Cite as:

Rosner, A., Schaffner, M., & von Helversen, B. (2021, November 8). When the Eyes Have It and When Not: How Multiple Sources of Activation Combine to Guide Eye Movements During Multiattribute Decision Making. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000833

# When the Eyes Have It and When Not: How Multiple Sources of Activation Combine to Guide Eye Movements During Multi-Attribute Decision Making

Agnes Rosner[1], Michael Schaffner[1], and Bettina von Helversen[1,2]

[1]Department of Psychology, University of Zurich

[2]Department of Psychology, University of Bremen

**Author Note**

Agnes Rosner https://orcid.org/0000-0001-9785-9660

Bettina von Helversen https://orcid.org/0000-0003-2004-6922

All data and materials are available at https://osf.io/zpc2s/.

All authors gratefully acknowledge funding from Grant 157432 from the Swiss National Science Foundation (SNSF). The first author additionally acknowledges the support of SNSF Grant 186032.

Study results were partially presented as talks at the TeaP 2018, ECEM 2019, and at the Virtual Process Tracing Seminar 2021.

The authors thank Ramona Obrist for collecting the data in Experiment 1 and Craig Hedge as well as Klaus Oberauer for helpful comments on a previous version of the manuscript.

Correspondence concerning this article should be addressed to Agnes Rosner, Department of Psychology, University of Zurich, Binzmuehlestrasse 14, Box 22, 8050 Zurich, Switzerland. Email: agnes.rosner@psychologie.uzh.ch

**Abstract**

Memory plays a major but underexplored role in judgment and decision making (JDM). Studying eye movements—especially how people look at empty spatial locations when retrieving from memory information previously associated with those locations—provides useful information about how memory influences JDM. This so-called looking-at-nothing behavior is thought to reflect memory-driven allocation of attention. However, eye movements are also guided toward salient visual stimuli, such as test items presented on a screen. It is unclear how these multiple sources of activation combine to guide looking-at-nothing in JDM. We investigated this question in two experiments in which participants solved multi-attribute categorization tasks using an exemplar-based decision strategy. In the first experiment, we tested how the occurrence and the strength of looking-at-nothing vary with the presentation format and the amount of training participants received. Looking-at-nothing occurred during categorizations when test-item information was presented auditorily and visually, but for the latter only after visual information was removed from the screen. It occurred both when training items were learned by heart and when they were presented 10 times on the screen. A second experiment revealed that an explicit instruction to imagine retrieval-relevant information during categorizations increased looking-at-nothing but did not change the decision-making process. The results shed light on the interaction between eye movements and attention to information in memory during JDM that can be explained in light of a shared priority map in memory. A detailed understanding of this interaction forms the basis for using eye movements to study memory processes in JDM.

*Keywords:* memory, attention, similarity, eye movements, decision making

**When the Eyes Have It and When Not: How Multiple Sources of Activation Combine to Guide Eye Movements During Multi-Attribute Decision Making**

Memory is prevalent in all stages of the judgment and decision-making (JDM) process. For instance, people can readily classify an unfamiliar painting as being the work of a known artist by retrieving similar past instances of the artist's work from memory. When deciding whether to take their car or bike to work, before taking the bike people might recall the weather conditions of similar days in the past and whether they were fortunate enough not to get wet. When judging the suitability of job applicants, people might consider similar applicants they have seen in the past and recall their performance. Usually, decision options or alternatives (e.g., job applicants) are characterized by one or more attributes (e.g., language skills) taking different values (e.g., French or Italian). These attributes frequently form the basis of the decision process as they can be used, for instance, to determine which past instances (e.g., job applicants with similar skills) are retrieved from memory.

