., devoting more attentional resources to the maintenance) and reduced engagement with the irrelevant information (Simon et al. 2016; Sörqvist and Marsh 2015; Sörqvist et al. 2016).

### 1.3 Individual Differences in Visual vs. Verbal Skills

The extent to which one can maintain multiple items from different modalities and ignore the distraction is related not only to the load involved in the given concurrent tasks but also to one's success in distributing attentional resources (Lavie 2010, for a review). For instance, individuals with low working memory capacity were found to be more prone to processing irrelevant distractors (Vogel, McCollough, and Machizawa 2005), that is, they are unsuccessful in focal task engagement. Furthermore, literature raised questions about whether people are similar in the way they attend to and memorize visual and verbal information. Can individual differences in visual imagery vs. verbal processing abilities predict visual vs. verbal memory performance? Can people with different visual and verbal skills benefit differently from text-based vs. visual-based instruction? For many decades, this discussion has been popular in psychological, educational, and psychometric literature (Kozhevnikov, Evans, and Kosslyn 2014; Richardson 1977; Sadler-Smith and

Riding 1999). More recently, it attracted the interest of cognitive neuroscientists.

Cognitive and learning styles models contrasted visual and verbal processing on the level of individual differences in abilities, strategies, and habitual preferences. According to the Visualizer-Verbalizer cognitive style model proposed by Richardson (1977), there are individual differences in people's abilities and preferences for verbal and visual strategies. That is, visualizers tend to think in pictures and apply visual thinking strategies, while verbalizers tend to think in words and rely on verbal strategies. However, Visualizer-Verbalizer model was criticized since learners with visual or verbal preferences did not perform well on the related cognitive ability tasks (Kollöffel 2012; Mayer and Massa 2003), e.g., visualizers did not outperform verbalizers on visual tasks. The instruments assessing visual vs. verbal cognitive styles were criticized for low internal reliability and low predictive validity (Antonietti and Giorgetti 1998; Boswell and Pickett 1991; McAvinue and Robertson 2007). The research examined assessments of visual vs. verbal styles in relation to performance on visual and verbal tests and found inconsistent results. Subsequently, the visual-verbal model was revised based on neuroscience evidence that distinguished between object and spatial visual processing (Mishkin, Ungerleider, and Macko 1983) and included 3 dimensions of style: visual-object, visual-spatial, and verbal (Blazhenkova and Kozhevnikov 2009). This model demonstrated improved psychometric properties and predicted performance measures better than the 2-dimensional model.

Neuroscience research provided some evidence that individual differences in visual and verbal processing can be reflected on a brain level. For example, Shin and Kim (2015) found that during the Stroop task, individuals with verbal cognitive style showed increased activation in the left dorsolateral prefrontal cortex, left fusiform gyrus, and left precuneus. A MEG study demonstrated that while visual areas showed greater activation among visualizers, frontal language areas in the middle frontal and left inferior gyri showed greater activation among verbalizers (Nishimura et al. 2015, 2020).

Other studies provided evidence supporting neural efficiency. That is, the high object-processing ability was associated with efficient use of visual-object resources, and this efficient use of resources lead to reduced neural activity in the object-processing regions (Motes, Malach, and Kozhevnikov 2008). Better spatial ability was associated with less neural activity in task-relevant brain regions (Lamm et al. 1999; Vitouch et al. 1997). Reichle et al. (2000) showed that individuals with better verbal ability had less activation in task-relevant Broca's area of the brain, while those with better visual–spatial skills had less activation in the left parietal cortex, associated with spatial processing. These results not only demonstrate the difference between the networks of cortical regions supporting verbal, visual-spatial, and visual-object processing, but also demonstrated the differences in neural underpinnings of individual differences in verbal, visual-spatial, and visual-object abilities, and suggested that they are associated with using strategies minimizing cognitive workload. Relatedly, individuals with extremely low visual imagery were found to activate a more widespread set of brain regions than those with high vividness (Fulford et al. 2018). Recent research also found other neural signatures of visual imagery vividness extremes, also known as "aphantasia" and "hyperphantasia" (Milton et al. 2021).

### 1.4 Our Research

Our research aimed to investigate the effects of visual and verbal distractions on the concurrent maintenance of visual and verbal information in working memory. In particular, we tested the following hypotheses:

**(H1) Visual vs. Verbal Modality Effects** (Visual processing is more powerful than verbal):

***H1A.Predominant Modality*** (Superior recognition performance of visual over verbal information)

Based on previous research that demonstrated the picture superiority effect, i.e., better learning and retaining of visual over verbal information (Bevan and Steger 1971; Shepard 1967; Thibodeau, Levy, and de Lemos 2021), we expected superior memory performance for visual over verbal information during the concurrent maintenance. Additionally, we aimed to compare memory for visual over verbal information when it is functioning as a distractor. Similarly, we expected that the visual distractor would be better memorized than the verbal one.

***H1B.Distracting Modality*** (Visual distraction has greater detrimental effects on memory of maintained items than verbal distraction)

If visual representation is more powerful than a verbal one, then we would observe a greater negative influence of visual than verbal distraction. Thus, we expected that visual modality distraction is more detrimental to memory performance. Specifically, we expected overall lower recognition performance under visual distraction than under verbal one. To our knowledge, even though the overall effects of visual and verbal distraction influence on visual vs. verbal processing were investigated in the previous research (Bae and Luck 2019; Kim, Kim, and Chun 2005; Oberauer et al. 2018), these disruptive effects of visual vs. verbal distraction on concurrent visual-verbal processing were not directly compared.

**(H2) Congruency Effects** (Memory depends on congruency of the processed information):

***H2A.Meaning Congruency*** (Memory is better for congruent than for incongruent information, i.e., meaning matching between the memory items improves their recognition).

Previous literature showed that meaning incongruency (Kiyonaga and Egner 2014; MacLeod 1991; Pan et al. 2022; Stroop 1935; Thierry and Price 2006) have a detrimental effect on memory performance. Therefore, we expected to observe a poorer memory performance when the meanings of the concurrently maintained or distractor items are incongruent than when they are congruent.

***H2B.Modality Congruency*** (Memory is decreased when the processed items share the same, visual or verbal, modality, i.e., the distractor matches with the modality of the maintained item).

Previous literature showed that distraction becomes more profound when the maintained items and the distractors highly match, i.e., congruency effect (Oberauer et al. 2018). This congruency influences the speed and accuracy of the performance (Jha, Fabian, and Aguirre 2004; Yoon, Curtis, and D'Esposito 2006) and is determined by the modality type (Bae and Luck 2019). Modality congruency (Bae and Luck 2019; Kim, Kim, and Chun 2005; Oberauer et al. 2018) was found to have a detrimental effect on memory performance. In our research, we manipulated the modality of the distractor. Since the modality of the concurrently maintained items was always different, we only examined the effects of distractor modality congruency on the concurrently maintained items. We expected to observe a modality congruency effect, i.e., a decreased memory performance when the additional information was processed in the same modality: visual or verbal. In particular, we expected that the visual distractor disrupts visual memory more than verbal memory and vice versa. As modality-specific cognitive resources are shared, the same modality conflicting distractor should impede performance more than the other modality conflicting distractor. Therefore, we expected to observe worse memory performance when modalities of memorized items are the same.

***H2C.Combined Effects of Meaning and Modality Congruency*** (Meaning incongruency reduces performance most when the modality of the distractor and maintained items are matching).

The efficiency of concurrent processing of semantically conflicting information coming from visual and verbal channels may be further affected by the matching/mismatching of these modalities. For example, a greater Stroop interference was observed with the modality-similar distraction (verbal load) compared to -different distraction (spatial) when people were asked to compare the meaning of the Stroop items (color of the patch vs. meaning of the colored word) (Kim, Kim, and Chun 2005). Tikhonenko et al. (2021) also showed that when the meaning of the objects was different, but the modalities were the same, the memory was impaired the most. When both meaning and modality congruency effects are in effect, the cognitive demand is increased. Therefore, we expected that combined effects of meaning, and modality congruency should further impede memory performance.

**(H3) Individual Differences** (Individual differences in visual and verbal processing are related to visual and verbal memory performance):

***H3A.Assessments of Individual Differences in Visual and Verbal Processing*** (Higher scores on visual processing assessments are related to better visual memory performance, whereas higher scores on verbal processing assessments are related to better verbal memory performance).

People vary in their memory capacities and styles and habitual preferences. Their success in the attentional adjustment to task-relevant information is related to their working memory capacity (Unsworth and Engle 2005). Additionally, self-reported visual-object, visual-spatial and verbal styles were found to predict relevant memory performance (Koć-Januchta et al. 2017; Kraemer et al. 2017; McCunn and Cilli-Turner 2020; Milz et al. 2016), e.g., verbal working memory performance found to be related to verbal styles. In this regard, we expected to observe a correlation between participants' performance on the memory task and individual difference measures, as well as the engagement with the distractor (processing and duration).

Generally, we hypothesized that people with greater visual or verbal abilities would benefit from having more cognitive resources in task-related domains, and thus show better performance in memorization of concurrent information. We expected that better domain-specific abilities such as better visual-object imagery or fluency in using language would be related to visual and verbal recognition performance. Additionally, based on the findings that individuals with low working memory capacity are more prone to processing irrelevant distractors (Vogel, McCollough, and Machizawa 2005), we expected to observe relationships between visual and verbal working memory and the effects of distraction.

***H3B. Distractor Processing Efficiency*** (Individual efficiency in processing the visual or verbal distractor is related to visual or verbal memory performance).

Similarly, we expected to observe the relationships between the individual differences in visual vs. verbal distractor processing efficiency (subjective ratings and time spent attending to a distractor) and visual vs. verbal recognition performance. However, this question was rather exploratory.

We conducted four experiments to test these hypotheses. In the first experiment, we compared concurrently processed visual and verbal memory performance under visual-object, visual-spatial and verbal distractions. In Experiment 2, we removed all distractors and tested the differences between the recognition of the concurrently processed visual and verbal memory items. Experiment 3 was a pilot experiment that tested the procedure implemented in Experiment 4. It included only one condition and used only visual distraction. Participants received either memory or imagery instruction, and we investigated how these different instructions affect the performance. Additionally, we examined the possible differences in reported strategies. In the most comprehensive Experiment 4, we compared concurrently processed visual and verbal memory performance under visual and verbal distraction. Unlike Experiment 1, Experiments 3 and 4 introduced a greater featural similarity between the concurrently maintained items and the distractor. Moreover, all experiments examined visual vs. verbal memory performance in relation to individual differences measures. See the detailed design for all the experiments in Figure 1.1 .

Our memory task generally included 1) encoding phase, when participants were concurrently presented with either congruent or incongruent visual and verbal information, 2) distraction phase, when participants, while maintaining the initially encoded items, were presented with a new distracting visual or verbal task, and 3) recognition phase, when participants recalled previously learned information.

Figure 1.1 Research Design for the Experiments

| | Concurrent Items<br>Distractor Items | Meaning Congruency | Individual Differences Measures | DV |
|---|---|---|---|---|
| **Experiment 1** | Text + Picture<br><br>Visual-object<br>Visual-spatial<br>Verbal | 1/2 Concurrent-Congruent<br><br>1/2 Concurrent-Incongruent | *Distractor Task Performance:*<br>Visual-spatial<br>Visual-object<br>Visual-object, visual-spatial, and verbal style questionnaire | Picture Retrieval Accuracy<br><br>Text Retrieval Accuracy |
| **Experiment 2** | Text + Picture<br><br>No distraction | 1/2 Concurrent-Congruent<br><br>1/2 Concurrent-Incongruent | Visual-object, visual-spatial, and verbal style questionnaire | Picture Retrieval Accuracy<br><br>Text Retrieval Accuracy |
| **Experiment 3** | Text + Picture<br><br>Visual-object | All Incongruent | Strategy use questionnaire<br><br>Distractor Task Performance | Picture Retrieval Accuracy and RT<br><br>Text Retrieval Accuracy and RT<br><br>Distractor processing efficiency (Vividness and RT) |
| **Experiment 4** | Text + Picture<br><br>Visual-object<br>Verbal | 1/3 All-congruent<br><br>1/3 All-incongruent<br><br>1/3 Distractor-incongruent | Visual-spatial<br>Visual-object<br>Verbal<br>Visual and verbal WM<br><br>Distractor Task Performance | Picture Retrieval Accuracy and RT<br><br>Text Retrieval Accuracy and RT<br><br>Distractor Retrieval Accuracy and RT<br><br>Distractor processing efficiency (Vividness and RT) |

# 2. EXPERIMENT 1

Experiment 1 aimed to investigate differences in performance for concurrently maintained picture and text memory. The picture and text memory performance were tested under different distractions, visual-object, verbal and visual-spatial. We investigated both the influence of concurrently maintained picture and text memory and the possible influence of distractions on this concurrent processing.

## 2.1 Method

### 2.1.1 Participants

Participants were students recruited via Sabancı University SONA-system (Sona Systems, https://sona-systems.com/default.aspx) and received bonus course credits for their participation. All the participants received consent forms. Fifty of them were excluded due to three reasons: having color blindness (N =1), incorrect answers to all three attention check questions (N = 47), and proceeding with the experiment with devices other than a computer (N = 3). Eventually, the data from 70 participants (52 Female, 18 Male; Mage= 21.81, SDage= 1.78) were included in the analysis.

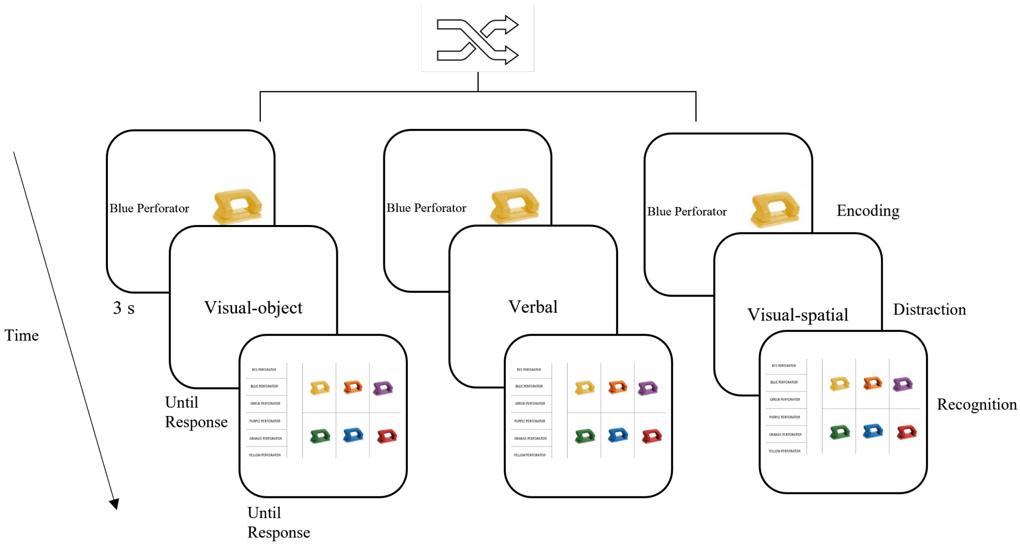
### 2.1.2 Materials

#### 2.1.2.1 Memory Task

This task was designed to assess memory for visual vs. verbal information and their possible interference and how it can be affected by the distractors of different modalities. Each trial consisted of encoding, distraction, and recognition phases (Figure

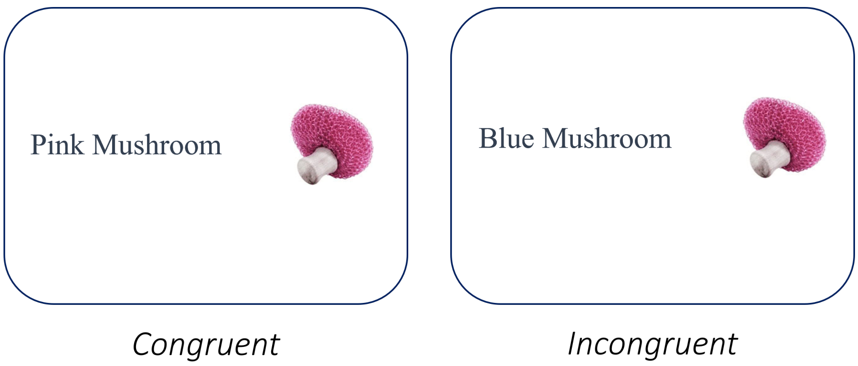
2.1). During the encoding phase, participants memorized two items simultaneously presented on the left and on the right of the screen. The stimuli were presented in a visual (picture) and verbal (text) format and referred to a simple object and its color (e.g., “blue iron”). The colored object stimuli set was taken from Brady et al. (Brady et al. 2013); https://bradylab.ucsd.edu/stimuli.html).

Figure 2.1 Experiment 1 - Memory Task trial structure: encoding, distraction, and recognition phases




For the purposes of this study, we only included six colors: blue, purple, red, green, orange, and yellow. The contrasts of red and orange objects were sharpened to avoid confusion between the colors. The location of the picture and text was manually counterbalanced across trials and trials were randomized. There were 72 trials, half of which were congruent in which the color of the object matched with the name of the color in the text (congruent), and half of them were incongruent (incongruent). The incongruent matches were balanced across six colors (Figure 2.2).

Figure 2.2 Experiment 1- Memory Task conditions: congruent and incongruent



The encoding phase was followed by a visual-object (Animals' Tails Task, ATT), verbal (Object Spatial Imagery and Verbal Questionnaire, OSIVQ), or visual-spatial (Mental Rotation Test, MRT) distractor. Each type of distractor was presented 24 ( of the trials) times, in random order. The recognition phase was composed of picture recognition and text recognition, presented in random order and counterbalanced across trials. During the recognition phase, text and picture items, for each of them, participants selected the answer among the 6 alternatives representing 6 different colors.

#### 2.1.2.2 Individual Difference Measures

The distractors included 24 Animal Tails Task items (Blazhenkova et al. 2022), and 48 survey questions, 45 were questions from Object-Spatial Imagery and Verbal Questionnaire (Blazhenkova and Kozhevnikov 2009), 24 Mental Rotation Task items (Peters et al., 1995 redrawn by Vanderberg  Kuse, 1978) and 3 items in OSIVQ were attention check questions: "I have grandchildren and they do not like playing lego", "I have three eyes and one mouth", "Currently I am doing this study". The distractor tasks served as different types of distraction (visual-object, verbal and visual-spatial) during the maintenance of the stimulus presented in the encoding phase. Additionally, we examined scores on these individual differences measures in relation to our main experimental variables.

OSIVQ (Blazhenkova and Kozhevnikov 2009) is a questionnaire assessing individual differences in abilities, preferences, and experiences in object imagery (i.e., mental visualization of pictorial object properties such as color, shape, and texture), spatial imagery (i.e., mental visualization of spatial relations and transformations, and verbal information processing (i.e., fluency in manipulating verbal information, using language). Participants rated their agreement with 45 statements describing object visualization (e.g., "I have a photographic memory"), spatial visualization (e.g., "I have excellent abilities in technical graphics"), and verbal processing (e.g., "I have difficulty expressing myself in writing") on a scale from 1= "totally disagree" to 5 = "totally agree". Blazhenkova and Kozhevnikov 2009 reported the internal reliabilities for the object, spatial, and verbal scales as .83, .79, and .74, respectively.