Little is known about how memory processes influence JDM (Weber, Goldstein, & Barlas, 1995). One reason is that there have been no suitable methods for observing the retrieval process independently of the decision outcome. Recent research suggests that eye movements provide useful information about the role of memory during JDM (Orquin & Mueller Loose, 2013). In particular, studies based on the so-called looking-at-nothing (LAN) behavior have been able to observe retrieval processes almost unobtrusively (Ferreira, Apel, & Henderson, 2008; Richardson, Altmann, Spivey, & Hoover, 2009). LAN describes how people look back at empty spatial locations when retrieving information from memory that has been associated with those locations during encoding. Researchers have used this phenomenon to trace what information is retrieved from memory during a wide range of JDM tasks (Jahn & Braatz, 2014; Klichowicz, Strehlau, Baumann, Krems, & Rosner, 2020; Krefeld-Schwalb & Rosner, 2020; Pärnamets, Johansson, Gidlöf, & Wallin, 2016; Platzer, Bröder, & Heck, 2014; Renkewitz & Jahn, 2012; Rosner & von Helversen, 2019; Scholz, Krems, & Jahn, 2017; Scholz, von Helversen, & Rieskamp, 2015). In the pioneering study by Renkewitz and Jahn (2012), participants first learned attribute information about decision options that were arranged within spatial frames at distinct spatial locations on a computer screen. Later they were asked to choose between the options. During decision making, eye movements to the emptied spatial locations reflected information search in memory for the learned attribute information. This study demonstrates that eye movements based on the LAN behavior can reveal what information is activated in memory during JDM.

Besides memory activation, eye movements are guided by visual features present in the stimulus environment (Awh, Belopolsky, & Theeuwes, 2012; Chun, Golomb, & Turk-Browne, 2011; Henderson, 2017; Kowler, 2011; Theeuwes, 2010). That is, a decision maker's attention is generally drawn to visually salient attributes of presented test items (Orquin, Bagger, & Mueller Loose, 2013; Rehder & Hoffman, 2005). Most studies applying LAN in JDM have used designs that limit the influence of visual information from the very beginning by, for instance, presenting test stimuli auditorily. However, most of the traditional JDM paradigms present test items visually. Little is known about how memory retrieval interacts with visual information processing in determining LAN behavior in JDM tasks. In addition, memory activation can also be modulated, for instance, through instructions given to participants or training of information that needs to be kept in memory. Therefore, a better understanding of how retrieval processes on the one hand and visual information processing on the other influence eye movement behavior is crucial if one wants to use LAN as a process measure to study memory-based JDM.

The main goal of this research was to devise and test a theoretical framework for exploring how memory retrieval interacts with visual information processing to produce LAN in memory-based JDM. The framework builds on the idea of a "shared priority map" that predicts where people will look during memory retrieval (Hedge & Leonards, 2013; Hedge, Oberauer, & Leonards, 2015; Theeuwes, Belopolsky, & Olivers, 2009). The key idea is that during JDM, information that is kept in memory and visual information compete for activation in a shared spatially organized representation held in memory. The spatial position in the memory representation that receives the highest activation will determine the target of the next fixation in the visual world. A secondary goal was to understand to what extent LAN can be used as a measure of memory processes in JDM tasks.

In the following, we first review the literature on LAN and LAN's relation to memory retrieval. Then, we detail the categorization decision task we used as a testbed and the assumed memory processes, introduce the framework and our predictions for LAN during the decision process, and report two experiments to test our predictions.

**LAN and Memory Retrieval**

LAN has been intensively studied in relation to memory retrieval (e.g., Altmann, 2004; Altmann & Kamide, 2009; Bone et al., 2019; de Groot, Huettig, & Olivers, 2016; Foerster, 2018; Johansson, Holsanova, Dewhurst, & Holmqvist, 2012; Johansson, Holsanova, & Holmqvist, 2006; Johansson & Johansson, 2014, 2020; Johansson, Oren, & Holmqvist, 2018; Jones, Kuipers, Nugent, Miley, & Oppenheim, 2018; Kinjo, Fooken, & Spering, 2020; Kumcu & Thompson, 2020; Laeng, Bloem, D'Ascenzo, & Tommasi, 2014; Martarelli & Mast, 2011, 2013; Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Scholz, Klichowicz, & Krems, 2018; Scholz, Mehlhorn, & Krems, 2016; Spivey & Geng, 2001; Wantz, Martarelli, & Mast, 2016; Wynn, Ryan, & Buchsbaum, 2020). In an experiment by Richardson and Spivey (2000), participants listened to semantic statements while a visual symbol appeared in one spatial location on a computer screen. For example, participants heard the sentence "Claire gave up her tennis career when she injured her shoulder" while a spinning cross was presented in the top-left quadrant of the screen. After listening to four different sentences, each associated with a different spatial position, the retrieval phase followed. That is, the screen went blank and participants answered a question about one of the presented sentences. Participants exhibited LAN; they looked more often to the area associated with the probed sentence during the retrieval phase than to any other area.