MRT (Peters et al., 1995 redrawn by Vanderberg  Kuse, 1978) measures spatial visualization ability. The test includes 24 items, in which participants have to find the two rotated versions of a given 3D abstract shape among the four options. Scores are calculated by summing the number of correct answers in which both of the correct answers are selected. Before the test, participants read the instructions and had 4 example trials with answer keys. They could not proceed until they complete this phase. Vandenberg and Kuse (1978) reported Cronbach's alpha of the test as .88.

ATT (Blazhenkova et al. 2022) aims to test the ability to mentally visualize and scale missing parts of objects. This task is based on a similar task used by Farah et al. (1988), but the modified version uses semi-covered visuals rather than verbal descriptions. Participants are presented with 36 visuals of animals with covered tails and asked to determine whether an animal's tail is long or short relative to its body size. Long is defined as a tail being longer than half of a body length; short is defined as a tail being shorter than half of a body length. Before the actual test trials, participants had two example trials with instructions. They could not proceed until they complete this phase. Cronbach's alpha for the test, based on our data was .61.

### 2.1.3 Procedure

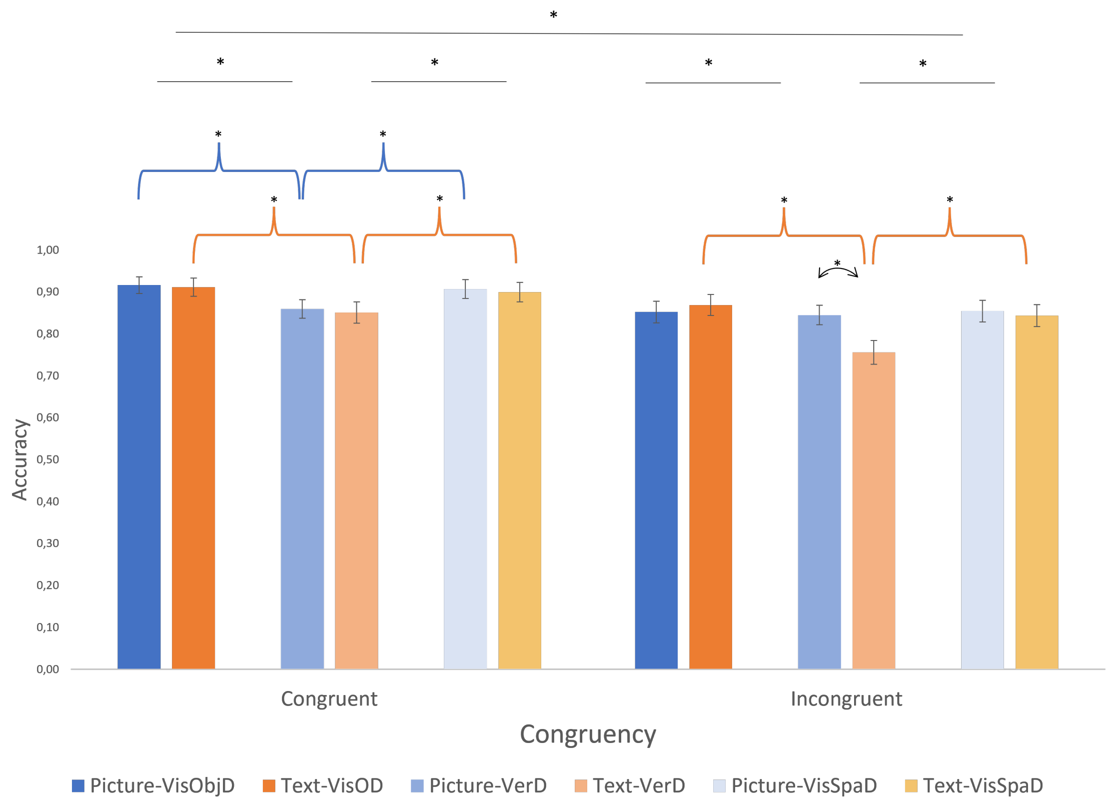
Participants completed the study online on their own computers, via Qualtrics (Qualtrics, Provo, UT) and accessed the experiment link via SONA systems. They were not assisted or watched by the experimenter. They approved the consent form and were instructed that they were required to use their PCs instead of phones. They read the instructions, completed sample trials for MRT and ATT, and proceeded with the Memory Task that included either one MRT, one ATT, or two survey items within each trial. Following the completion of the study, credits were granted to those who participated in the study for compensation.

## 2.2 Results

### 2.2.1 Recognition Type x Distractor Modality x Congruency: Accuracy

Repeated measures ANOVA was performed to reveal the effects of recognition type (picture, text), distractor modality (visual-object, verbal, visual-spatial), and meaning congruency (incongruent, congruent) on the accuracy of recognition performance (Figure 2.3).

Figure 2.3 Experiment 1- Accuracy for the recognition of picture and text items under visual-object, verbal, and visual-spatial distraction, on different congruency conditions




The effect of recognition type was significant, F (1, 69) = 8.915, p = .004, $\eta_p^2 = .114$. Picture recognition accuracy was greater than text recognition accuracy (Mdiff = .017, SE = .006), p = .004.

There was a significant effect of distractor modality, F (2, 68) = 18.803, p < .001, $\eta_p^2 = .214$ so that participants' scores under verbal distraction condition were lower than under visual-object distraction (Mdiff = .060, SE = .011), p < .001, and visual-spatial distraction (Mdiff = .049, SE = .011), p < .001, conditions. There was a significant effect of meaning congruency, F (1, 69) = 23.454, p < .001, $\eta_p^2 = .254$, so that participants' scores on meaning-congruent trials were greater than incongruent trials (Mdiff = .054, SE = .011), p < .001.

The interaction between recognition type and distractor modality was also significant, F (1.823, 125.756) = 14.148, p < .001, $\eta_p^2 =.170$. Picture recognition accuracy was lower under verbal distraction than under visual-object distraction (Mdiff = .032, SE = .011), p = .003, and visual-spatial distraction (Mdiff = .029, SE = .012), p = .016, while there was no difference in picture recognition accuracy under visual-spatial and visual-object distraction p = .713. The text recognition accuracy was also lower under verbal distraction than under visual-object distraction (Mdiff = .087, SE = .014), $p < .001$, or visual-spatial distraction (Mdiff = .068, SE = .013), $p < .001$. There was no significant difference in text recognition accuracy under visual-object vs. visual-spatial distractions p = .071, that is text recognition accuracy tended to be greater under visual-object than under visual-spatial distraction.

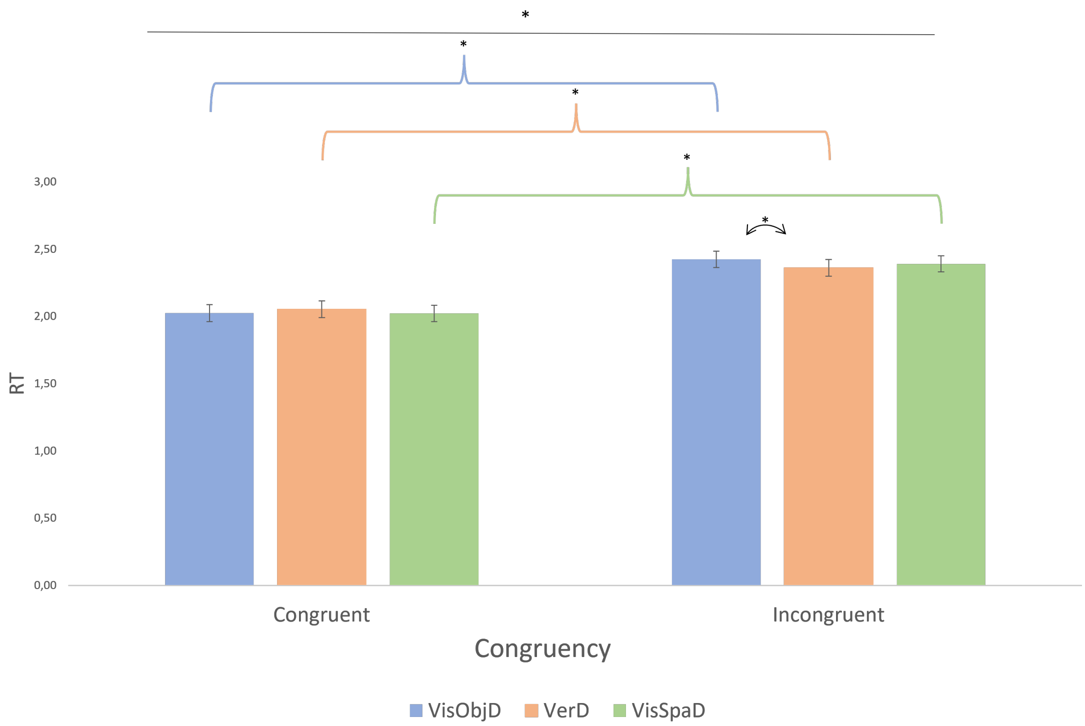
The interaction between distractor modality and congruency was not significant, F(2, 68) = .002, p = .998, $\eta_p^2 = =$ .001. The interaction between recognition type and congruency was significant F(1, 69) = 4.937, p = .030, $\eta_p^2$ =.067, suggesting that picture recognition accuracy was greater than text recognition accuracy (Mdiff = .028, SE = .010), p = .005 on incongruent trials. No other significant differences were observed.

There was a significant three-way interaction between recognition type, distractor modality, and congruency, F(1.707, 117.759) = 11.480, $p < .001$ $\eta_p^2 = .143$, that is the interaction between the recognition type and distractor modalities differed depending on the congruency. On the congruent trials, text recognition accuracy did not differ from picture recognition accuracy under visual-object, visual-spatial, or verbal distractors, all p's >.05. On the incongruent trials, picture recognition accuracy was significantly greater than text recognition accuracy under verbal distraction (Mdiff = .089, SE = .018), $p < .001$. No other differences were observed.

### 2.2.2 Congruency x Distractor Modality: RT

We performed repeated measures ANOVA to reveal the effects of distractor modality (visual-object, visual-spatial, verbal), and meaning congruency (incongruent, congruent) on the speed of recognition performance (Figure 2.4). As the participants were asked to simultaneously indicate the concurrently maintained picture and text items in the recognition phase, separate RT comparison for picture and text recognition was not possible.

Figure 2.4 Experiment 1- Reaction time for the recognition of picture and text items under visual-object, verbal, and visual-spatial distraction, in different congruency conditions.




Note. RT for recognition is combined for text and picture items.

There was no significant effect of distractor modality, F(2, 138) =.515, p = .599, so participants' speed of recognition under visual-object, visual-spatial, or verbal distractors did not differ. There was a significant effect of meaning congruency, F (1, 69) = 92.077, p < .001, $\eta_p^2 = .572$, so that participants were faster on congruent trials than incongruent trials (Mdiff = .358, SE = .037), p < .001.

The interaction between distractor modality and meaning congruency was not significant F(2, 138) = 2.311, p = .103, suggesting that the effect of visual-object, verbal, and visual-spatial distractors on the speed of the performance did not differ across different meaning congruency conditions.

### 2.2.3 Individual Differences

We performed Pearsons' correlational analysis to check whether recognition performance was related to individual differences in visual-object, visual-spatial, and verbal processing. Results showed that neither task score was correlated with individual difference measures (Table 2.1).

Table 2.1 Experiment 1- Correlations between individual difference measures and memory performance

| | Congruent Picture Recognition accuracy | Congruent Text Recognition accuracy | Incongruent Picture Recognition accuracy | Incongruent Text Recognition accuracy |
|---|---|---|---|---|
| Animal Tails Test | -.12 | -.11 | -.11 | -.08 |
| OSIVQ visual-object | .02 | .03 | .02 | .06 |
| OSIVQ visual-spatial | -.03 | -.02 | -.04 | -.04 |
| OSIVQ verbal | -.04 | -.02 | .01 | .04 |
| Mental Rotation Test | -.02 | -.03 | .04 | -.04 |

Note. * $p$ <.05, ** $p$ <.01.

## 2.3 Conclusions

In the first experiment, H1 was only partially supported. We found that picture recognition accuracy was greater than text recognition accuracy. However, verbal distraction was more detrimental to the performance than visual distraction. Thus, visual modality did not appear to be more powerful in all respects. Further in support for H2A, meaning congruency effect was observed in the data. We found that when the meaning of the maintained items was congruent, the performance was better both in terms of accuracy and reaction time. However, modality congruency effect (H2B) was only observed for verbal modality. That is, verbal distractor was more detrimental to text than picture memory. Additionally, in partial support of H2C, we observed that verbal modality congruency effect was enhanced when meaning is also incongruent. However, no such combined effect of meaning and modality congruency was observed for visual modality. Finally, we did support H3, as we did not observe significant correlation between individual difference measures and the memory task performance. The results of Experiment 1 in relation to the hypotheses are summarized in Figure 2.5.

Figure 2.5 Summary of the Results of Experiment 1

| **H1. Visual vs. Verbal Modality Effects** | | **H2.Congruency Effects** | | | **H3. Individual Differences** | |
|---|---|---|---|---|---|---|
| ***H1A. Predominant Modality*** | ***H1B. Distracting Modality*** | ***H2A. Congruency Effect for Meaning*** | ***H2B. Congruency effect for Modality*** | ***H2C. Combined Effects of Meaning and Modality Congruency*** | ***H3A. Assessments of individual Differences in Visual and Verbal Skills*** | ***H3B. Assessments of Distractor Processing Efficiency*** |
| In support of the picture superiority effect, we found superior recognition accuracy for visual over verbal concurrently processed information.<br><br>*RT comparison was not possible due to task design.* | Scores under verbal distraction were lower than under visual-object or visual-spatial distraction, suggesting that verbal distraction was more detrimental than visual to the accuracy of memory.<br><br>The recognition speed under visual-object, visual-spatial, or verbal distractors did not differ. | Recognition accuracy on congruent trials was greater than on incongruent trials.<br><br>Consistently, RT was faster on congruent trials than on incongruent trials. | The text recognition accuracy was lower under verbal distraction than under either visual-object or visual-spatial distraction (verbal modality congruency effect).<br><br>However, picture recognition accuracy was also lower under verbal distraction than under visual (both, object and spatial) distraction (reversed visual modality congruency effect).<br><br>Additionally, text recognition accuracy was lower than picture recognition accuracy under verbal distraction (verbal modality congruency effect), but no such difference was observed under any of the visual distractors.<br><br>*RT comparison was not possible due to task design.* | On the meaning incongruent trials but not on meaning congruent, text recognition accuracy was significantly lower than picture recognition accuracy under verbal distraction (verbal modality congruency effect is enhanced when meaning is also incongruent).<br><br>No parallel effect under visual distractor.<br><br>*RT comparison was not possible due to task design.* | No significant correlations between individual differences measures (OSIVQ, MRT, ATT) and memory task performance. | *The analysis was not possible due to task design.* |

We must acknowledge several limitations of our study. First, we did not measure RT separately for picture and text recognition, which prevented the separate analysis of picture vs. text speed of recognition. Second, we observed the ceiling effect, as the task appeared relatively easy. Besides, the distractors from different modalities included very different tasks (both, in terms of task nature, and possible difficulty and completion time), that also differed from the primary memory task. This limits the comparisons of distractor modality effects. Notably, while OSIVQ and MRT are validated measures of imagery and verbal processing, ATT was a novel task.

# 3. EXPERIMENT 2

Experiment 2 aimed to investigate differences in performance for concurrently maintained picture and text memory. Unlike in Experiment 1, we did not include any distraction tasks between the encoding and recognition phases.

## 3.1 Method

### 3.1.1 Participants

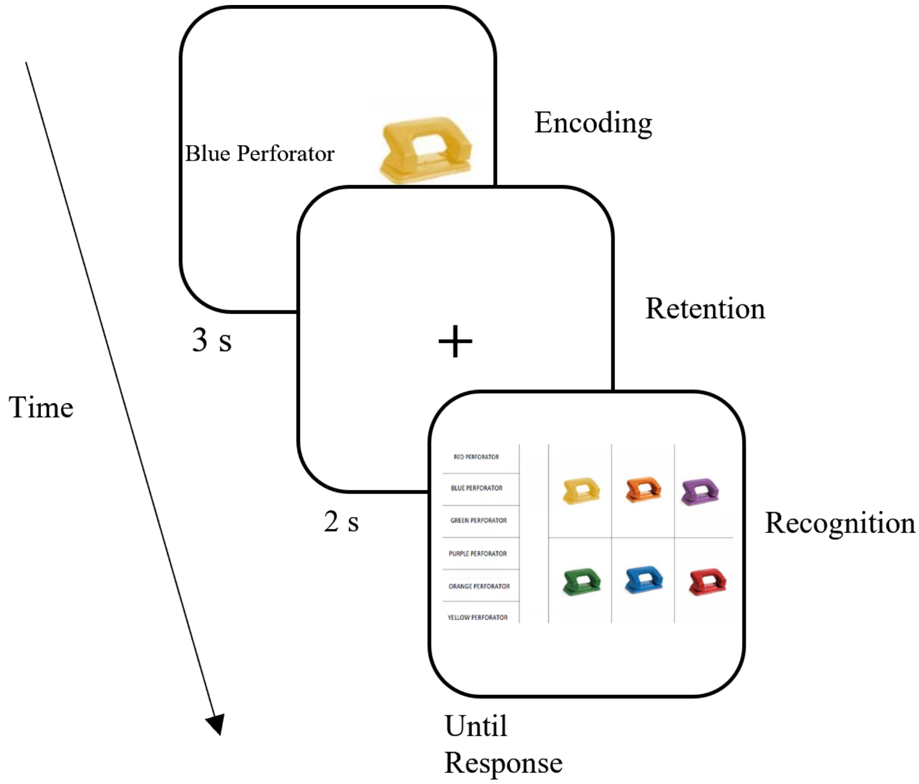
Participants were students recruited via Sabancı University SONA-system (Sona Systems, https://sona-systems.com/default.aspx) and received bonus course credits for their participation. All the participants received consent forms. Participants completed the study online, via Qualtrics (Qualtrics, Provo, UT). We excluded participants who are colorblind (N = 0), who did not use a computer (N = 2), and whose reaction times exceeded one standard deviation above and below the mean since reaction time analysis was sensitive to the outliers (N = 6). Eventually, 177 participants (121 Females, 56 Males; Mage= 21.68, SDage= 1.71) were included in the analysis.

### 3.1.2 Materials

#### 3.1.2.1 Memory Task

This task was the same as in the first study, except that we replaced the distractor task with a fixation cross presented for 3 seconds (Figure 3.1).

Figure 3.1 Experiment 2- Memory Task trial structure: encoding, retention, and recognition phases




Note. The original language of the task was Turkish.