The LAN behavior can be explained as resulting from an overlap between processes engaged at encoding and retrieval of information stored in memory. During the encoding of the information, spatial information of where the information was presented (e.g., a quadrant on a computer screen) is stored along with semantic information (e.g., Claire's injury) in an episodic memory representation (in the form of object–location bindings). Retrieving parts of the episodic trace, for example, by probing parts of the stored information, leads to the reactivation of associated location information. This, in turn, elicits an eye movement to the location where a visual object was presented during encoding, even if it is no longer present during the retrieval or decision-making phase (for an overview see Wynn, Shen, & Ryan, 2019).

**Memory-Based JDM**

Memory processes have been investigated in a variety of decision and judgments tasks (e.g., Albrecht, Hoffmann, Pleskac, Rieskamp, & von Helversen, 2020; Bröder, Newell, & Platzer, 2010; Erickson & Kruschke, 1998; Hoffmann, von Helversen, & Rieskamp, 2013, 2014, 2016; Juslin, Jones, Olsson, & Winman, 2003; Juslin, Karlsson, & Olsson, 2008; Juslin, Olsson, & Olsson, 2003; Medin & Schaffer, 1978; Nosofsky, 1988; Nosofsky & Palmeri, 1997; Olsson, Enkvist, & Juslin, 2006; Persson & Rieskamp, 2009; Rieskamp & Otto, 2006; Rouder & Ratcliff, 2004; von Helversen & Rieskamp, 2008, 2009). The influence of memory in these tasks is frequently described by exemplar-based accounts (Hahn & Chater, 1998; Pothos, 2005). There are a variety of exemplar-based models for JDM, but the common idea is that new objects are classified by relying on memory for similar instances, so-called exemplars. For instance, when evaluating the suitability of a job applicant, an exemplar-based account would assume that people retrieve similar job candidates (exemplars) from memory and how well they performed in a previous interview (criterion). The decision is then based on the category of the most similar exemplar. Theoretical accounts postulate that exemplars are stored as episodic memory traces in long-term memory (Dougherty, Gettys, & Ogden, 1999; Estes, 1986; Hintzman, 1984, 1986; Nosofsky, 1988). Similarity is determined by matches or distances between features of the target object and the exemplar. The features or attributes of the target object then function as retrieval cues for the exemplars. The higher the similarity between an exemplar and a test item, the higher its activation in memory.

Scholz et al. (2015) studied exemplar-based categorizations with LAN. They tested the hypothesis that if people use their memories about past instances to make decisions, LAN should reflect the similarity between decision options and exemplars stored in memory. Participants first memorized attribute values of four job candidates and how well they performed in a previous interview. In subsequent test trials, they judged the suitability of new candidates that varied in their similarity (i.e., the number of shared attribute values) to the previously learned exemplars. Test items were presented auditorily while participants saw only the empty rectangles of the trained exemplars on the screen. Results showed that when using an exemplar-based decision strategy, participants fixated longer on the previous location of exemplars that resembled the new candidates than on the location of dissimilar exemplars. A later study using a judgment task established a direct link between LAN and the resulting behavioral judgment (Rosner & von Helversen, 2019). The more participants looked toward high-performing exemplars, the higher their judgments, suggesting that LAN is directly linked to exemplar-based memory processes in JDM. It reflects the similarity between new and past instances stored in memory. Before we explain in more detail why similarity and other task factors might influence LAN, we first introduce the framework of a shared priority map.

## A Shared Priority Map of Attention During JDM

There is general agreement that both what people see in the world (i.e., salient pieces of information) and what they retrieve from memory determine where they look (for an overview see Kowler, 2011). How the cognitive system handles this issue is arguably a difficult computation (Itti & Koch, 2001). On the basis of previous research, Theeuwes and colleagues (2009) proposed a solution by assuming that different sources of activation create neural signals in a spatially organized representation held in memory, a so-called shared priority map. The spatial location that receives the highest neural activation then determines the target for the next fixation (Figure 1).