### 3.1.3 Procedure

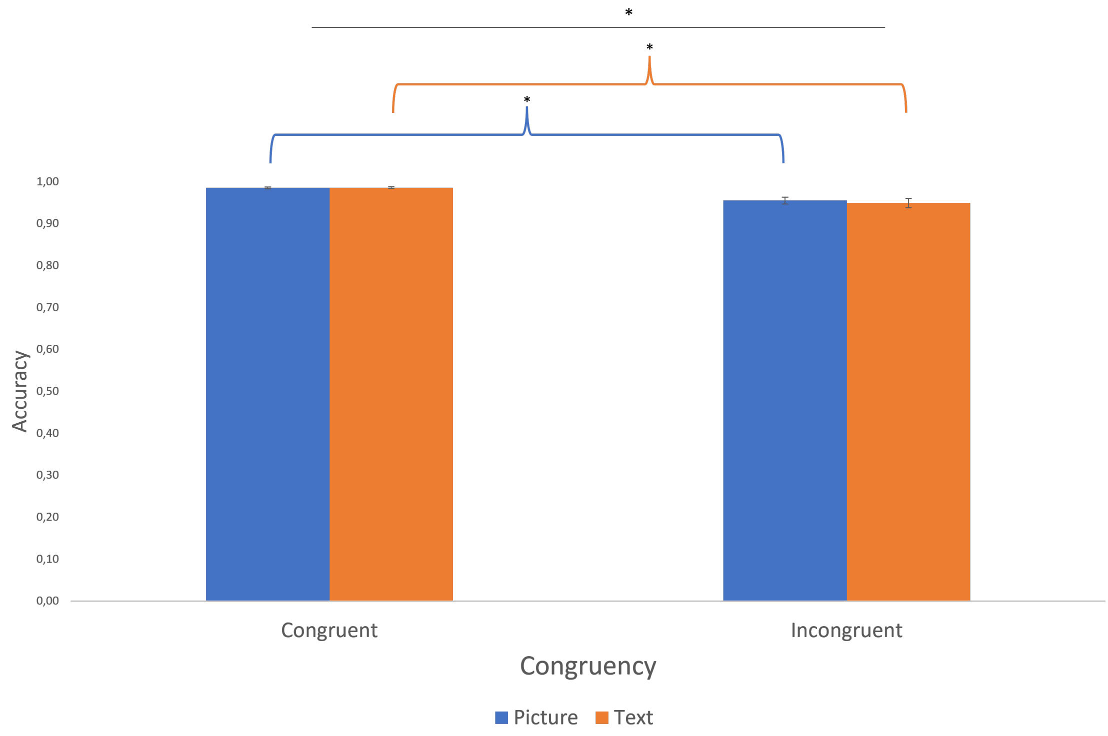
The same general procedure as in the first study was followed. After the completion of the Memory task, participants were asked to complete an individual difference questionnaire assessing visual-object, visual-spatial, and verbal processing (the OS-IVQ).

## 3.2 Results

### 3.2.1 Picture vs. Text Memory: Accuracy

Repeated measures ANOVA was performed to reveal the effect of meaning congruency (congruent, incongruent) and recognition type (picture, text) on the accuracy of recognition performance. There was a significant effect of meaning congruency, F (1, 176) = 19.717, p < .001, $\eta_p^2 = .101$. Participants' scores on all-congruent trials were significantly greater than on incongruent trials (Mdiff = .033, SE = .007), p < .001. There was no significant effect of recognition type, p = .660. The interaction between meaning congruency and recognition type was not significant, p = .611 (Figure 3.2).

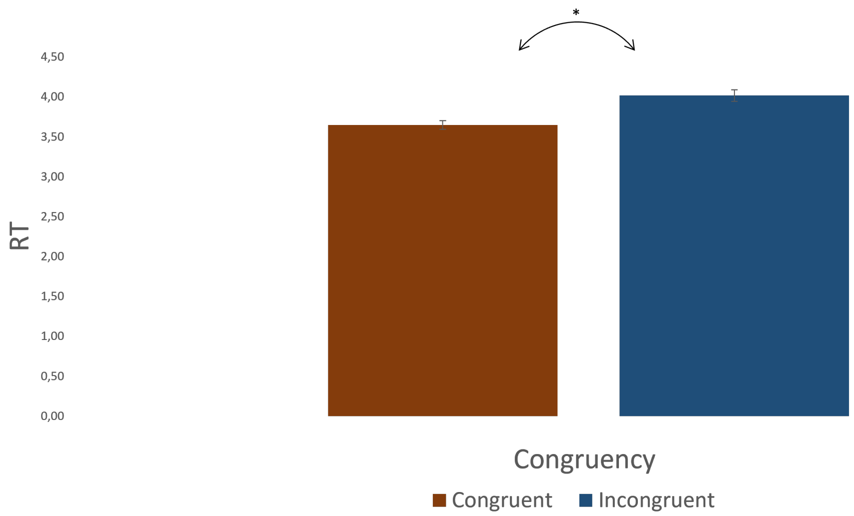
Figure 3.2 Experiment 2- Accuracy for the recognition of picture and text items on different congruency conditions




### 3.2.2 Congruent vs. Incongruent Conditions: RT

A paired sample t-test was performed to compare the total RTs on congruent and incongruent trials. Results revealed that reaction time was faster on congruent trials than on incongruent trials (Mdiff = .368, SE = .053), t (176) = -6.894, p < .001 (Figure 3.3).

Figure 3.3 Experiment 2- Reaction time on different congruency conditions




### 3.2.3 Individual Differences

We performed a correlational analysis to check whether recognition performance was related to individual differences in visual-object, verbal and visual-spatial processing. Results showed that neither of the subscales of OSIVQ was correlated with task performance (ps = .054) (Table 3.1)

Table 3.1 Experiment 2- Correlations between individual difference measures and memory performance

| | Congruent Picture Recognition accuracy | Congruent Text Recognition accuracy | Incongruent Picture Recognition accuracy | Incongruent Text Recognition accuracy |
|---|---|---|---|---|
| OSIVQ visual-object | .10 | .01 | .05 | -.01 |
| OSIVQ visual-spatial | .09 | -.06 | .02 | -.02 |
| OSIVQ verbal | -.13 | -.03 | -.09 | -.08 |

Note. * $p$ <.05, ** $p$ <.01.

## 3.3 Conclusions

In the second experiment, we did not observe superior picture memory performance in the absence of the distraction, thus H1 was not supported. We replicated our findings in Experiment 1 regarding the meaning congruency effect both in terms of RT and accuracy. Thus, H2A was consistently supported. Similarly, we did not observe significant correlations between the individual difference measures and the memory task performance. H3 was not supported. The results of Experiment 2 in relation to the hypotheses are summarized in Figure 3.4.

Figure 3.4 Summary of the Results of Experiment 2

| **H1. Visual vs. Verbal Modality Effects** | | **H2.Congruency Effects** | | | **H3. Individual Differences** | |
|---|---|---|---|---|---|---|
| ***H1A. Predominant Modality*** | ***H1B. Distracting Modality*** | ***H2A. Congruency Effect for Meaning*** | ***H2B. Congruency effect for Modality*** | ***H2C. Combined Effects of Meaning and Modality Congruency*** | ***H3A. Assessments of individual Differences in Visual and Verbal Skills*** | ***H3B. Assessments of Distractor Processing Efficiency*** |
| No superiority effect was observed in the accuracy of recognition.<br><br>*RT comparison was not possible due to task design.* | *There was no distraction in this experiment.* | Recognition accuracy on congruent trials was greater than on incongruent trials.<br><br>Consistently, RT was faster on congruent trials than on incongruent trials. | *The comparison was not possible because there were no distractors.* | *The comparison was not possible because there were no distractors.* | No correlation between the subscales of OSIVQ (visual-object, visual-spatial, verbal) and task performance. | *There was no distraction in this experiment.* |

Our experiment had several limitations. First, we did not measure RT separately for picture and text recognition due to task design. Second, the accuracy approached the ceiling. The current task appeared to be even easier compared to Experiment 1, as there was no distraction and the interval between the encoding and recognition was rather short (2 sec).

# 4. EXPERIMENT 3

Experiment 3 aimed to investigate differences in performance for concurrently maintained picture and text memory under visual distraction, on different instruction conditions. In this experiment, we separately measured speed and accuracy of recognition to address the limitations in the previous experiments. We increased the task difficulty by reducing stimulus presentation duration and increasing featural similarity between the maintained item and the distractor. The picture and text memory performance were tested under visual distraction and the represented meaning of all memory items were different (all-incongruent condition). This experiment was conducted to test the procedure for the following, most comprehensive Experiment 4.

## 4.1 Method

### 4.1.1 Participants

Participants were students (57 Female, 17 Male) recruited via Sabancı University SONA-system (Sona Systems, https://sona-systems.com/default.aspx) and received bonus course credits for their participation, or via in-person invitation and received no compensation. All the participants received consent forms. Due to experimenter error, the information about ages was not collected. However, as the participants came from the same subject pool, as for other studies (students of Sabanci University), their ages or other demographic characteristics are not expected to differ.

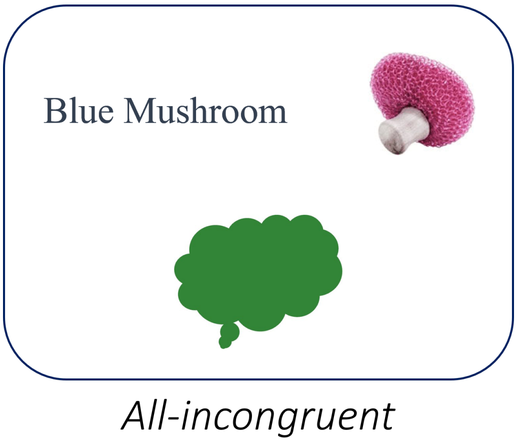
### 4.1.2 Materials

#### 4.1.2.1 Memory Task

Our task assessed the impact of visual distraction on the concurrent maintenance of visual and verbal working memory representations. This task was modified from Experiment 1 in several ways. First, to increase the complexity of the task which appeared to be quite easy, the concurrent text and picture items were presented more briefly (1.75 instead of 3 seconds). Second, it included a visual distractor that matched the picture item not only in modality, but also in format. In Experiment 3, the distractor item was exactly in the same format as the maintained item and the participants were asked to engage in manipulation of the maintained item, whereas, in Experiment 1, the distraction was a completely different task. Li and Cowan (2022) concluded that modality is not the only factor leading to distraction, but the degree of feature similarity (even within the same modality) also leads to distraction in working memory (e.g., Farrell Oberauer, 2014; Nairne, 1990; Oberauer Lin, 2017). Third, we manipulated imagery vs. memory instruction during the distraction phase.

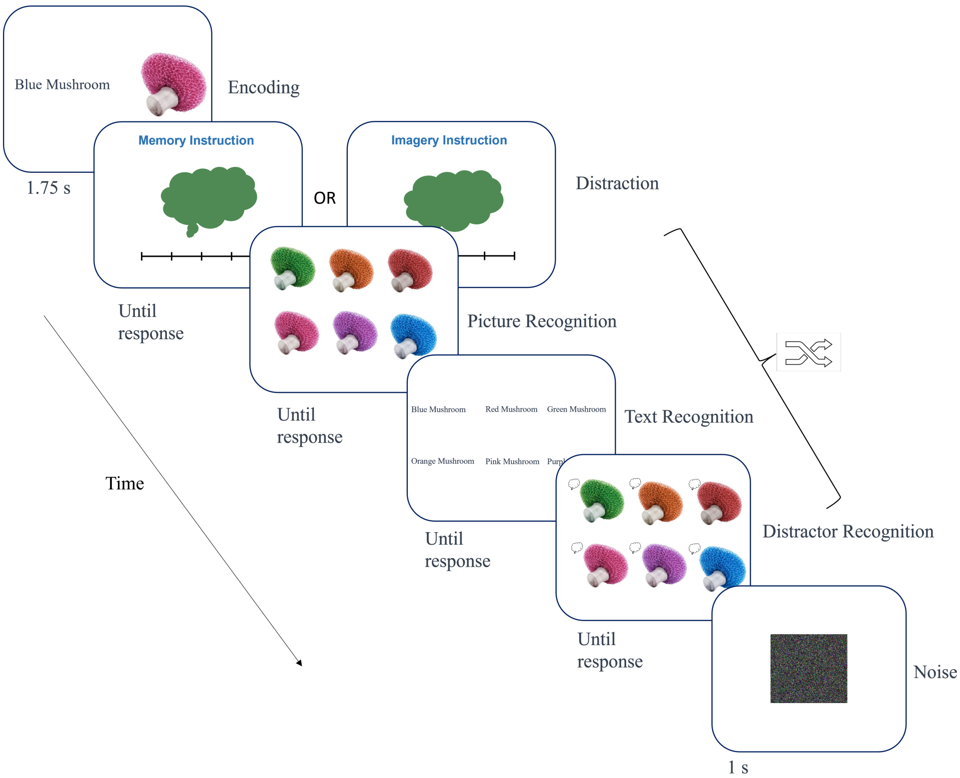
During the encoding phase, participants were concurrently presented with visual (picture) and verbal (text) items on the left and on the right parts of the screen, as in Experiments 1 and 2. The encoding phase was followed by a visual distractor, presented in a callout. During this phase, participants were given a colored callout. The participants in the imagery instruction condition were instructed "Please imagine the previously presented object being in the new color of the callout and rate how vividly you imagined the color of the object". The participants in the memory instruction condition were instructed: "Please maintain the new color of the callout in your mind and rate how well you maintained the color of the object". The distractor was always incongruent in meaning with the concurrent items, which were also incongruent (Figure 4.1).

Figure 4.1 Experiment 3- Memory Task condition: all-incongruent




The recognition phase was composed of three parts: picture recognition, text recognition, and distractor recognition, and all these different recognition types were presented in random order, on different presentation screens. At the end of each trial, a colored mask was presented as a refresh period. Overall, the task consisted of 37 trials and took approximately 20 minutes to complete (Figure 4.2).

Figure 4.2 Experiment 3- Memory Task trial structure: encoding, distraction, and recognition phases




Note. The original language of the task was Turkish.

#### 4.1.2.2 Strategy Use Questions

To explore the strategies participants, applied in the different stages of the experiment, the strategy questions were included. They aimed to tap different (e.g., visual vs. verbal) strategies, used in the encoding phase (e.g. "I mostly focused on pictures and did not pay attention to verbal descriptions."), distraction phase (e.g., "Colors shown in the callouts influenced my maintenance of the initially shown pictures."), and recognition phase (e.g. "It was easier to recall the initially shown color word in the verbal description than the color of the picture."). To improve participants' understanding of these questions, a picture of a stimulus from the relevant stage is presented with the question. Participants rated their agreement with each item on a scale from 1= "strongly disagree" to 4 = "strongly agree".

### 4.1.3 Procedure

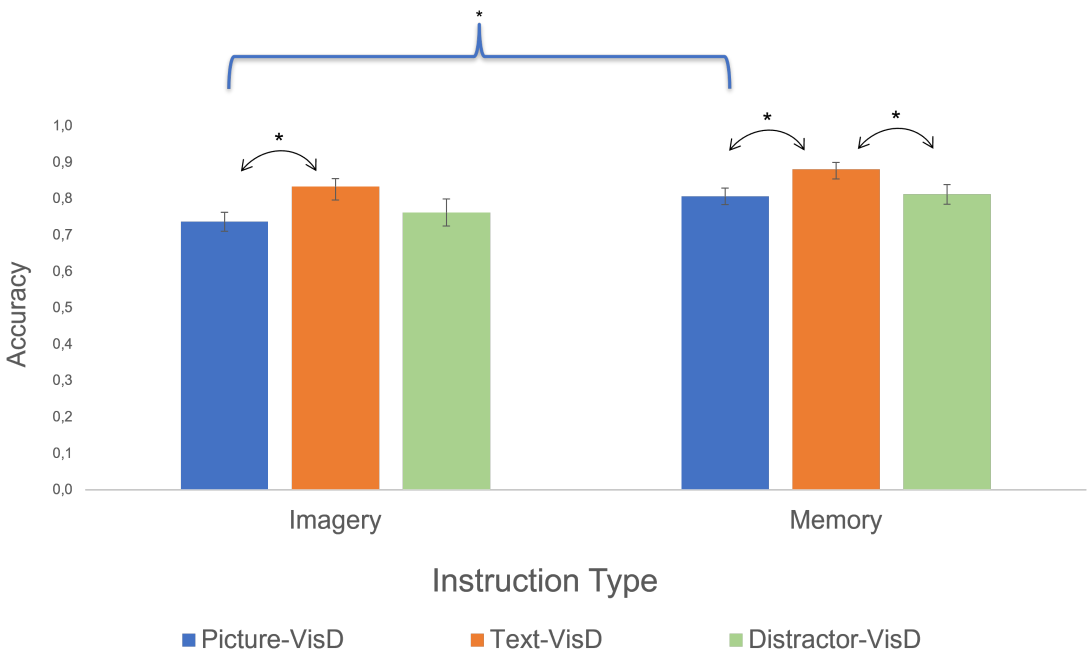
The study was run in 2 parts. An introductory zoom meeting and the main experiment (Memory Task) were conducted in an online setting. Participants were required to be in a silent place and use their PCs instead of phones. Participants were invited via zoom link to the first part of the study. The researcher individually instructed them about the task. Next, they received an email with the link to the second part of the study. Afterward, credit was granted to those who participated in the study for compensation.

## 4.2 Results

### 4.2.1 Picture vs. Text Memory: Accuracy

Repeated measures ANOVA was performed to reveal the effects of recognition type (picture concurrent, text concurrent, distractor) and instruction type (imagery, memory) on the accuracy of recognition performance. The Greenhouse-Geisser correction was applied as the data violated the assumption of sphericity (Figure 4.3).

Figure 4.3 Experiment 3- Accuracy for the recognition of picture, text, and distractor items under visual distraction, on memory and imagery instruction conditions




Note. VisD - visual distractor Note: All-incongruent condition.

The results demonstrated a significant effect of recognition type, F(1.529, 110.102) = 12.143, $p < .001$, $\eta_p^2 = .144$, so that participants' text recognition accuracy was significantly greater than picture recognition accuracy (Mdiff= .086, SE = .013), $p < .001$, and then visual distractor recognition accuracy (Mdiff= .070, SE = .022), p = .002. No other significant differences were observed. The effect of instruction type was only marginally significant, p = .067, that is, accuracy in the memory instruction condition (M = .833, SD = .115) tended to be higher than in the imagery instruction condition (M = .777, SD = .141), t(72) = -1.861, p = .067.

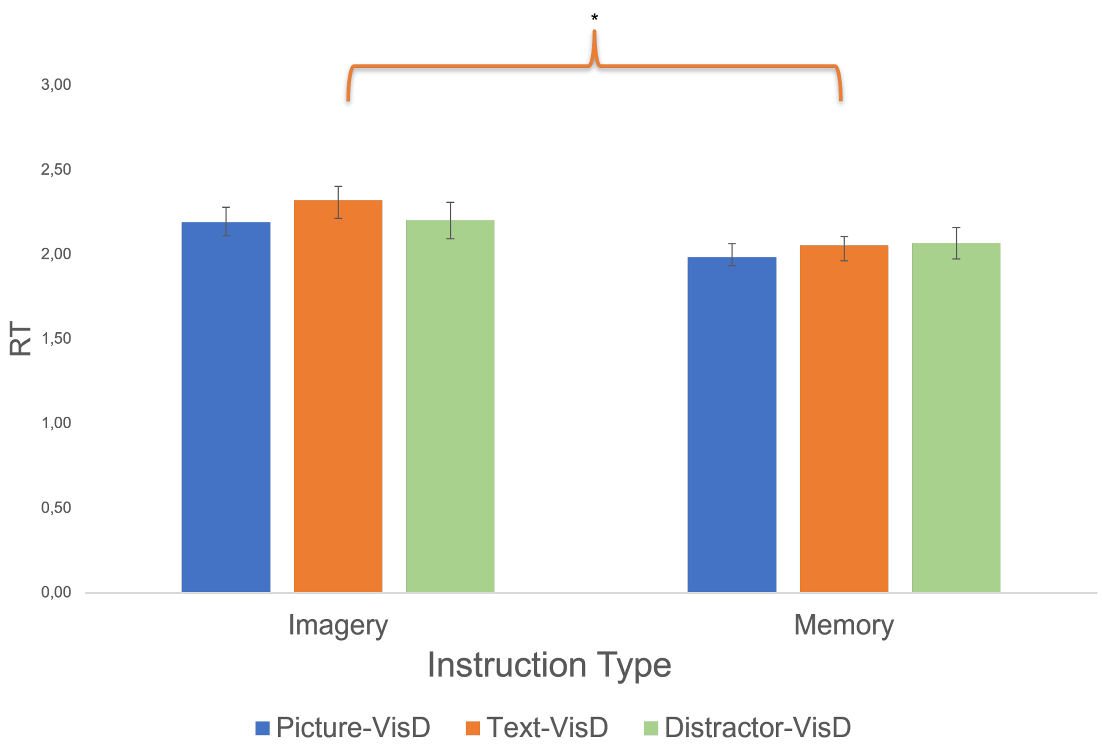
The interaction between recognition type and instruction type was not significant, p = .738, such that the difference between the accuracy of the recognized picture, text, and distractor items was not different between memory vs. imagery conditions. Additionally, an independent samples t-test demonstrated that picture recognition accuracy was significantly greater in the memory (M = .806, SD = .136) than in the imagery instruction condition (M = .736, SD = .156), t(72) = -2.060, p = .043.

### 4.2.2 Picture vs. Text Memory: RT

Repeated measures ANOVA was performed to reveal the effects of recognition type (picture concurrent, text concurrent, distractor) and instruction type (imagery, memory) on the speed of recognition performance (Figure 4.4). The results did not demonstrate a significant effect of recognition type, p = .204.

There was a trending effect of instruction type, F(1, 72) = 3.955, p = .051, $\eta_p^2$ = .052, that is RT in memory instruction condition tended to be faster. The interaction between recognition type and instruction type was not significant, p = .485. Independent samples t-test demonstrated that text recognition RT was significantly shorter in memory (M = 2.320, SD = .495). than in the imagery instruction condition (M = 2.053, SD = .310), t(72) = 2.784, p = .007 (Figure 4.4)

Figure 4.4 Experiment 3- Reaction time for the recognition of picture, text, and distractor items under visual distraction, on memory and imagery instruction conditions




### 4.2.3 Individual Differences in the efficiency of processing distractor

We compared the engagement with a distractor in Imagery vs. Memory Instruction conditions. Independent samples t-test demonstrated that distractor ratings were significantly greater in memory (M = 3.904, SD = .767) than in the imagery (M = 3.404, SD = .746). Distractor rating duration was significantly shorter in memory (M = 4.867, SD = 1.237) than in the imagery (M = 5.692, SD = 1.884) instruction condition, t(72) = 2.226, p = .029, suggesting that imagery instruction condition is more effortful than memory instruction condition.

Further, we performed correlation analysis to reveal the relationship between the performance on our memory test and mean ratings and duration (of processing the distractor item) (Table 4.1). We found that increase in visual distractor vividness rating increased the picture r = .258, p = .027, and visual distractor r = .269, p = .020, recognition accuracy (but not RT).

Distractor rating duration was found to be positively correlated with RT for the picture, r = .445, p < .001, text r = .525, p < .001, and visual distractor r = .528, p < .001. This finding suggests that greater engagement with the visual distractor increased the recognition pace. Additionally, we found a single common trait among all RT measures, that is all of the measures are intercorrelated.

Table 4.1 Experiment 3- Correlations between individual difference measures and memory performance.

| | | Picture Recognition accuracy | Picture RT | Text Recognition accuracy | Text RT | Visual Distractor Recognition accuracy | Visual Distractor RT |
|---|---|---|---|---|---|---|---|
| Distractor ratings | Imagery | .394* | -.298 | .094 | -.243 | .320 | -.381* |
| | Memory | -.018 | .068 | .187 | .014 | .150 | .228 |
| | Overall | .258* | -.18 | .190 | -.22 | .269* | -.12 |
| Distractor rating duration | Imagery | .163 | .448** | .155 | .468** | .076 | .493** |
| | Memory | -.092 | .376* | .032 | .528** | .009 | .581** |
| | Overall | .005 | .445** | .056 | .525** | .020 | .528** |

Note. * $p$ <.05, ** $p$ <.01.

### 4.2.4 Strategy Use

Independent samples t-test comparing self-reported strategies in imagery vs. memory instruction conditions, demonstrated just some marginal differences, and only one item "I mentally added the object in the new color to the previously maintained object" received significantly greater ratings in imagery (M = 3.405, SD = 1.066) than in memory (M = 2.595, SD = 1.343) instruction condition, t(72) = 2.876, p = .005 (Table 4.2).

Table 4.2 Experiment 3- Independent Samples t-test Results for Strategy Use Questions

| | Imagery Instruction | | Memory Instruction | | *t (72)* | *p* |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | | |
| I memorized the color of the picture by naming it verbally (e.g., silently repeating the word 'blue' to myself). | 4.351 | 0.789 | 4.568 | 0.867 | -1.121 | .266 |
| I memorized the color word in the verbal description by visualizing it (e.g., imagining the 'orange' color in my mind's eye). | 2.595 | 1.301 | 3.189 | 1.525 | -1.805 | .075 |
| I verbally named the new color (e.g., silently saying the word 'green' to myself). | 4.378 | 0.794 | 4.514 | 0.989 | -0.648 | .519 |
| I visualized the new color (e.g., imagining the 'green' color in my mind's eye). | 3.108 | 1.370 | 2.622 | 1.401 | 1.510 | .135 |
| I imagined (mentally visualized) changing the color of the initial object in my mind. | 3.378 | 1.139 | 2.811 | 1.371 | 1.937 | .057 |
| I mentally added the object in the new color to the previously maintained object. | 3.405 | 1.066 | 2.595 | 1.343 | 2.876 | .005 |
| It was easier to recall the initially shown color of the picture than the color word in the verbal description. | 2.946 | 1.471 | 2.622 | 1.401 | 0.971 | 0,335 |
| It was easier to recall the initially shown color word in the verbal description than the color of the picture. | 3.324 | 1.292 | 3.811 | 1.151 | -1.710 | 0,092 |
| It was easier to recall the initially shown object than the added item. | 3.324 | 1.375 | 3.216 | 1.250 | 0.354 | 0,725 |

## 4.3 Conclusions

In the third experiment, different than previous experiments, we observed word superiority effect, that is text memory performance was greater than picture memory performance. Thus, H1 was not supported. We attribute this finding of a superior text memory performance to the detrimental effect of visual distractor on picture memory performance. Note, we had only visual distraction in study 3. Our design did not allow us to test H2A, meaning congruency effect, but we did test modality congruency effect. In support for H2B, we found that visual distractor deprived memory accuracy for picture than text, suggesting visual modality congruency effect.

Finally, in support of H3B, we found a positive correlation between the visual distractor ratings and picture recognition accuracy, indicating that greater ability in vivid visual imagery is related to better visual but not verbal memory performance, even though the imagined and maintained items were representing different meaning. Possibly, vividness reflects a general individual trait indicating better visual processing capacity, rather than situational attention to a distractor that limits the recourses for processing other memory items. We also found positive correlation between the distractor rating duration and RT for the picture, text, and visual distractor, suggesting that processing speed could be more or less stable individual trait across situations. The results of Experiment 3 in relation to the hypothesis are summarized in Figure 4.5.

Additionally, our results revealed differences between the imagery and memory instruction conditions, indicating that the imagery instruction condition was the most difficult. In imagery instruction condition, participants consistently showed slower responses and lower accuracy, which was significant only for accuracy of picture recognition and speed of text recognition. Additionally, we found one significant and some trending differences between the conditions in the reported strategies. Participants in the imagery instruction condition reported significantly higher agreement with statements indicating visualization strategy, e.g. “I mentally added the object in the new color to the previously maintained object”. These results suggest that memory and imagery instructions encouraged somewhat different strategies. We assumed that manipulating the maintained visual items in imagery appeared to be more effortful, probably because participants had to do a mental manipulation rather than simply adding a new item to their memory.

However, alternative explanations are also possible. The difference between memory and imagery instruction condition at least partially might be due to greater load on the imagery instruction condition. Participants were additionally asked to visualize the maintained item in the new color and the task became more attention demanding. In this sense, the difference between the conditions might not necessarily guarantee people are “visualizing” and that’s why it is more effortful. Rather, it might be a sign of greater load on memory. At the same time, there is no guarantee that people are not visualizing in memory condition. As noted by numerous studies, visual memory and visual imagery are related constructs (Keogh and Pearson 2011; Tong 2013).

Notably, we observed the differences in the reported strategies which indicate that greater load doesn’t fully explain the difference between the conditions. In any case, the data for imagery and memory conditions appeared quite similar to each other, suggesting commonalities in our memory task performance when instruction was emphasizing using imagery or not.

In Experiment 3, we addressed the limitations of Experiments 1 and 2. We separately measured speed of and accuracy of recognition. We increased the task difficulty to eliminate ceiling effect by decreasing the stimulus presentation duration (in the encoding phase) and by increasing the featural similarity between the maintained item and the distractor. However, we must acknowledge the limitations of this study as well. That is, Experiment 3 included only one all-incongruent condition, and one a distractor which restricted the possibilities of comparisons between the effects of visual vs. verbal distractors and different types of congruency effects. This was further tested in Experiment 4.

Figure 4.5 Summary of the Results of Experiment 3

| H1. Visual vs. Verbal Modality Effects | | H2.Congruency Effects | | | H3. Individual Differences | |
|---|---|---|---|---|---|---|
| ***H1A. Predominant Modality*** | ***H1B. Distracting Modality*** | ***H2A. Congruency Effect for Meaning*** | ***H2B. Congruency effect for Modality*** | ***H2C. Combined Effects of Meaning and Modality Congruency*** | ***H3A. Assessments of individual Differences in Visual and Verbal Skills*** | ***H3B. Assessments of Distractor Processing Efficiency*** |
| Contrary to our expectations, the text recognition accuracy was greater than both picture recognition accuracy and visual distractor recognition accuracy, indicating a word superiority effect.<br><br>No superiority effect was observed in terms of RT. | *There was only visual distraction in this experiment.* | *Meaning congruency effects were not tested as all the colors were mismatching (incongruent).* | Picture recognition accuracy was lower than text recognition accuracy under visual distraction, both in memory and imagery conditions (visual modality congruency effect).<br><br>However, this effect was not observed for RT. | *The comparison was not possible because there was just one type of distraction (visual) and one congruency condition (all-incongruent).* | *The comparison was not possible because we did not included individual difference measures.* | A positive correlation between the visual distractor ratings and recognition accuracy of the picture, and visual distractor. So, the vividness of imagery is related to visual memory.<br><br>Positive correlation between the distractor rating duration and RT for the picture, text, and visual distractor. Likely, there is an individual speed trait that affects all. |

# 5. EXPERIMENT 4

Experiment 4 aimed to investigate differences in performance for concurrently maintained picture and text memory. The picture and text memory performance were tested under visual and verbal distractors. The task was the same as in imagery instruction condition of Experiment 3; it included three meaning congruency types (all-incongruent, distractor-incongruent, all-congruent) and two types of distractors (visual and verbal).

## 5.1 Method

### 5.1.1 Participants

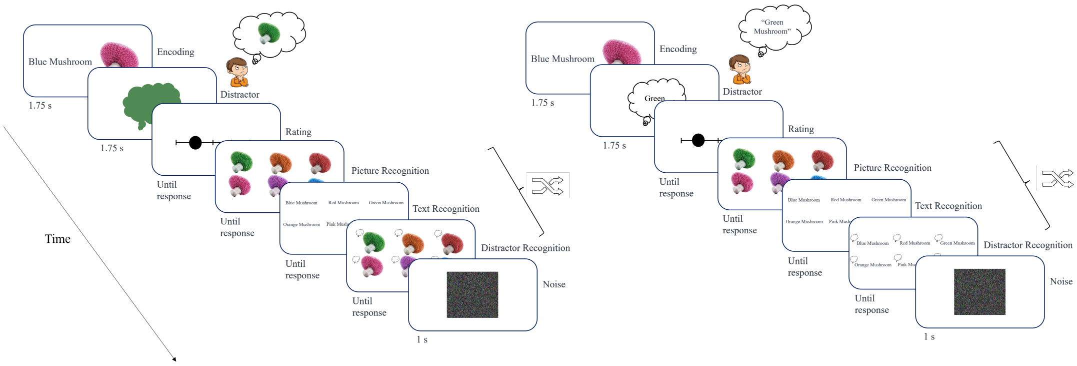
Participants were students (25 Female, 22 Male) recruited via Sabancı University SONA-system (Sona Systems, https://sona-systems.com/default.aspx) and received bonus course credits for their participation or via in-person invitation and received no compensation. All the participants received consent forms.

### 5.1.2 Materials

#### 5.1.2.1 Memory Task

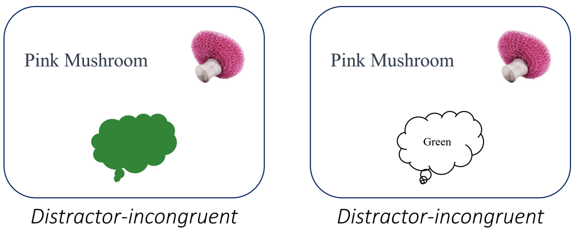
This task assessed the impact of visual and verbal distractors on the concurrent maintenance of visual and verbal working memory representations. The task design was the same as in Experiment 3 with a few modifications. First, in addition to visual distraction we included verbal distraction; both were presented to all the participants. During verbal distraction, participants were presented with a color name written inside a callout and were instructed: “Please mentally replace the previously seen color name with the color written in the callout". Visual and verbal distractors were presented in 2 separate blocks consisting of the same number of trials, and the order of these blocks was randomized (See Figure 5.1).

Figure 5.1 Experiment 4- Memory Task trial structure: encoding, distraction, and recognition phases




Second, we separated the rating task from the distraction phase, so that the distractor itself was presented for s limited time (1.75), while in Experiment 3 they appeared on the same page. Third, we added a distractor-incongruent condition in which the concurrently maintained items were the same in meaning but the distractor was incongruent with those items (Figure 5.2)

Figure 5.2 Experiment 4- Memory Task condition: Distractor-incongruent




The Memory Task consisted of overall 198 trials (99 with visual distraction, 99 with verbal distraction). In each of the distraction blocks, there were 3 conditions with 33 trials: 1) all the colors were matching (all-congruent), 2) all the colors were mismatching (all-incongruent), and 3) the distractor colors were mismatching with picture and text colors (distractor-incongruent). The task took approximately 1 hour to complete.

The randomization was applied using Psychopy/Pavlovia code for the following: 1) the spatial location (left vs. right) of the concurrently presented picture and the text in the encoding phase, 2) the color of each item in a specific trial, 3) the order of the presented objects, 4) the order of pictures, text, and distractor recognition. Half of the participants started with the visual distraction block and the other half started with the verbal distraction block.

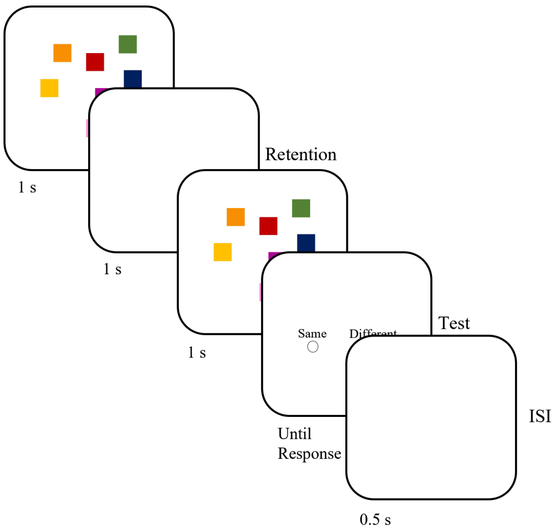
In addition, we included individual differences measures to examine their relationship with memory task performance. As Experiment 4 was the most comprehensive in terms of design, we included more tests in the battery of individual difference measures to examine their possible relationships with our memory task performance in different conditions. In particular, we included the validated visual vs. verbal processing style instrument, OSIVQ, same as in Experiments 1 and 2. Same as in Experiment 1, we included MRT assessing visual-spatial manipulation ability and WM. ATT, however, was not a validated measure of object imagery, but we considered it as an appropriate object-visual distractor in Experiment 1. We did not include this task in Experiment 4, rather we included more relevant, largely used visual and verbal WM capacity tasks, namely Change Detection Task (CDT) and Letter Span Task (LST). As our main memory task involved processing conflicting information and introduced additional distractor, we decided to include WM capacity measures in different modalities (visual-spatial, visual-object, and verbal) to examine how they were related to our memory task performance.

#### 5.1.2.2 Working Memory Capacity Tasks

Change Detection Task (CDT) assessed individual differences in visual (color) working memory capacity. We created the adapted version of the task developed by Luck and Vogel (1997); unlike the authors of the original task, we did not generate the color patterns with a code, and we included 24 trials instead of 32 as the available stimulus set only included 24 trials. Rather, the ready stimulus set was retrieved from Psychopy/Pavlovia (Pierce, et al., 2019; Project ID:206465).

In each trial, participants were presented with a pattern of colored squares, located at different parts of the screen (See Figure 19). The number of squares varied from trial to trial between 1 and 12. Then, half of the trials were presented with the same, and the other half of the trials with different patterns. Participants indicated whether the two patterns were the same or different in terms of the exact correspondence of the colors in the pattern (the number of squares and their spatial locations did not change within a trial). We averaged all correct responses to the trials to compute the scores.