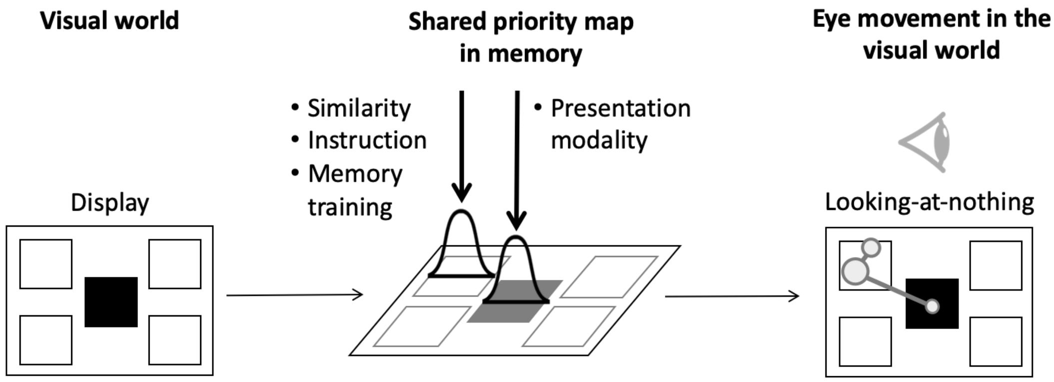


*Figure 1.* Schematic overview of factors influencing attentional allocation during memory-based judgment and decision making (JDM). Left: A typical information board used to study eye movements during JDM. Middle: Assumed memory processes. The spatial location that receives the highest neural activation determines the target for the next fixation. Right: Expected eye movements based on looking-at-nothing. Gray dots represent fixations, with larger dots indicating longer fixations.

Empirical evidence for an interaction effect of memory-driven and stimulus-driven attentional allocation on eye movements was found, for instance, by Hedge and Leonards (2013). In their study, participants had to trace the position of an invisible dot in an empty 3 × 3 grid with several updating steps. Each updating operation was indicated by an arrow cue appearing in the center of the screen that stayed visible until participants requested the next update. For instance, a blue dot was initially located in the top-right position of the grid. The arrow cue, always presented in the center of the screen, pointed toward the bottom of the screen. The arrow indicated that the invisible blue dot moved to the middle right position on the grid. Thus, the participant needed to mentally update the position of the dot from the top-right to the middle-right grid position. People looked both at the visible arrow location and at empty previous and new locations of the updated object. This suggests that eye movement behavior was guided by two signals. First, by information that was visible on the screen (the arrow indicating the next updating operation), and second by information retrieved from memory leading to LAN. In Hedge et al. (2015), the authors explained these findings in light of a shared priority map. Visual information presented in the center of the screen leads to an activation peak in the center of the shared priority map. The previous and next locations of the to-be-updated object are activated in working memory and thereby create activations in the previous and next locations represented in the shared priority map (i.e., the top-right and middle-right grid positions). The location on the map with the highest resulting activation then determines the spatial location toward which the next eye movement is directed in a winner-takes-all fashion.

From the perspective of our proposed framework, the results of Scholz et al. (2015) can be explained by the target object activating similar exemplars in memory, which increases activations of the spatial locations of these exemplars in the shared priority map and ultimately leads the eyes to the empty spatial locations of the most similar exemplar. That is, the higher the similarity between a test item and an exemplar, the higher the activation of locations of similar exemplars in memory (e.g., Dougherty et al., 1999) on the shared priority map and, consequently, the more LAN can be observed. In Scholz et al. (2015), the screen was almost devoid of any visual information, because information about the test items was presented auditorily; eye movements thus reflected the memory-driven activations stemming from retrieving similar instances from memory.

In the following we consider how manipulating different sources of activation in the shared priority map should influence LAN according to the framework. We first consider two manipulations likely to affect the strength of memory-driven activations—memory training and task instructions—and then one manipulation likely to affect the strength of activations competing with memory-driven activations—the presentation format.