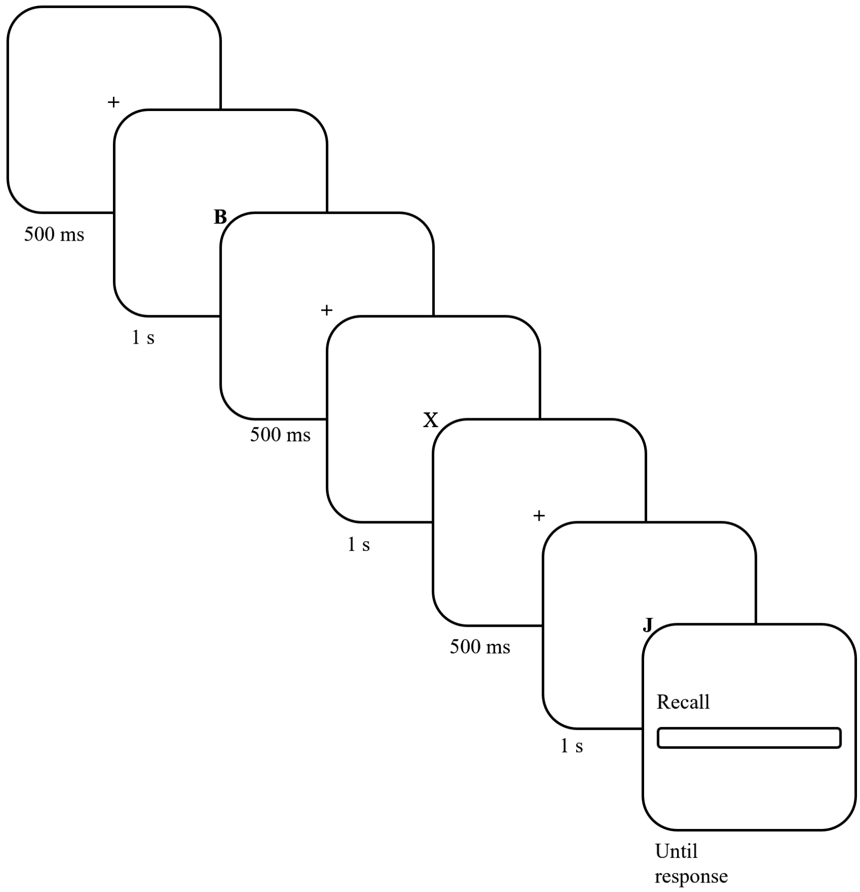
Figure 5.3 Experiment 4- Change Detection Task trial structure




Note. ISI: interstimulus interval

Letter Span Task (LST) was used to test the verbal working memory capacities of the participants. Originally, this task was developed by Kane et al. (2004). In this task, participants were randomly presented with the sequences of uppercase letters for 1 second with a 500 ms blank period between the trials. The length of each set of letters varied from three to eight and was randomly distributed across the task and overall there were 18 trials. Participants were instructed to repeat each letter aloud as the letters appeared on the screen. The letter set was composed of nine letters (B, F, H, J, L, M, Q, R, and X) randomly distributed across stimulus set conditions and each letter was repeated equally throughout the test. Different from the Kane et al. (2004) task, we presented a text box and asked participants to type the letter in the correct order in the recall phase, instead of providing a response sheet and asking them to fill the empty rows in the correct order. For scoring, we summed the fully correct responses. While computing the scores, we considered the answer as correct only when all letters were correctly recalled. As indicated earlier, within each span, ranging from 3 to 8, we included 3 trials. As suggested by Kane et al. (2004), we computed the average proportion-correct scores within each set and averaged the proportion-correct scores across all set sizes. For instance, in three trials of a 5-digit span, if the participant successfully recalled 1 of the trials then we considered that one as the correct answer and granted 5 points for the 5-digit span condition. If none of the 3 trials were successfully recalled, then the participant received 0 points. After calculating the proportion correct scores of each span, we averaged them to come up with the final Verbal Working Memory Capacity score.

Figure 5.4 Experiment 4- Letter Span Task trial structure




#### 5.1.2.3 Imagery Measures

As in Experiment 1, we assessed individual differences in abilities, preferences, and experiences in object imagery, spatial imagery, and verbal information processing with OSIVQ (Blazhenkova and Kozhevnikov 2009) and assessed spatial visualization ability with MRT (Peters et al., 1995 redrawn by Vanderberg Kuse, 1978). Different from Experiment 1, we only included the first part MRT.

### 5.1.3 Procedure

The study was run in 2 parts at 2 different time points. The main experiment (Memory Task) was prepared via Psychopy/Pavlovia (Pierce, et al., 2019) and conducted in the lab setting. The second part of the experiment (Individual Differences Battery) was created in Qualtrics (Provo, UT) and was conducted online in settings convenient for the participants. They were required, however, to use their PCs instead of phones. Individual Differences Battery included the tasks assessing visual working memory capacity (Change Detection Task; (Luck and Vogel 1997)), verbal working memory capacity (Letter Span Task; (Kane et al. 2004)), and individual differences in imagery and verbal processing (Object-Spatial Imagery and Verbal Questionnaire; (Blazhenkova and Kozhevnikov 2009)).

Participants were first invited to the laboratory to participate in the first part of the study. The researcher individually instructed each participant about the task. Following their completion of the first part, participants received an email with the link to the second part of the task and were asked to complete this part within a week.

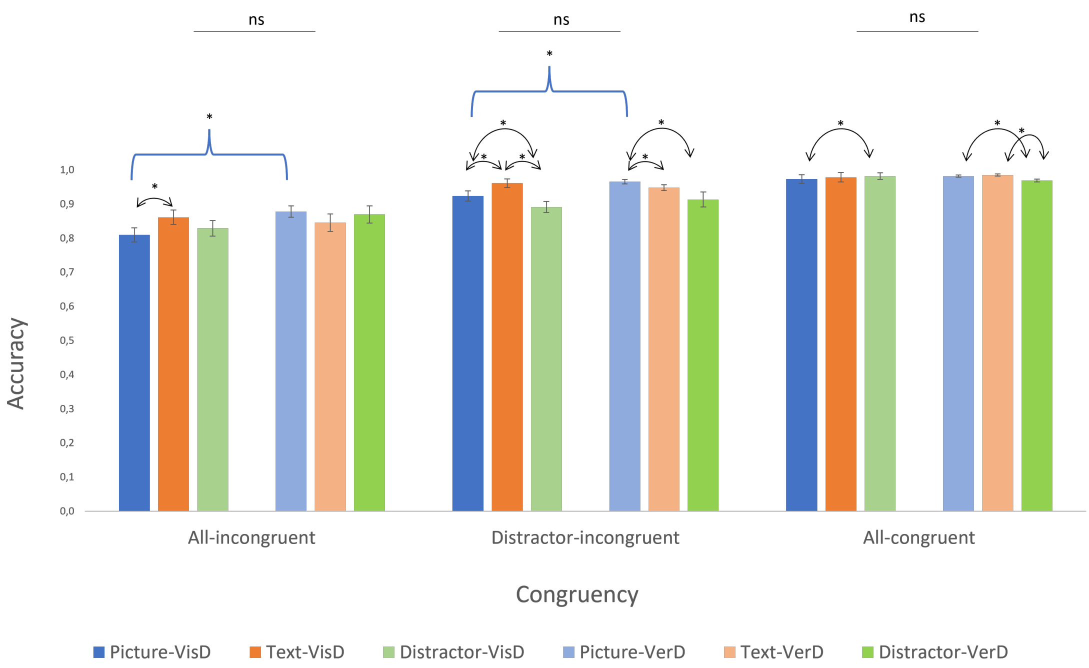
## 5.2 Results

### 5.2.1 Recognition Type x Distractor Modality x Congruency: Accuracy

Repeated measures ANOVA was performed to reveal the effects of recognition type (picture concurrent, text concurrent, distractor), distractor modality (visual, verbal), and meaning congruency (all-incongruent, all-congruent, distractor-incongruent) on the accuracy of recognition performance. The Greenhouse-Geisser correction was applied as the data violated the assumption of sphericity (Figure 5.5

The results demonstrated a significant effect of recognition type, F(1.660, 76.344) = 4.109, p = .027, $\eta_p^2$ = .082 so that participants' text recognition accuracy was significantly greater than distractor recognition accuracy (Mdiff= .021, SE = .007, p = .016). No other significant differences were observed. There was no effect of distractor modality; the accuracy when visual vs. verbal distractions were used did not significantly differ, F(1, 46) = 1.204, p = .278, $\eta_p^2$ = .026. The effect of congruency was significant, F(1.310, 60.253) = 76.257, p < .001, $\eta_p^2$ = .624, so that accuracy was greater on all-congruent trials than on all-incongruent trials (Mdiff = .129, SE = .014), and on distractor-incongruent trials (Mdiff = .044, SE = .007), p < .001. In addition, participants' accuracy was greater on all-incongruent trials than on distractor-incongruent trials (Mdiff = .085, SE = .010), all p's < .001 for pairwise comparisons.

Figure 5.5 Experiment 4- Accuracy for the recognition of picture and text items under visual-object, and verbal distraction, on different congruency conditions




Note. VisD - visual distractor, VerD - verbal distractor.

The interaction between recognition type and distractor modality was significant, F(1.599, 126.289) = 8.142, p = .01, $\eta_p^2 = .150$ such that the picture recognition accuracy was greater under verbal distraction than under visual distraction (Mdiff = .040, SE = .014, p = .007). No other significant differences were observed.

The interaction between recognition type and congruency was significant, F(2.745, 1.456) = 5.130, p = .03, $\eta_p^2 = .10$ such that there were no significant differences between the recognition types on all-incongruent trials. Text recognition accuracy was greater than distractor recognition accuracy on all-congruent trials (Mdiff = .006, SE = .111, p <.034). Distractor recognition accuracy was lower than picture recognition accuracy (Mdiff = .095, SE = .017, p < .001), and the text recognition accuracy (Mdiff = .105, SE = .015, p < .001) on distractor-incongruent trials. The interaction between distractor modality and congruency was not significant, F(1.456, 66.963) = 2.108, p = .142, $\eta_p^2 = .044$.

There was a significant three-way interaction between recognition type, distractor modality, and congruency, F(4, 184) = 4.44, p < .01, $\eta_p^2 = .09$, the interaction between the recognition type and the congruency differed across distractor modalities.

On the all-congruent condition, picture recognition accuracy was not different from text recognition accuracy neither under visual distraction (Mdiff = .006, SE = .003, p = .117) nor under verbal distraction (Mdiff = .003, SE = .003, p = .341). Picture recognition accuracy was lower than visual distractor recognition accuracy (Mdiff = .008, SE = .004), p = .036 but greater than verbal distractor recognition accuracy (Mdiff = .013, SE = .005, p = .007). Text recognition accuracy was not different from visual distractor recognition accuracy (Mdiff = .003, SE = .005, p = .506), but greater than verbal distractor recognition accuracy (Mdiff = .016, SE = .004, p < .001).

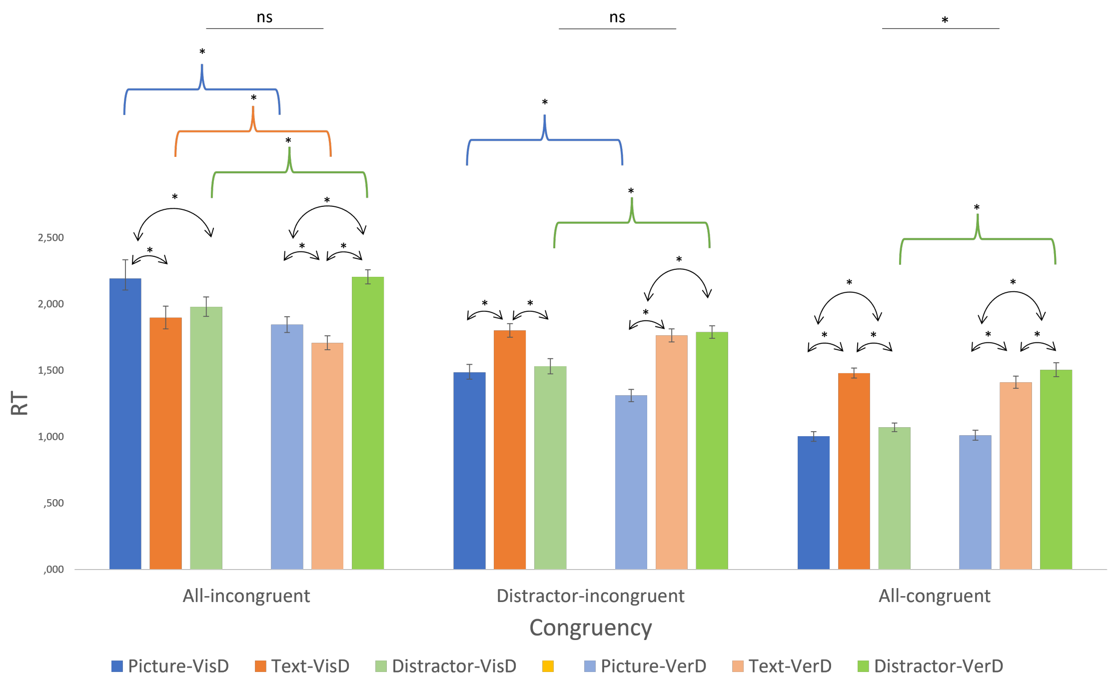
On the distractor-incongruent condition, text recognition accuracy was greater than picture recognition accuracy under visual distraction (Mdiff = .038, SE = .008), p < .001, but lower under verbal distraction (Mdiff = .017, SE = .007, p = .016). The picture recognition accuracy was greater than visual (Mdiff = .032, SE = .009, p = .001) and verbal (Mdiff = .052, SE = .021), p = .017 distractor recognition accuracy. Text recognition accuracy was greater than visual distractor (Mdiff = .070, SE = .010), p < .001 but not different from verbal distractor recognition accuracy (Mdiff = .035, SE = .020, p = .085).

On the all-incongruent condition, text recognition accuracy was greater than picture recognition accuracy under the visual distraction (Mdiff = .052, SE = .012), p < .001. No other significant difference was observed.

### 5.2.2 Recognition Type x Distractor Modality x Congruence: RT

Repeated measures ANOVA was performed to reveal the effects of recognition type (picture concurrent, text concurrent, distractor), distractor modality (visual, verbal), and congruency (all-incongruent, all-congruent, distractor-incongruent) on the recognition RT (Figure 5.6) There was a significant effect of recognition type, F(2, 45) = 27.438, p < .001, $\eta_p^2 = .549$, so that the picture recognition was faster than text recognition (Mdiff= .202, SE = .031), p < .001 and distractor recognition (Mdiff= .206, SE = .029), p < .001. There was no effect of distractor modality; the difference between the RT when visual vs. verbal distractions were used F(1, 46) = 0.072, p = .790, $\eta_p^2 = .002$. The effect of congruency was significant, F(1.373, 63.179) = 143.870, p < .001, $\eta_p^2 = .758$, so that participants were the fastest on all-congruent trials (M = 1.248, SE = .032), the slowest on all-incongruent trials (M = 1.973, SE = .058), and intermediate on distractor-incongruent trials (M = 1.615, SE = .041), all p's < .001 for pairwise comparisons.

Figure 5.6 Experiment 4- RT for the recognition of picture, text, and distractor items under visual vs. verbal distraction and on different congruency conditions.




Note. VisD - visual distractor, VerD - verbal distractor.

The interaction between recognition type and distractor modality was significant, F(2, 45) = 47.112, p < .001, $\eta_p^2$ = .677, such that the picture recognition was faster under verbal distraction than under visual distraction (Mdiff = 1.71, SE = .055), p = .003; text recognition was marginally faster under verbal distraction than under visual distraction (Mdiff = .099, SE = .055) p = .056; visual distractor recognition was faster than verbal distractor recognition (Mdiff = .306, SE = .050), p < .001.

The interaction between recognition type and congruency was significant, F(2.878, 132.398) = 48.816, p < .001, $\eta_p^2$ = .515, suggesting that the picture RT fastest on all-congruent trials (M = 1.009, SE = .030), the slowest on all-incongruent trials (M = 2.020, SE = .087), and intermediate on distractor-incongruent trials (M = 1.399, SE = .049), all p's < .001 for pairwise comparisons; text RT faster on all-congruent trials than on all-incongruent trials (Mdiff = .358, SE = .049) and on distractor-incongruent trials (Mdiff = .337, SE = .032), all p's < .001 for pairwise comparisons; distractor RT fastest on all-congruent trials (M = 1.289, SE = .038), the slowest on all-incongruent trials (M = 2.094, SE = .051), and intermediate on distractor-incongruent trials (M = 1.661, SE = .044), all p's < .001 for pairwise comparisons.

The interaction between distractor modality and congruency was significant, $F(1.350, 62.095) = 8.238$, $p = .003$, $\eta_p^2 = .152$, such that RT was fastest on all-congruent trials (M = 1.186, SE = .029) and slowest on all-incongruent trials (M =2.025, SE = .089) and intermediate on distractor-incongruent trials (M = 1.607, SE = .051) under the visual distraction, all p's < .001 for pairwise comparisons; RT was fastest on all-congruent trials (M = 1.310, SE = .042) and slowest on all-incongruent trials (M = 1.921, SE = .044) and intermediate on distractor-incongruent trials (M = 1.623, SE = .041) under the verbal distraction all p's < .001 for pairwise comparisons.

There was a marginally significant three-way interaction between recognition type, distractor modality, and congruency, $F(2.675, 123, 027) = 2.704$, $p = .055$, $\eta_p^2 = .056$. That is, the interaction between the recognition modality and the congruency marginally differed across distractor modalities.

On the all-incongruent condition, picture RT was slower than the text RT both under visual distraction (Mdiff = .294, SE = .092), p = .003 and under verbal distraction (Mdiff = .137, SE = .056), p = .019. Picture RT was slower than the visual distractor RT (Mdiff = .213, SE = .106), p = .050, but faster than verbal distractor RT (Mdiff = .360, SE = .066), p < .001. Text RT was faster than verbal distractor RT (Mdiff = .497, SE = .047), p < .001, but not different from visual distractor RT, p = .326.

On the all-congruent condition, picture RT was faster than text RT under visual distraction (Mdiff = .400, SE = .040), p < .001, and under verbal distraction (Mdiff = .497, SE = .029), p < .001. Picture RT was faster than the visual distractor RT (Mdiff = .067, SE = .033), p = .048 and verbal distractor RT (Mdiff = .494, SE = .030), p < .001. Text RT was slower than visual distractor RT (Mdiff = .409, SE = .029), p < .001 but faster than the verbal distractor RT (Mdiff = .094, SE = .030), p = .003.

On the distractor-incongruent condition, picture RT was faster than text RT under visual distraction (Mdiff = .316, SE = .045), p < .001, and under verbal distraction (Mdiff = .454, SE = .041), p < .001. Picture RT was not different than the visual distractor RT (Mdiff = .046, SE = .036), p = .209 but significantly faster than verbal distractor RT (Mdiff = .478, SE = .043), p < .001. Text RT was slower than visual distractor RT (Mdiff = .270, SE = .049), p < .001 but not different than verbal distractor RT (Mdiff = .024, SE = .040), p = .551.