***Memory Training***

In memory-based JDM, participants usually work through a fixed number of training rounds. In each round each exemplar is judged once and receives outcome feedback (Hoffmann et al., 2013, 2014, 2016; Juslin et al., 2008; Juslin, Olsson, & Olsson, 2003). For instance, in the study by Hoffmann et al. (2014), exemplars were presented 10 times before participants entered the judgment phase. The use of exemplar memory becomes more likely, the more accurately participants can recall the exemplars (Hoffmann et al., 2014; Hoffmann, von Helversen, Weilbächer, & Rieskamp, 2018; Johansen & Palmeri, 2002). This enhancement may be explained through a more distinct memory representation (Rouder & Ratcliff, 2004) of the seen exemplars or by a larger number of stored memory traces for exemplars (Dougherty et al., 1999; Nosofsky & Palmeri, 1997). Since memory traces compete for retrieval in each trial (Nosofsky & Palmeri, 1997), a larger number of memory traces will reduce errors in decisions based on the retrieval of exemplars. Accordingly, if exemplars are learned by heart during training, this should increase their memory strength and activation in memory.

In most LAN studies in the JDM literature, information about exemplars was learned by heart. For instance, participants had to correctly retrieve all attribute values in two consecutive rounds before proceeding with the experiment (Scholz et al., 2015; see also Jahn & Braatz, 2014; Platzer et al., 2014; Renkewitz & Jahn, 2012, for similar procedures). This way participants had sufficient memories for both the exemplar information and the associations with spatial locations on the screen. Whether an increase in memory strength will influence LAN is unclear. Several studies have shown that LAN reliably occurs without extensive memory training for objects and their presentation locations (Hoover & Richardson, 2008; Johansson et al., 2012, 2006; Kumcu & Thompson, 2020; Laeng & Teodorescu, 2002; Martarelli & Mast, 2011, 2013; Scholz et al., 2018, 2016). So far, only indirect evidence suggests that LAN may be sensitive to how well information is learned. It has been observed that LAN is more likely to occur when giving a correct than an incorrect response (Martarelli & Mast, 2011; Scholz et al., 2016). Last but not least, the relationship may have an inverted-U shape, as LAN can diminish when information is overlearned or responses can be given very quickly (Jones et al., 2018; Scholz, Mehlhorn, Bocklisch, & Krems, 2011; Wantz et al., 2016). Given the framework of the shared priority map and the idea that stronger representations lead to stronger activations in memory, we assume that the more strongly the exemplars are represented in memory, the stronger their activations on the shared priority map will be and consequently, the more likely LAN is to reflect similarity-based activations from memory.

***Task Instructions***

On the one hand, there is generally wide agreement that exemplar retrieval takes place automatically (e.g., Johansen & Palmeri, 2002), that is, without an explicit instruction to use an exemplar-based strategy (Hoffmann et al., 2013, 2014, 2016; Juslin et al., 2008; Juslin, Jones et al., 2003; Scholz et al., 2015). On the other hand, using exemplar memory during JDM can also be a deliberate decision strategy (Karlsson, Juslin, & Olsson, 2008) and can be explicitly instructed (e.g., Albrecht et al., 2020; Olsson et al., 2006; Scholz et al., 2015). Deliberate thinking about exemplars can make exemplar retrieval more likely, by increasing memory activations (Dougherty et al., 1999). Consequently, an explicit task instruction may activate similar exemplars in memory more strongly in comparison to leaving the process spontaneous.

Eye movements more generally seem to be sensitive to task instructions. Shih, Meadmore, and Liversedge (2012) presented participants with photographs of household items with or without participants being aware that their memories about the seen objects will be tested. Under an explicit memory instruction, older adults exhibited more memory-enhancing viewing behavior, indicating that eye movements and memory can be modulated through task instructions. It is unclear if LAN is sensitive to task instructions. In studies on mental imagery in which LAN has often been observed, participants received an explicit instruction to imagine a seen object with their “inner eye,” before being probed on some of the object features (Bone et al., 2019; Brandt & Stark, 1997; Johansson et al., 2006; Kosslyn, 1994; Martarelli & Mast, 2013; Wantz et al., 2016). However, leaving the instruction implicit can also elicit LAN (Altmann, 2004; Renkewitz & Jahn, 2012; Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Rosner & von Helversen, 2019; Scholz et al., 2015).