### 5.2.3 Individual Differences

We performed Pearsons' correlation analysis to examine the relationships between the performance on our memory test and individual differences assessments, including measures of processing efficiency of the distractor item (Table 5.2) as well as OSIVQ, MRT, Visual Working Memory Capacity, and Verbal Working Memory Capacity scores (Table 5.1). We found that participants with higher visual-object ability (those with higher OSIVQ-object scores) showed consistently faster recognition of picture (r= -.482), text (r= -.305) and distractor (r= -.299), on all-congruent trials when the distractor was visual. That is, there was no conflict in the meaning of the color, and during the distractor phase, participants viewed the same color again. Our data suggest that object visualizers benefited from this most, as reflected in their improved processing speed for all, not only visual items. However, no such effect was observed for the verbal distractor, or for those who were verbalizers (those with higher OSIVQ-verbal scores) or spatial visualizers (those with higher OSIVQ-spatial scores).

There were only few other significant but inconsistent correlations, which we can't clearly explain. In particular, higher OSIVQ-object scores negatively correlated with text RT for all-incongruent trials under visual distractor (r= -.342). This may indicate some text processing advantage for object visualizers when they deal with conflicting information under visual distraction. Higher OSIVQ-spatial scores negatively correlated with text recognition accuracy for distractor-incongruent trials under visual distractor (r= -.304) but positively with text recognition RT for distractor-incongruent trials under verbal distractor (r=.349). This may indicate some difficulty with text processing in spatial visualizers when there is a conflicting distractor. Higher OSIVQ-verbal scores positively correlated with distractor recognition accuracy for distractor-incongruent trials under visual distractor (r=.448), which may indicate their higher susceptibility for visual distraction. However, these correlations with OSIVQ spatial and verbal scales did not show a consistent pattern and should be interpreted with caution.

Further, we observed that greater visual-spatial task (MRT) performance was negatively correlated with picture recognition RT (r= -.358) under visual distraction in all-congruent condition. That is, individuals with higher visual-spatial ability (and better visual-spatial WM) benefited when the visual distractor was representing the same color as both memory items. It might suggest that similar to those with higher visual-object ability, people with high-spatial ability might have some processing advantage for visual information.

Visual WM capacity was found to be correlated both with picture (r= .303) and text (r= .308) accuracy scores, when all items were incongruent and under verbal distraction. That is, the performance under incongruent verbal distractor increased when that person had greater visual WM capacity. However, this increase was not specific to picture memory, but applied to text memory as well. Verbal WM capacity was found to be correlated both with visual distractor recognition accuracy (r= .395), when all items were incongruent.

Table 5.1 Correlations between individual differences in visual and verbal skills and memory performance.

| | | | Picture recognition Accuracy | Picture RT | Text recognition Accuracy | Text RT | Distractor recognition Accuracy | Distractor RT |
|---|---|---|---|---|---|---|---|---|
| OSIVQ_obj | Visual Distractor | All-incongruent | -.063 | -.275 | -.004 | -.342* | -.023 | -.101 |
| | | Distractor-incongruent | -.138 | -.232 | -.235 | -.090 | .012 | -.253 |
| | | All-congruent | .136 | -.482** | .218 | -.305* | -.081 | -.299* |
| | Verbal Distractor | All-incongruent | .188 | -.229 | -.048 | -.206 | -.046 | -.195 |
| | | Distractor-incongruent | .055 | -.267 | -.151 | -.078 | -.018 | -.224 |
| | | All-congruent | -.044 | -.117 | .120 | -.085 | .110 | -.236 |
| OSIVQ_spa | Visual Distractor | All-incongruent | -.008 | .162 | .095 | .097 | -.061 | .233 |
| | | Distractor-incongruent | -.204 | .147 | -.304* | .288 | -.110 | .115 |
| | | All-congruent | -.109 | .147 | -.264 | .120 | -.285 | .148 |
| | Verbal Distractor | All-incongruent | .289 | .224 | .038 | .115 | -.057 | .131 |
| | | Distractor-incongruent | .005 | .148 | -.214 | .349* | -.081 | .130 |
| | | All-congruent | .201 | .144 | .119 | .196 | -.114 | .043 |
| OSIVQ_verbal | Visual Distractor | All-incongruent | .103 | -.154 | .136 | -.248 | .105 | -.147 |
| | | Distractor-incongruent | .171 | -.175 | .178 | -.281 | .448** | -.163 |
| | | All-congruent | .131 | -.112 | .127 | -.021 | -.012 | .003 |
| | Verbal Distractor | All-incongruent | .109 | -.239 | .034 | -.148 | -.027 | -.272 |
| | | Distractor-incongruent | .220 | -.117 | .145 | -.131 | -.030 | -.107 |
| | | All-congruent | -.038 | .102 | -.036 | .017 | .046 | .074 |
| MRT | Visual Distractor | All-incongruent | .269 | -.047 | .246 | -.012 | .291 | -.024 |
| | | Distractor-incongruent | .204 | -.179 | .164 | .065 | .185 | -.184 |
| | | All-congruent | .241 | -.358* | .265 | -.013 | .280 | -.172 |
| | Verbal Distractor | All-incongruent | .081 | .036 | .080 | .217 | .092 | .194 |
| | | Distractor-incongruent | .014 | -.098 | .089 | .076 | .151 | .094 |
| | | All-congruent | .021 | -.073 | .044 | -.019 | -.082 | -.071 |
| Visual WMC | Visual Distractor | All-incongruent | .033 | -.230 | .044 | -.229 | .119 | -.074 |
| | | Distractor-incongruent | .028 | -.029 | -.096 | -.012 | -.009 | -.069 |
| | | All-congruent | -.037 | -.069 | -.089 | -.090 | -.090 | -.081 |
| | Verbal Distractor | All-incongruent | .303* | .062 | .308* | .039 | .195 | .047 |
| | | Distractor-incongruent | .182 | -.057 | .260 | .108 | -.031 | .050 |
| | | All-congruent | -.125 | .020 | -.126 | -.062 | -.005 | -.037 |
| Verbal WMC | Visual Distractor | All-incongruent | .266 | -.047 | .119 | -.110 | .395** | .077 |
| | | Distractor-incongruent | .044 | .056 | -.056 | .058 | .078 | -.069 |
| | | All-congruent | .012 | .060 | .015 | -.136 | .004 | .037 |
| | Verbal Distractor | All-incongruent | .154 | .092 | .197 | -.071 | .193 | .013 |
| | | Distractor-incongruent | -.046 | -.006 | .173 | -.003 | .195 | -.129 |
| | | All-congruent | -.041 | -.072 | .013 | -.240 | .194 | .072 |

Note. * $p < .05$, ** $p < .01$.

## 5.3 Conclusions

Experiment 4 provided evidence for the picture superiority effect and a greater detrimental effect of visual distraction (H1A and H1B), which was observed only for RT data. Further, it demonstrated the congruency effect for meaning, both in RT and accuracy measures (H2A). As for the congruency effect for modality (H2B), the data appeared somewhat more complex than the prediction. Visual modality congruency effect was observed both for RT and accuracy measures, that is picture recognition performance was worse under visual than verbal distractor. Moreover, picture recognition performance was worse than text recognition accuracy under visual distraction. This data confirms our expectations that the same modality distractor is more detrimental to performance as compared to the different-modality distractor.

Contrary to the predictions, we did not find evidence that text recognition performance was worse than picture recognition accuracy under verbal distraction. Instead, text recall was better under verbal distraction than under visual distraction. Additionally, text RT was slower than picture RT under verbal distraction. The latter finding partially supports the verbal modality congruency effect. However, this seems to reflect an overall greater negative effect of visual over verbal distraction.

The combined effects of meaning and modality congruency (H2C) appeared to be even more complex, though it generally supported the expectations. When there was no conflict in meaning among the items (all-congruent trials) or when a distractor (visual or verbal) provided additional support for the encoded information, the speed of visual (both picture and visual distractor) processing improved compared to text memory speed. Possibly, visual memory can be aided by meaning-congruent information more readily than verbal one. On the distractor-incongruent trials, when the distractor created a new meaning but the concurrent items had the same meaning, the distractor was more impeding to the recognition accuracy of the same modality item.

A different pattern was observed for RT. Visual processing speed (both for picture and visual distractor) was faster compared to the text memory speed. On the all-incongruent trials when all meanings were incongruent, picture memory accuracy was reduced, and RT increased under the congruent (visual) modality distractor compared to the verbal distractor. Therefore, the meaning incongruency and modality congruency effects had combined effects on memory performance. However, this finding was observed only for picture memory under visual distractor, but not for text memory under verbal distractor.

Furthermore, the analysis of individual differences in visual and verbal processing did not show consistent and strong associations between memory performance and individual differences measures. Thus, H3A was not supported. The only relatively consistent finding is that Object Imagery (Object OSIVQ) under visual distractor was negatively associated with RT for both text, picture, and distractor recognition, mostly for all-congruent trials. So, when visual information was additionally enhanced, object visualizers benefited all memory items in terms of RT.

Interestingly, we observed correlations mostly under visual rather than verbal distractor and mostly for visual rather than verbal cognitive style or WM capacity measures which might suggest that visual processing seems to be more sensitive to individual variability than the verbal processing.

Finally, the analyses of distractor processing efficiency showed that in all-congruent trials, the congruent visual distractor aided memory for all modality items, but more so for visual ones. Visual distractor rating durations were positively related to RT for all recognitions, possibly reflecting individual higher processing speed. Further, these correlations were higher for all-congruent conditions, parallel to accuracy data. However, there was no parallel effect for verbal distraction. Verbal distractor vividness was not related to better verbal (either text or verbal distractor itself) memory accuracy. When the distractor was verbal and incongruent, surprisingly, pictures but not text were memorized better. This may indicate that the verbal distractor did not impede visual processing. The results of Experiment 4 in relation to the hypothesis are summarized in Figure 5.7.

Table 5.2 Correlations between distractor processing efficiency and memory performance

| | | | Picture recognition Accuracy | Picture RT | Text recognition Accuracy | Text RT | Distractor recognition Accuracy | Distractor RT |
|---|---|---|---|---|---|---|---|---|
| Distractor ratings | Visual Distractor | All-incongruent | .314* | .022 | .147 | .053 | .384** | -.109 |
| | | Distractor-incongruent | .244 | -.270 | .310* | .005 | .413** | -.229 |
| | | All-congruent | .717** | -.747** | .723** | -.067 | .717** | -.308* |
| | Verbal Distractor | All-incongruent | .162 | -.385** | .046 | -.255 | .199 | -.059 |
| | | Distractor-incongruent | .356* | -.444** | -.041 | -.159 | .259 | -.143 |
| | | All-congruent | -.082 | -.210 | -.146 | -.174 | -.110 | -.205 |
| Distractor rating duration | Visual Distractor | All-incongruent | .016 | .559** | -.006 | .430** | .100 | .446** |
| | | Distractor-incongruent | .071 | .463** | .017 | .329* | .016 | .415** |
| | | All-congruent | -.112 | .548** | -.143 | .580** | -.126 | .539** |
| | Verbal Distractor | All-incongruent | .177 | .487** | .121 | .386** | .108 | .086 |
| | | Distractor-incongruent | -.071 | .279 | .133 | .314* | .093 | .190 |
| | | All-congruent | -.297* | .714** | -.034 | .472** | .173 | .631** |

Note. * $p$ <.05, ** $p$ <.01.

Experiment 4 was the most comprehensive and balanced in terms of experimental design, as we manipulated visual and verbal distractors and included different meaning congruency conditions. However, we must acknowledge that, again, we observed the ceiling effect, as the task appeared relatively easy, which is a limitation.

Figure 5.7 Summary of the Results of Experiment 4

| **H1. Visual vs. Verbal Modality Effects** | | **H2.Congruency Effects** | | | **H3. Individual Differences** | |
|---|---|---|---|---|---|---|
| ***H1A. Predominant Modality*** | ***H1B. Distracting Modality*** | ***H2A. Congruency Effect for Meaning*** | ***H2B. Congruency effect for Modality*** | ***H2C. Combined Effects of Meaning and Modality Congruency*** | ***H3A. Assessments of individual Differences in Visual and Verbal Skills*** | ***H3B. Assessments of Distractor Processing Efficiency*** |
| Overall, the accuracy of text recognition was not different from picture recognition.<br><br>Picture recognition was faster than text recognition, supporting picture superiority.<br><br>Visual and verbal distractors' recognition accuracy did not differ.<br><br>Visual distractors were retrieved faster than verbal distractors. | Accuracy of recognition (overall, for picture and text combined) did not differ under verbal vs. visual distractors.<br><br>RT of recognition (overall, for picture and text combined) was faster under verbal than visual distraction. | The performance, both in terms of RT and accuracy, was the best in the all-congruent condition, the worst in the all-incongruent condition, and intermediate in the distractor-incongruent (partial congruency) condition. | Picture recognition accuracy was lower under visual distraction than under visual distraction (visual modality congruency effect).<br><br>Consistently, picture recognition was slower under visual distraction than under visual distraction (visual modality congruency effect).<br><br>However, text recognition was marginally faster under verbal distraction than under visual distraction (reversed verbal modality congruency effect). No such effect for accuracy. | On all-congruent trials, visual processing speed (both picture and visual distractor) was improved (processed faster) compared to text memory speed. No such effect for verbal memory under visual distraction. No such effect for RT.<br><br>On distractor-incongruent trials, the distractor was more impeding the recognition accuracy of the same modality item. This was observed for both picture and text memory, under visual and verbal distractors, correspondingly (combined congruency effects of meaning and modality).<br><br>Visual processing speed (both for picture and visual distractor) was always faster compared to the text memory speed.<br><br>On all-incongruent trials, picture memory accuracy was reduced, and RT increased under the congruent modality (visual) distractor compared to the verbal distractor. However, this finding was observed only for picture memory under visual distractor, but not for text memory under verbal distractor. | Overall, we did not find consistent and strong associations, clearly supporting our hypothesis. | For all-congruent trials, visual distractor vividness ratings correlated with all the measures of recognition accuracy, including picture, visual distractor, and text.<br><br>Visual distractor ratings were negatively correlated with RT for picture and visual distractor recognition.<br><br>Visual distractor rating durations were positively related to RT for all recognitions.<br><br>For partial and full meaning incongruency conditions, visual distractor vividness ratings correlated with recognition accuracy of visual distractor itself, and to a lesser degree with picture and text accuracy.<br><br>Visual distractor rating durations were positively related to all RTs.<br><br>Verbal distractor ratings positively correlated with the accuracy of picture recognition, but negatively with the speed of picture recognition. |

## 6. GENERAL DISCUSSION

Our research aimed to investigate recognition of concurrently encoded visual and verbal information presented under visual or verbal distraction. In four experiments, we manipulated the meaning (same vs. different) and modality (visual vs. verbal) of the concurrently encoded and distractor items and examined their recognition performance. Specifically, we focused on congruency effects for meaning and modality, and their interactions. Additionally, we investigated the relationship between memory performance and individual differences in visual and verbal processing.

Our research contributes to the body of literature that demonstrated evidence of both differences and interactions between visual and verbal processing. Consistent with these studies, we show that meaning of visual and verbal information may influence each other, and that they may be limited the use of shared cognitive resources, coordinated by the same executive component. The novelty of our study is that we did not only show that conflicting meaning representations might have detrimental effects on performance, but also showed how modality and meaning congruency interact with each other. Additionally, we showed that visual and verbal congruency effects are not exactly parallel to each other, which further contributes to the literature on visual vs. verbal processing differences.

**(H1) Visual vs. Verbal Modality Effects** (Visual processing is more powerful than verbal):

We hypothesized that visual processing is more powerful than verbal. Overall, our results provided partial support for H1 (both, A and B), indicating some evidence for the picture superiority effect, as well as a somewhat greater visual over verbal disruptive effect on memory. The underlying reason for expecting stronger visual representation is explained with greater physical distinctiveness of visual information (Mintzer Snodgrass, 1999). Indeed, research found that when the distinctiveness of verbal information was increased by adding different font, color to text, the superior visual memory performance disappeared (Ensor et al., 2019). Consistent with previous research, we expected that visual memory will be better than verbal. In addition, we tested whether visual superiority is also reflected in greater disruptive effect.

***H1A.Predominant Modality*** (Superior recognition performance of visual over verbal information)

Overall, H1A was partially supported. Consistent with literature that demonstrated the picture superiority effect (Bevan and Steger 1971; Shepard 1967; Thibodeau, Levy, and de Lemos 2021), we found some support for superior visual over verbal recognition for information learned during the concurrent presentation as well as distraction. In Experiments 1, 2, and 4, we found evidence for picture superiority effect either in accuracy or RT measures. In Experiment 1, we indeed observed that picture recognition accuracy was greater than text recognition accuracy. A consistent but only marginally significant trend was revealed in Experiment 2. In Experiment 4, even though we did not observe the picture superiority effect in terms of accuracy, possibly, due to the ceiling effect, we found it in the speed of the performance. That is picture items were recognized faster than text items.

Additionally, in Experiment 4, we found that visual distractors were recognized faster than verbal distractors, while their recognition accuracy did not differ. This finding further indicates the advantage of visual over verbal memory and a greater engagement in processing visual over verbal information. However, contrary to the expectations, Experiment 3 showed word superiority in the accuracy of recognition. Notably, this experiment used only a visual distractor, while in other experiments, distractors were either balanced or absent. The same condition (all-incongruent items, visual distractor) in Experiment 4 yielded similar to Experiment 3 results. This conflicting data could be possibly explained by using a visual distractor. That is, same-modality distraction impeded visual memory and therefore benefited text over picture memory. Our findings suggest that picture superiority could be affected by additional factors such as the modality (visual vs. verbal) of a distractor, as discussed in further sections.

***H1B.Distracting Modality*** (Visual distraction has greater detrimental effects on memory of maintained items than verbal distraction)

We found only limited support for H1B. Consistent with literature indicating that visual information is more salient and memorable (Van der Cruyssen et al. 2020), we further expected that visual modality could be more attention-demanding, and therefore more distracting than verbal one. Indeed, Experiment 4 demonstrated that visual distraction was overall more detrimental to performance, but this effect was observed only in the speed of recognition of concurrent items. However, we did not observe an overall superior visual over verbal disruptive effect on memory in Experiment 1. Instead, we observed the opposite effect: the accuracy of recognition under verbal distraction was lower than under visual-object or visual-spatial distraction, suggesting that verbal distraction was more detrimental than visual to the accuracy of memory of maintained items. The recognition speed did not depend on the type of distractor. The conflicting findings of Experiment 1, though, should be treated with caution, since all the distractors were different not only in modality but also in format, and verbal one could be just a more challenging task.

**(H2) Congruency Effects** (Memory depends on congruency of the processed information):

Consistent with expectations, we found congruency effects in meaning, that is meaning matching between the memory items improved their recognition, and in modality, that is memory decreased when the processed items shared the same, visual or verbal, modality. Further, we found partial support for the combined effects of meaning and modality congruency. Meaning incongruency reduced the recognition performance accuracy most when the modality of the distractor and the maintained items were matching. We found that modality congruency impeded memory more when meaning was also incongruent. However, this combined effect was observed mostly for visual modality and accuracy measures.

***H2A.Meaning Congruency*** (Memory is better for congruent than for incongruent information, i.e., meaning matching between the memory items improves their recognition).