To our knowledge, nobody has yet explored if task instructions make a difference in LAN. Given the framework of the shared priority map and research showing that effortful deliberations can increase memory activation for exemplars, we assume that an explicit task instruction to mentally imagine the most similar exemplar would increase activations for this exemplar. Thus, with explicit activations LAN should more strongly reflect similarity-based activations from memory.

***Presentation Modality***

In the majority of previous LAN studies, test items were presented auditorily to leave the screen almost devoid of any visual information (the blank screen paradigm, e.g., Altmann, 2004), which is in line with literature suggesting a tight link between eye movements and the processing of auditorily presented items (Huettig, Mishra, & Olivers, 2012; Huettig, Olivers, & Hartsuiker, 2011). However, this procedure stands in contrast to more traditional paradigms used to study (exemplar-based) decision processes. There, new test items are presented visually, as either verbal descriptions or pictures (Bröder et al., 2010; Hoffmann et al., 2013, 2014, 2016; Juslin, Jones et al., 2003; von Helversen & Rieskamp, 2009).

What does this mean for observing LAN in JDM? With an auditory stimulus presentation it is possible that language processing contributes to LAN in addition to the retrieval of similar exemplars from memory. Although, Scholz et al. (2015) showed that when people used a rule-based strategy that did not draw on memory retrieval to the same extent as an exemplar-based strategy, people did not show LAN even when test items were presented auditorily, which suggests that language processing on its own is not enough to drive LAN. Nevertheless, language processing may add to activations stemming from memory retrieval and thus influence whether LAN occurs and how much it reflects similarity-based activation.

One way to remove additional activations resulting from processing auditorily presented information when asking participants to rely on the retrieval of exemplars during JDM could be to present information about test items as visually displayed verbal descriptions. However, given that new visual information is a strong cue for inducing visual attention (e.g., Chun et al., 2011), visual information may compete with memory-driven activations on the shared priority map. In turn, LAN to the exemplar locations may become weaker or vanish completely because people might look only at the screen location where this salient and relevant visual information is presented. Indeed, a decrease in LAN for visual materials has been reported by Jones et al. (2018), who tested LAN in a cued recall task to study the ability to learn new phonological associations. LAN during the recall of learned associations did not occur in about one third of all trials. Accordingly, when decision items are presented visually, activations resulting from the visual presentation format may win out in guiding eye movements, leading to a decrease in eye movements guided by memory retrieval. That is, even though people would still be retrieving exemplars from memory, this might no longer be reflected in their eye movements. However, the framework of the shared priority map also suggests that if visual information is removed after it was perceived, memory processes may guide eye movements again, thus increasing LAN behavior. Indeed, recently Wynn et al. (2020) found that eye movements reflected memory processes after the removal of visual stimuli from screen.

**The Present Research**

The goal of this research was to test to what extent eye movements during JDM are driven by different sources of activation. We explored this question in two experiments using a categorization decision task. Participants needed to classify objects described on several attributes into one of two categories. In this task, people often rely on an exemplar-based decision strategy, making it ideal for our purposes (Hoffmann et al., 2016; Juslin et al., 2008; Juslin, Jones et al., 2003; Karlsson et al., 2008; Nosofsky & Palmeri, 1997; Persson & Rieskamp, 2009; von Helversen & Rieskamp, 2009).

In Experiment 1 we considered presentation format (auditory vs. visual) and memory training (exemplar training and/or criterion learning). In Experiment 2, we investigated the effect of explicit mental imagery by instructing participants either to use an intuitive strategy or to retrieve exemplars deliberately. In both experiments we varied the similarity between the decision object under evaluation and the trained exemplars to investigate how strongly LAN reflects similarity-based memory activations.

Previous LAN research has mainly focused on testing the quality of LAN. Here, we extended this approach and considered two main dependent variables: in terms of quality, LAN strength and, in addition, a quantity measure, LAN occurrence.

LAN strength measures the extent to which people look more at screen locations associated with the retrieved pieces of information ("relevant locations") than to locations that are irrelevant for a given trial (e.g., Johansson et al., 2012, 2006; Martarelli & Mast, 2011; Scholz et al., 2016) and thus reflects the memory-driven activation strength of information. Accordingly, in line with previous research (Rosner & von Helversen, 2019; Scholz et al., 2015), in the present study LAN strength should be a function of exemplar similarity. More precisely, we expected that the higher the similarity between a test item and an exemplar, the more participants would look to the location of the most similar exemplar and the less they would look to less similar (i.e., irrelevant) exemplar locations. Furthermore, taking into consideration the findings of Rosner and von Helversen (2019), we assumed that the higher the LAN strength for a category decision, the more likely it would be that this category would be chosen. Our new predictions are that LAN strength might also be higher with exemplars that are more readily available in memory, which is the case with more memory training, and with an explicit imagery instruction.

To measure LAN quantity, we looked at LAN occurrence. We defined LAN as occurring if participants looked at least once toward one of the exemplar locations during decision making.[1] If, as proposed in the framework of the shared priority map, the location of the visually presented stimulus material on the screen wins the competition for participants' attention, we should see LAN reduced in its quantity. That is, participants will not look at the exemplar locations, regardless of the extent to which they retrieve exemplars from memory. Conversely, LAN will be more likely to occur the less visual information is presented on the screen and when visual information is removed from the screen.

## Experiment 1A and B

The participants' task was to decide whether to invite job candidates for an interview or reject them based on information on four attributes. They first learned (via feedback) in a criterion training phase which previous job candidates (the training exemplars) had been invited and then continued with a test phase, in which they had to make decisions for new candidates. Experiment 1 manipulated the amount of training participants received prior to this training phase. They either first learned the attributes of the exemplars by heart (exemplar learning) or started directly with the decision training phase (only criterion learning). Training

[1] Richardson and Spivey (2000) applied this measure to test if participants answered more correctly when looking at least once to the relevant location. Here, we aimed to use it as an indicator of whether the behavior is shown at all.

was varied between participants. Furthermore, we manipulated the presentation format during the decision-making phase. In one condition, information about new job applicants (test candidates) was presented auditorily while the screen contained only empty rectangles. In the other conditions, test candidates were presented visually, that is, as verbal descriptions in the center of the screen. As illustrated in Figure 2, we had three conditions in Experiment 1A (visual with exemplar learning, auditory, and visual without exemplar learning); in the visual conditions, visual presentation lasted until participants decided. In Experiment 1B we added two more visual conditions, in which visual presentation of attribute values ended after a fixed amount of time to test whether LAN would occur when visual stimuli were removed. Presentation format was varied between participants. Test candidates differed in their similarity in attribute values (defined as the number of matches) to the training exemplars. The factor similarity thus varied within participants. Eye movements were recorded during the test phase. At the end of the experiment, all participants were tested on their long-term memories about the location of attribute values (location memory test). The experiment lasted on average 33.2 min (*SD* = 12.4). The study design and methods were approved by the ethics committee of the University of Zurich. All materials and data are available at https://osf.io/zpc2s/.