Our data consistently supported H2A, which hypothesized that the meaning (color identity) mismatch between the concurrently encoded and maintained items would impede memory performance was overall supported.

Our results are consistent with a large body of previous research that demonstrated memory advantage when the meanings of the encoded items are congruent (Hershman, Beckmann, and Henik 2022; Kiyonaga and Egner 2014; MacLeod 1991; Stroop 1935). Across all the experiments when meaning congruency was manipulated (1, 2, and 4), we found that memory performance on congruent trials was better than on incongruent trials, both in terms of RT and accuracy. Additionally, we observed intermediate performance in the case of partial congruency, i.e., when only the distractor meaning was incongruent with the maintained items (Experiment 4). This finding indicates that memory performance may change gradually in relation to the level of congruency between the maintained information.

Higher performance on congruent trials than on incongruent trials can be explained by an attention-demanding filtering process that is needed to resolve the conflict between the meaning of the maintained visual and verbal information in WM (Kiyonaga and Egner 2014; Pan, Han, and Zuo 2019). When the newly processed information (distractor) is incongruent with the maintained information, or when there is a conflict in the meaning of the maintained information, this attention-demanding filtering process takes away the limited attentional resources from the maintained items, which leads to a decrease in their memory performance (Pan et al. 2022). Since the limited attentional resources (Lavie 2010) are divided between the different types of conflicting information, meaning incongruency might be detrimental to memory performance.

What happens to the distractor recognition performance when its meaning is incongruent with the concurrently maintained items? It's not only the memory for the concurrently processed items but also the memory for the distractor that is negatively affected by the meaning incongruency. Our analysis revealed that distractor recognition was better when the meaning of a distractor was congruent with the concurrently maintained items, both in terms of accuracy and RT. This reduction in the memory for a distractor on incongruent trials may indicate reduced engagement with the distractor during the maintenance of conflicting information.

Other research also showed that the effect of distractors during the encoding and maintenance was reduced when the memory was highly loaded (Bollinger et al. 2009; Konstantinou et al. 2014; Roper and Vecera 2014; Rose et al. 2005). Similarly, perception research reported reduced attention to a distractor when the attention was devoted to a cognitively demanding task involving conflict, e.g., flanker compatibility (Green and Bavelier 2003). Such that, protection of the maintained items from the distracting information may occur during the increased focal-task engagement (i.e., devoting more attentional resources to the maintenance), resulting in a reduced engagement in processing of irrelevant distracting information (Sörqvist and Marsh 2015; Sörqvist et al. 2016). Consistently, in our experiment, we found that, as the attentional demand of the task increased (incongruency level increased), the performance on the distractor recognition gradually reduced, i.e., increased focal-task (maintenance) engagement.

***H2B.Modality Congruency*** (Memory is decreased when the processed items share the same, visual or verbal, modality, i.e., the distractor matches with the modality of the maintained item).

Our results provided partial support for H2B, suggesting that memory performance decreases when the additional information is processed in the same modality. Indeed, we observed modality congruency effects in memory performance, both for picture and text memory. The effect was examined only in the experiments in which the distractor modality congruency was manipulated (Experiments 1 and 4).

In Experiment 1, the modality congruency effect was observed only for verbal modality. That is, text recognition was lower under verbal distraction than under visual-object or visual-spatial distraction. We should acknowledge though, that verbal distraction in Experiment 1 was more detrimental than visual distraction for both visual and verbal maintained information, though, it was more disruptive for text memory compared to picture memory. Therefore, it disrupted modality-similar (text) memory more than modality-different (picture) memory. We should note, however, that the distractors in Experiment 1 included very different tasks, i.e., verbal distractors required completing 3 different items of a questionnaire, which could be more cognitively demanding than performing 1-item of visual-object or visual-spatial task. Therefore, the conclusions from this experiment are rather limited.

Complementary, Experiment 4 provided evidence for the visual modality congruency effect. Picture recognition performance was worse under visual than verbal distraction both in terms of accuracy and RT. That is, the visual distractor was more detrimental to the picture rather than text memory performance. However, we did not observe a parallel effect for a verbal distractor: it did not significantly lower text accuracy recognition. RT analyses showed that recognition was always slower under visual distraction than under verbal one, but more so for picture memory. Thus, the negative impact of visual distraction was more evident for the same (visual) modality. We obtained additional support for the visual modality congruency effect in Experiment 3, where picture recognition accuracy was lower than text recognition accuracy under visual distraction, both in memory and imagery condition. However, this effect was not observed for RT.

These results are in line with the literature suggesting that congruency between the modality of the maintained and distractor items have a detrimental effect on memory performance (Bae and Luck 2019; Kim, Kim, and Chun 2005; Oberauer et al. 2018; Tikhonenko, Brady, and Utochkin 2021).

***H2C. Combined Effects of Meaning and Modality Congruency*** (Meaning incongruency reduces performance most when the modality of the distractor and maintained items are matching).

H2C was partially supported, mostly for visual modality, and for accuracy measures. That is, consistent with expectations that memory performance declines more when meaning incongruency and modality congruency appear in combination, we indeed found that modality congruency impeded memory more when meaning was also incongruent. Overall, while the meaning congruency effect was quite robust, the modality congruency effect depended on meaning congruency (congruent, incongruent), item modality (visual, verbal), and item type (maintained information or distractor).

In line with our expectations, Experiment 1 showed a combined negative effect of modality and meaning congruency on picture recognition accuracy, however, it was not observed for text recognition accuracy. Further, results of Experiment 4 showed that when there is no conflict in meaning (all-congruent trials), and any-modality distractor provides additional support for the encoded information since the distractor and maintained items were representing the same color, then visual (picture and visual distractor) processing RT were faster than verbal (text) memory. We also observed the same effect when the distractor was not matching in meaning with text and picture items (distractor-incongruent trials). Our results suggest that irrespective of whether there is a conflict with the distractor, pictures are just processed faster.

On the other hand, as expected, on distractor-incongruent trials, the combination of meaning incongruency and modality congruency impeded both picture and text memory performance. However, on all-incongruent trials, when all meanings are incongruent, this combination reduced picture accuracy and increased RT. This was not observed, however, for text performance. That is, picture memory seems to be more fragile than text memory due to the combined effects of meaning incongruency and modality congruency, but only in the most challenging condition.

Interestingly, processing time seems to be affected in a more complex way than accuracy. That is the meaning support either from concurrent or distractor items gave an advantage for pictures but not for text memory. However, in the case of full meaning conflict, RT behaved consistently with accuracy measures, indicating combined congruency effects. Our results are consistent with other literature suggesting that meaning and modality effects could be combined (Tikhonenko, Brady, and Utochkin 2021)

**(H3) Individual Differences** (Individual differences in visual and verbal processing are related to visual and verbal memory performance):

***H3A.Assessments of individual differences in visual and verbal processing*** (Higher scores on visual processing assessments are related to better visual memory performance, whereas higher scores on verbal processing assessments are related to better verbal memory performance).

Overall, H3A was not supported. We did not find significant correlations between memory task performance and individual differences in visual and verbal processing measures in Experiments 1 and 2. Even though Experiment 4 showed some significant correlations, they were rare, inconsistent, and not strong.

These results are inconsistent with our predictions and other literature that showed that working memory capacity related to the ability of attentional filtering of relevant information (Unsworth and Engle 2005), while differences in visual-object, visual-spatial and verbal styles were related to visual-object, visual-spatial, and verbal memory performance, correspondingly (Kraemer et al. 2017; McCunn and Cilli-Turner 2020). The inconsistency with the literature could be explained by the different experimental manipulations applied. For instance, Kraemer et al. 2017 findings were mainly relying on a spatial task. Participants were presented with a route through the city and asked about the landmarks as a visual task and judgment of relative direction as a verbal task as it is considered measuring the labeling the visual information. However, in our study we directly tested visual and verbal maintenance and distractibility of those modalities. Further investigation might extend our understanding of the relationship between the WM task performance and individual differences in visual vs. verbal WM capacity and processing styles.

It should be also noted that, we observed correlations mostly under visual rather than verbal distractor and mostly for visual rather than verbal cognitive style or WM capacity measures. Although the observed correlations were not consistent and strong, visual processing seems to be more sensitive to individual variabilities than verbal one.

***H3B. Distractor processing efficiency*** (Individual efficiency in processing the visual or verbal distractor is related to visual or verbal memory performance).

Our results provided partial support for H3B. As expected, we observed correlations between participants' performance on the memory task and the assessments of engagement with the distractor. Both in Experiment 3 and the related condition (all-incongruent condition, under visual distraction) of Experiment 4, we found an increase in the picture and visual distractor recognition accuracy (but not RT) when the visual distractor vividness rating increased. Experiment 4 showed that visual distractor vividness was, overall, associated with higher accuracy of distractor recognition. A similar trend was observed for RT, indicating that higher visual distractor vividness was related to faster recognition of the distractor, though it was significant only in the all-congruent condition. Besides, the association between vividness and accuracy increased with increasing congruency. These results suggest that higher processing efficiency of the distractor (and lower meaning conflict) improved distractors' memory. Greater distractor engagement mostly improved memory for the distractor itself, but also for other items, mostly when the distractor was visual.

Inconsistent with the expectation that a greater engagement with a distractor would lead to a reduced memory performance for the concurrent but not for the distractor items, in both Experiments 3 and 4, we found that the duration of the engagement with the visual distractor was related to a better recognition performance not exclusively for the distractor item. According to the time-based resource-sharing model (Barrouillet, Portrat, and Camos 2011, for a review), concurrent attentional processes limit the WM storage capacity as the same time was consumed by the concurrent items. The poorer performance on meaning incongruent trials could be due to the attention to the distractor consuming the same time that deprives the attention devoted to the internally maintained items (Kiyonaga and Egner 2013).

The model suggested that greater time consumed by the distractor, greater impairment in performance, and limited time for processing multiple items simultaneously, lead to greater impairment of WM performance (Camos and Barrouillet 2010; Vergauwe 2010). In our data, we observed that a greater time devoted to the distractor processing positively correlated with all RT measures for different memory items performance. Consistent with this model, this may indicate that greater engagement with the distractor, i.e., time consumed, leads to a drop in the memory performance, since simultaneous processing of several memory items excessively consumes shared attentional resources. The engagement with a distractor led to reduced performance, as was shown in the increased time for recognition. Interestingly, this was mostly observed for a visual distractor, which further indicates that it has more negative influence than the verbal distractor. However, since the greater engagement with a distractor on the all-congruent trials was also found to increase the recognition speed of all items, including the distractor itself, this further investigation is needed.

# 7. CONCLUDING REMARKS, LIMITATIONS AND FUTURE DIRECTIONS

Our work revealed noteworthy differences between visual and verbal memory during concurrent processing and distraction. It contributes to the literature on visual and verbal memory and their interactions. Our data suggest that visual processing is more powerful than verbal. In particular, we found evidence for superior visual over verbal memory, as well as for a greater visual than verbal disruptive effect on memory. Next, we found congruency effects in meaning, that is meaning matching between the memory items improved their recognition, and in modality, that is memory decreased when the processed items shared the same, visual or verbal, modality. Another contribution of our research is the description of how meaning and modality effect interact. We found partial support for the combined effects of meaning and modality congruency. Meaning incongruency reduced the recognition performance accuracy most when the modality of the distractor and the maintained items were matching. We found that modality congruency impeded memory more when meaning was also incongruent. However, this combined effect was observed mostly for visual modality and accuracy measures. We did not find consistent and strong associations between memory performance and existing assessments of individual differences. However, for our memory task, we found correlations between individual efficiency in processing visual or verbal distractors and visual or verbal memory performance.

We acknowledge several limitations of our work and would like to provide future directions for both improving the weaknesses of our research and contributing to the literature with novel findings. First, we observed ceiling effect across all studies, which indicates that the task was too easy. We believe that our design can be improved by increasing its difficulty, such as increasing the retention period and making it harder for participants to maintain more conflicting items in their memory. Additionally, the current task may involve not only manipulation of the color meaning, but also the object itself. A further development of this research could expand the current task to visual-spatial domain. That is, the locations (e.g., below, above, left etc.) could be manipulated.

Notably, since the studying relationships with individual differences measures was additional, exploratory research question, we tried including different measures to examine the potential relationships between our main memory task and these measures. The most comprehensive Experiment 4 used the most comprehensive battery of assessments. However, having different assessments in different studies is a limitation of our research. We want to acknowledge that using the same assessments would aid the comparison between the studies.

We included strategy use questions in Experiment 3 but not Experiment 4 since the main aim in Experiment 3 was to test the experimental procedure for Experiment 4 and examine whether imagery instruction would yield different results from memory instruction for the same task. Strategy questions mainly assessed the differences between imagery and memory institutions, and therefore were not relevant for Experiment 4, which only used imagery instruction. However, future studies may further explore strategies used when performing memory task that require processing conflicting visual and verbal information.

Furthermore, we did not calculate sample size prior to the experiments and recruited participants from the pools available at the time of the experiments. In Experiments 1 and 2, we used larger samples as they were run online, so later we were able to apply strict exclusion criteria to contain more reliable data. In Experiments 3 and 4, the participants were run individually in the lab settings, which was more time consuming. We carefully informed participants individually and made sure they all received the correct instructions and did not experience any problem during the experiment, in a tight schedule.

Additionally, we want to note that even though we observed some differences in recognition of distractor vs. concurrent items, we cannot attribute the found differences in the memory to the time when they were presented. That is, concurrent items we not only presented first, but they also presented together with another memory item, whereas distractor items not only appeared later but also were presented alone. Future studies may disentangle these factors by administering two concurrent distractors and manipulating their meaning or showing all the items sequentially one after the other. This would contribute to the literature on order effects in memory performance.

Another limitation of our study relates to its ecological validity. Interaction between visual and verbal processing was demonstrated in more complex context (Arndt, Schüler, and Scheiter 2019; Schüler, Arndt, and Scheiter 2015) corresponding to daily life. However, in our study we individually presented objects and words without a given context. Further research might consider including more realistic stimuli (e.g., textbook with visual illustrations or navigation map containing visual and verbal information) while testing the possible interactions between visual and verbal processing. This will improve the potential applications of our study for a range of fields such as education, design, human-computer interaction, or development of visual and verbal aids targeted to the needs and processing capacities of different individuals.

## BIBLIOGRAPHY


Antonietti, Alessandro, and Marisa Giorgetti. 1998. "The verbalizer-visualizer questionnaire: A review." *Perceptual and Motor skills* 86(1): 227–239.

Arndt, Jana, Anne Schüler, and Katharina Scheiter. 2019. "Investigating the influence of simultaneous–versus sequential–text-picture presentation on text-picture integration." *The Journal of Experimental Education* 87(1): 116–127.

Baars, Bernard J, and Nicole M Gage. 2010. *Cognition, brain, and consciousness: Introduction to cognitive neuroscience.* Academic Press.

Baddeley, AD, GJ Hitch, and RJ Allen. 2021. "A multicomponent model of working memory." *Working memory: State of the science* pp. 10–43.

Baddeley, Alan D, and Graham Hitch. 1974. "Working memory." In *Psychology of learning and motivation.* Vol. 8 Elsevier pp. 47–89.

Bae, Gi-Yeul, and Steven J Luck. 2019. "What happens to an individual visual working memory representation when it is interrupted?" *British Journal of Psychology* 110(2): 268–287.

Barlow, Horace B, Colin Blakemore, and John D Pettigrew. 1967. "The neural mechanism of binocular depth discrimination." *The Journal of physiology* 193(2): 327.

Barrouillet, Pierre, Sophie Portrat, and Valérie Camos. 2011. "On the law relating processing to storage in working memory." *Psychological review* 118(2): 175.

Bellhouse-King, Mathew W, and Lionel G Standing. 2007. "Recognition memory for concrete, regular abstract, and diverse abstract pictures." *Perceptual and motor skills* 104(3): 758–762.

Bevan, William, and Joseph A Steger. 1971. "Free recall and abstractness of stimuli." *Science* 172(3983): 597–599.

Blazhenkova, Olesya, and Maria Kozhevnikov. 2009. "The new object-spatial-verbal cognitive style model: Theory and measurement." *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 23(5): 638–663.

Blazhenkova, Olesya, Junko Kanero, Irem Duman, and Ozgenur Umitli. 2022. "Read and Imagine: Visual Imagery Experience Evoked by Native vs. Foreign Language.".

Bollinger, Jacob, Edrick Masangkay, Theodore P Zanto, and Adam Gazzaley. 2009. Age differences in N170 amplitude modulation by selective attention and working memory load. In *Poster presented at the annual meeting of the Society for Neuroscience, Chicago, IL.*

Bonner, Michael F, and Russell A Epstein. 2021. “Object representations in the human brain reflect the co-occurrence statistics of vision and language.” *Nature Communications* 12(1): 1–16.

Boswell, Donald L, and Jeffery A Pickett. 1991. “A study of the internal consistency and factor structure of the Verbalizer-Visualizer Questionnaire.” *Journal of Mental Imagery* .

Brady, Timothy F, Talia Konkle, Jonathan Gill, Aude Oliva, and George A Alvarez. 2013. “Visual long-term memory has the same limit on fidelity as visual working memory.” *Psychological science* 24(6): 981–990.

Cabeza, Roberto, and Lars Nyberg. 2000. “Imaging cognition II: An empirical review of 275 PET and fMRI studies.” *Journal of cognitive neuroscience* 12(1): 1–47.

Camos, Valérie, and Pierre Barrouillet. 2010. “Factors of working memory development: The time-based resource-sharing approach.” In *Cognitive development and working memory.* Psychology Press pp. 165–190.

Cocchini, Gianna, Robert H Logie, Sergio Della Sala, Sarah E MacPherson, and Alan D Baddeley. 2002. “Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems.” *Memory & Cognition* 30(7): 1086–1095.

Commodari, Elena, Maria Guarnera, Andrea Di Stefano, and Santo Di Nuovo. 2020. “Children learn to read: how visual analysis and mental imagery contribute to the reading performances at different stages of reading acquisition.” *Journal of psycholinguistic research* 49(1): 59–72.

Cowan, Nelson. 1988. “Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system.” *Psychological bulletin* 104(2): 163.

Cowan, Nelson, and Candice C Morey. 2007. “How can dual-task working memory retention limits be investigated?” *Psychological science* 18(8): 686–688.

Cowan, Nelson, J Scott Saults, and Christopher L Blume. 2014. “Central and peripheral components of working memory storage.” *Journal of Experimental Psychology: General* 143(5): 1806.

Dehaene-Lambertz, Ghislaine, Christophe Pallier, Willy Serniclaes, Liliane Sprenger-Charolles, Antoinette Jobert, and Stanislas Dehaene. 2005. “Neural correlates of switching from auditory to speech perception.” *Neuroimage* 24(1): 21–33.

D’Esposito, Mark, John A Detre, Geoffrey K Aguirre, Matthew Stallcup, David C Alsop, Lynette J Tippet, and Martha J Farah. 1997. “A functional MRI study of mental image generation.” *Neuropsychologia* 35(5): 725–730.

Doherty, Jason M, Clement Belletier, Stephen Rhodes, Agnieszka Jaroslawska, Pierre Barrouillet, Valerie Camos, Nelson Cowan, Moshe Naveh-Benjamin, and

Robert H Logie. 2019. "Dual-task costs in working memory: An adversarial collaboration." *Journal of experimental psychology: learning, memory, and cognition* 45(9): 1529.

Durgin, Frank H. 2000. "The reverse Stroop effect." *Psychonomic bulletin & review* 7(1): 121–125.

Farah, Martha J, Katherine M Hammond, David N Levine, and Ronald Calvanio. 1988. "Visual and spatial mental imagery: Dissociable systems of representation." *Cognitive psychology* 20(4): 439–462.

Fougnie, Daryl, and Rene Marois. 2006. "Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms." *Psychological science* 17(6): 526–534.

Friederici, Angela D. 2011. "The brain basis of language processing: from structure to function." *Physiological reviews* 91(4): 1357–1392.

Friederici, Angela D, and Sonja A Kotz. 2003. "The brain basis of syntactic processes: functional imaging and lesion studies." *Neuroimage* 20: S8–S17.

Friederici, Angela D, and Stefan Frisch. 2000. "Verb argument structure processing: The role of verb-specific and argument-specific information." *Journal of Memory and language* 43(3): 476–507.

Fulford, Jon, Fraser Milton, David Salas, Alicia Smith, Amber Simler, Crawford Winlove, and Adam Zeman. 2018. "The neural correlates of visual imagery vividness–An fMRI study and literature review." *Cortex* 105: 26–40.

Ganis, Giorgio, William L Thompson, and Stephen M Kosslyn. 2004. "Brain areas underlying visual mental imagery and visual perception: an fMRI study." *Cognitive Brain Research* 20(2): 226–241.

Gazzaniga, M. 2009. *The cognitive neurosciences.* MIT press.

Gazzaniga, MS. 2004. "The cognitive neurosciences. 3rd.".

Green, C Shawn, and Daphne Bavelier. 2003. "Action video game modifies visual selective attention." *Nature* 423(6939): 534–537.

Heilbron, Micha, David Richter, Matthias Ekman, Peter Hagoort, and Floris P De Lange. 2020. "Word contexts enhance the neural representation of individual letters in early visual cortex." *Nature communications* 11(1): 1–11.

Hershman, Ronen, Lisa Beckmann, and Avishai Henik. 2022. "Task and information conflicts in the numerical Stroop task." *Psychophysiology* p. e14057.

Hubel, David H, and Torsten N Wiesel. 1962. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *The Journal of physiology* 160(1): 106.

Humphries, Colin, Jeffrey R Binder, David A Medler, and Einat Liebenthal. 2007. "Time course of semantic processes during sentence comprehension: an fMRI study." *Neuroimage* 36(3): 924–932.

Jackendoff, Ray. 2000. Foundations of language. In *Workshop in Leipzig.*

Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene, and Emmanuel Dupoux. 2003. "Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study." *Journal of Neuroscience* 23(29): 9541–9546.

Jha, Amishi P, Sara A Fabian, and Geoffrey K Aguirre. 2004. "The role of prefrontal cortex in resolving distractor interference." *Cognitive, Affective, & Behavioral Neuroscience* 4(4): 517–527.

Johnson, Elizabeth N, Michael J Hawken, and Robert Shapley. 2001. "The spatial transformation of color in the primary visual cortex of the macaque monkey." *Nature neuroscience* 4(4): 409–416.

Johnson, Marcia K, Carol L Raye, Karen J Mitchell, Sharon R Touryan, Erich J Greene, and Susan Nolen-Hoeksema. 2006. "Dissociating medial frontal and posterior cingulate activity during self-reflection." *Social cognitive and affective neuroscience* 1(1): 56–64.

Kahneman, Daniel, and Diane Chajczyk. 1983. "Tests of the automaticity of reading: dilution of Stroop effects by color-irrelevant stimuli." *Journal of Experimental Psychology: Human perception and performance* 9(4): 497.

Kane, Michael J, David Z Hambrick, Stephen W Tuholski, Oliver Wilhelm, Tabitha W Payne, and Randall W Engle. 2004. "The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning." *Journal of experimental psychology: General* 133(2): 189.

Kemmerer, David. 2014. *Cognitive neuroscience of language.* Psychology Press.

Keogh, Rebecca, and Joel Pearson. 2011. "Mental imagery and visual working memory." *PloS one* 6(12): e29221.

Kilts, Clinton D, Robin E Gross, Timothy D Ely, and Karen PG Drexler. 2004. "The neural correlates of cue-induced craving in cocaine-dependent women." *American Journal of Psychiatry* 161(2): 233–241.

Kim, So-Yeon, Min-Shik Kim, and Marvin M Chun. 2005. "Concurrent working memory load can reduce distraction." *Proceedings of the National Academy of Sciences* 102(45): 16524–16529.

Kiyonaga, Anastasia, and Tobias Egner. 2013. "Working memory as internal attention: Toward an integrative account of internal and external selection processes." *Psychonomic bulletin & review* 20(2): 228–242.

Kiyonaga, Anastasia, and Tobias Egner. 2014. "The working memory Stroop effect: When internal representations clash with external stimuli." *Psychological science* 25(8): 1619–1629.

Koć-Januchta, Marta, Tim Höffler, Gun-Brit Thoma, Helmut Prechtl, and Detlev Leutner. 2017. "Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study." *Computers in Human Behavior* 68: 170–179.

Kollöffel, Bas. 2012. "Exploring the relation between visualizer–verbalizer cognitive styles and performance with visual or verbal learning material." *Computers & Education* 58(2): 697–706.

Konstantinou, Nikos, Eleanor Beal, Jean-Remi King, and Nilli Lavie. 2014. "Working memory load and distraction: dissociable effects of visual maintenance and cognitive control." *Attention, Perception, & Psychophysics* 76(7): 1985–1997.

Kosslyn, Stephen M, Giorgio Ganis, and William L Thompson. 2001. "Neural foundations of imagery." *Nature reviews neuroscience* 2(9): 635–642.

Kotz, Sonja A, Stefano F Cappa, D Yves von Cramon, and Angela D Friederici. 2002. "Modulation of the lexical–semantic network by auditory semantic priming: An event-related functional MRI study." *Neuroimage* 17(4): 1761–1772.

Kozhevnikov, Maria, Carol Evans, and Stephen M Kosslyn. 2014. "Cognitive style as environmentally sensitive individual differences in cognition: A modern synthesis and applications in education, business, and management." *Psychological science in the public interest* 15(1): 3–33.

Kraemer, David JM, Victor R Schinazi, Philip B Cawkwell, Anand Tekriwal, Russell A Epstein, and Sharon L Thompson-Schill. 2017. "Verbalizing, visualizing, and navigating: The effect of strategies on encoding a large-scale virtual environment." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(4): 611.

Kroll, Judith F, and Jill S Merves. 1986. "Lexical access for concrete and abstract words." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12(1): 92.

Lamm, Claus, Herbert Bauer, Oliver Vitouch, and Reinhard Gstättner. 1999. "Differences in the ability to process a visuo-spatial task are reflected in event-related slow cortical potentials of human subjects." *Neuroscience Letters* 269(3): 137–140.

Lavie, Nilli. 2010. "Attention, distraction, and cognitive control under load." *Current directions in psychological science* 19(3): 143–148.

Lewis-Peacock, Jarrod A, Andrew T Drysdale, and Bradley R Postle. 2015. "Neural evidence for the flexible control of mental representations." *Cerebral Cortex* 25(10): 3303–3313.

Logie, Robert H, Valérie Camos, and Nelson Cowan. 2020. "The State of the Science of Working Memory." *Working Memory: The state of the science* p. 1.

Luck, Steven J, and Edward K Vogel. 1997. "The capacity of visual working memory for features and conjunctions." *Nature* 390(6657): 279–281.

Luo, Chun R. 1999. "Semantic competition as the basis of Stroop interference: Evidence from color-word matching tasks." *Psychological Science* 10(1): 35–40.

MacLeod, Colin M. 1991. "Half a century of research on the Stroop effect: an integrative review." *Psychological bulletin* 109(2): 163.

Makovski, Tal, Won Mok Shim, and Yuhong V Jiang. 2006. "Interference from filled delays on visual change detection." *Journal of vision* 6(12): 11–11.

Mayer, Richard E, and Laura J Massa. 2003. "Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference." *Journal of educational psychology* 95(4): 833.

Mazoyer, Bernard, Nathalie Tzourio-Mazoyer, Angélique Mazard, Michel Denis, and Emmanuel Mellet. 2002. "Neural bases of image and language interactions." *International Journal of Psychology* 37(4): 204–208.

McAvinue, Laura P, and Ian H Robertson. 2007. "Measuring visual imagery ability: A review." *Imagination, Cognition and Personality* 26(3): 191–211.

McCunn, Lindsay J, and Emily Cilli-Turner. 2020. "Spatial training and calculus ability: Investigating impacts on student performance and cognitive style." *Journal of Educational Research and Practice* 10(1): 20.

Mestres-Missé, Anna, Estela Camara, Antoni Rodriguez-Fornells, Michael Rotte, and Thomas F Münte. 2008. "Functional neuroanatomy of meaning acquisition from context." *Journal of Cognitive Neuroscience* 20(12): 2153–2166.

Milton, Fraser, Jon Fulford, Carla Dance, James Gaddum, Brittany Heuerman-Williamson, Kealan Jones, Kathryn F Knight, Matthew MacKisack, Crawford Winlove, and Adam Zeman. 2021. "Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia versus hyperphantasia." *Cerebral Cortex Communications* 2(2): tgab035.

Milz, Patricia, Pascal L Faber, Dietrich Lehmann, Thomas Koenig, Kieko Kochi, and Roberto D Pascual-Marqui. 2016. "The functional significance of EEG microstates—Associations with modalities of thinking." *Neuroimage* 125: 643–656.

Mishkin, Mortimer, Leslie G Ungerleider, and Kathleen A Macko. 1983. "Object vision and spatial vision: two cortical pathways." *Trends in neurosciences* 6: 414–417.

Motes, Michael A, Rafael Malach, and Maria Kozhevnikov. 2008. "Object-processing neural efficiency differentiates object from spatial visualizers." *Neuroreport* 19(17): 1727–1731.

Nelson, Douglas L, L Cermak, and F Craik. 1979. "Remembering pictures and words: Appearance, significance and name." *Levels of processing in human memory* pp. 45–76.

Nishimura, Kazuo, Takaaki Aoki, Michiyo Inagawa, Yoshikazu Tobinaga, and Sunao Iwaki. 2015. "Brain activities of visual thinkers and verbal thinkers: A MEG study." *Neuroscience letters* 594: 155–160.

Nishimura, Kazuo, Takaaki Aoki, Michiyo Inagawa, Yoshikazu Tobinaga, and Sunao Iwaki. 2020. "Mental rotation ability and spontaneous brain activity: a magnetoencephalography study." *Neuroreport* 31(13): 999–1005.

Oberauer, Klaus, Stephan Lewandowsky, Edward Awh, Gordon DA Brown, Andrew Conway, Nelson Cowan, Christopher Donkin, Simon Farrell, Graham J Hitch, Mark J Hurlstone et al. 2018. “Benchmarks for models of short-term and working memory.” *Psychological bulletin* 144(9): 885.

Obleser, Jonas, Jonas Zimmermann, John Van Meter, and Josef P Rauschecker. 2007. “Multiple stages of auditory speech perception reflected in event-related FMRI.” *Cerebral Cortex* 17(10): 2251–2257.

Paivio, Allan. 1965. “Abstractness, imagery, and meaningfulness in paired-associate learning.” *Journal of Verbal Learning and Verbal Behavior* 4(1): 32–38.

Paivio, Allan. 1971. “Imagery and language.” In *Imagery.* Elsevier pp. 7–32.

Pan, Yi, Yu Han, and Wuheng Zuo. 2019. “The color-word Stroop effect driven by working memory maintenance.” *Attention, Perception, & Psychophysics* 81(8): 2722–2731.

Pan, Yi, Zheyu Zhang, Xinkui Hu, and Wuheng Zuo. 2022. “Revisiting congruency effects in the working memory Stroop task.” *Attention, Perception, & Psychophysics* pp. 1–16.

Poldrack, Russell A, Anthony D Wagner, Matthew W Prull, John E Desmond, Gary H Glover, and John DE Gabrieli. 1999. “Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex.” *Neuroimage* 10(1): 15–35.

Postle, Bradley R, and Massihullah Hamidi. 2007. “Nonvisual codes and nonvisual brain areas support visual working memory.” *Cerebral cortex* 17(9): 2151–2162.

Reichle, Erik D, Patricia A Carpenter, and Marcel Adam Just. 2000. “The neural bases of strategy and skill in sentence–picture verification.” *Cognitive psychology* 40(4): 261–295.

Rhodes, Stephen, Agnieszka J Jaroslawska, Jason M Doherty, Clément Belletier, Moshe Naveh-Benjamin, Nelson Cowan, Valerie Camos, Pierre Barrouillet, and Robert H Logie. 2019. “Storage and processing in working memory: Assessing dual-task performance and task prioritization across the adult lifespan.” *Journal of Experimental Psychology: General* 148(7): 1204.

Richardson, Alan. 1977. “Verbalizer-visualizer: a cognitive style dimension.” *Journal of mental imagery* .

Roper, Zachary JJ, and Shaun P Vecera. 2014. “Visual short-term memory load strengthens selective attention.” *Psychonomic bulletin & review* 21(2): 549–556.

Rose, Michael, Carmen Schmid, Almut Winzen, Tobias Sommer, and Christian Büchel. 2005. “The functional and temporal characteristics of top-down modulation in visual selection.” *Cerebral Cortex* 15(9): 1290–1298.

Sack, Alexander T, Joan A Camprodon, Alvaro Pascual-Leone, and Rainer Goebel. 2005. “The dynamics of interhemispheric compensatory processes in mental imagery.” *Science* 308(5722): 702–704.

Sadler-Smith, Eugene, and Richard Riding. 1999. "Cognitive style and instructional preferences." *Instructional science* 27(5): 355–371.

Schwanenflugel, Paula J. 2013. "Why are abstract concepts hard to understand?" In *The psychology of word meanings*. Psychology Press pp. 235–262.

Schüler, Anne, Jana Arndt, and Katharina Scheiter. 2015. "Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations?" *Learning and Instruction* 35: 62–72.

Shepard, Roger N. 1967. "Recognition memory for words, sentences, and pictures." *Journal of verbal Learning and verbal Behavior* 6(1): 156–163.

Shin, Gyeonghee, and Chobok Kim. 2015. "Neural correlates of cognitive style and flexible cognitive control." *Neuroimage* 113: 78–85.

Shinkareva, Svetlana V, Vicente L Malave, Robert A Mason, Tom M Mitchell, and Marcel Adam Just. 2011. "Commonality of neural representations of words and pictures." *Neuroimage* 54(3): 2418–2425.

Simon, Sharon S, Erich S Tusch, Phillip J Holcomb, and Kirk R Daffner. 2016. "Increasing working memory load reduces processing of cross-modal task-irrelevant stimuli even after controlling for task difficulty and executive capacity." *Frontiers in human neuroscience* 10: 380.

Sörqvist, Patrik, and John E Marsh. 2015. "How concentration shields against distraction." *Current directions in psychological science* 24(4): 267–272.

Sörqvist, Patrik, Örjan Dahlström, Thomas Karlsson, and Jerker Rönnberg. 2016. "Concentration: The neural underpinnings of how cognitive load shields against distraction." *Frontiers in human neuroscience* 10: 221.

Stenberg, Georg. 2006. "Conceptual and perceptual factors in the picture superiority effect." *European Journal of Cognitive Psychology* 18(6): 813–847.

Stroop, J Ridley. 1935. "Studies of interference in serial verbal reactions." *Journal of experimental psychology* 18(6): 643.

Thibodeau, Paul, Isaac Levy, and Mikaela de Lemos. 2021. Exploring mental representation with a memory matching game. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 43.

Thierry, G., Price C. J. 2006. "Dissociating verbal and nonverbal conceptual processing in the human brain." *Journal of Cognitive Neuroscience* 18(6): 1018–1028.

Thierry, Guillaume, and Cathy J Price. 2006. "Dissociating verbal and nonverbal conceptual processing in the human brain." *Journal of Cognitive Neuroscience* 18(6): 1018–1028.

Tikhonenko, Platon, Timothy Brady, and Igor Utochkin. 2021. "Independent storage of real-world object features is visual rather than verbal in nature.".

Tong, Frank. 2013. "Imagery and visual working memory: one and the same?" *Trends in cognitive sciences* 17(10): 489–490.

Unsworth, Nash, and Randall W Engle. 2005. “Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven.” *Intelligence* 33(1): 67–81.

Van der Cruyssen, Ine, Franziska Regnath, Gershon Ben-Shakhar, Yoni Pertzov et al. 2020. “The Picture Superiority Effect in Memory Detection: Modality Effects in the Concealed Information Test.”.

Vandenberg, Steven G, and Allan R Kuse. 1978. “Mental rotations, a group test of three-dimensional spatial visualization.” *Perceptual and motor skills* 47(2): 599–604.

Vergauwe, Evie. 2010. Interference between processing and storage in working memory: a domain-general mechanism of time-based resource sharing? PhD thesis University of Geneva.

Vigneau, Mathieu, Virginie Beaucousin, Pierre-Yves Hervé, Hugues Duffau, Fabrice Crivello, Olivier Houde, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. 2006. “Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing.” *Neuroimage* 30(4): 1414–1432.

Visser, Maya, Elizabeth Jefferies, and MA Lambon Ralph. 2010. “Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature.” *Journal of cognitive neuroscience* 22(6): 1083–1094.

Vitouch, Oliver, Herbert Bauer, Georg Gittler, Michael Leodolter, and Ulrich Leodolter. 1997. “Cortical activity of good and poor spatial test performers during spatial and verbal processing studied with slow potential topography.” *International journal of psychophysiology* 27(3): 183–199.

Vogel, Edward K, Andrew W McCollough, and Maro G Machizawa. 2005. “Neural measures reveal individual differences in controlling access to working memory.” *Nature* 438(7067): 500–503.

Wang, Jing, Julie A Conder, David N Blitzer, and Svetlana V Shinkareva. 2010. “Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies.” *Human brain mapping* 31(10): 1459–1468.

Wickens, Delos D. 1973. “Some characteristics of word encoding.” *Memory & Cognition* 1(4): 485–490.

Yoon, Jong H, Clayton E Curtis, and Mark D’Esposito. 2006. “Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex.” *Neuroimage* 29(4): 1117–1126.

Yui, Lin, Roslin Ng, and Hiran Perera-WA. 2017. Concrete vs abstract words–what do you recall better? A study on dual coding theory. Technical report PeerJ Preprints.

Zwaan, Rolf A, Gabriel A Radvansky, Amy E Hilliard, and Jacqueline M Curiel. 1998. “Constructing multidimensional situation models during reading.” *Scientific studies of reading* 2(3): 199–220.