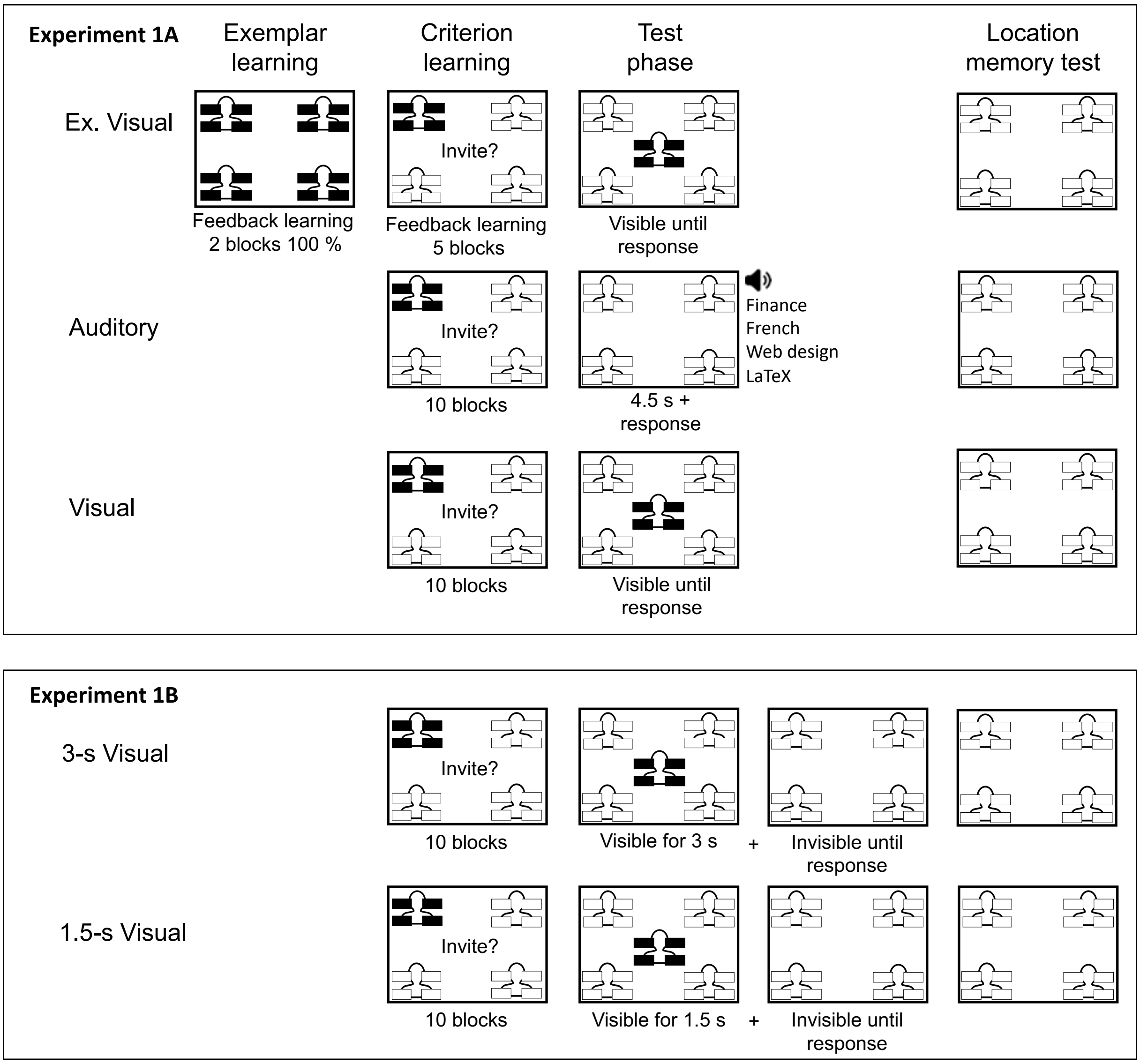


*Figure 2.* Experimental procedure. All participants worked through the criterion learning phase, test phase, and location memory test. Additionally, participants in the visual with exemplar learning condition (Ex. Visual condition, top row) received exemplar training. The presentation modality during the test phase varied between conditions. Solid black frames contained attribute values. Finance, French, Web design, and LaTeX refer to attributes. See text for details.

## Participants

Previous studies found large effects (e.g., $\eta_p^2 \approx .29$) regarding the influence of similarity on eye movements (Scholz et al., 2015). A sample of at least 23 participants in each condition was needed to ensure that we would have appropriate power to find an effect of medium size of similarity on LAN ($\eta_p^2 \approx .07$ requires N=23 to reach a power of 80%; see Faul, Erdfelder, Lang, & Buchner, 2007).

Overall, 134 participants at the University of Zurich took part in the study for course credit or financial compensation [15 Swiss francs (CHF) per hour]. One participant did not meet the learning criterion in the first experimental phase and due to a technical error, data of another participant were corrupted. In total, we could analyze the data of 132 participants (Experiment 1A: 78 participants, 23 male, $M_{\text{age}}$ = 26.8 years, range 19–57 years; Experiment 1B: 54 participants, 18 male, $M_{\text{age}}$ = 26.2 years, range 18–52 years[2]). All participants had normal or corrected-to-normal vision. Mean tracking accuracy in the test trials was very high at 0.7° of visual angle. All participants signed informed consent forms.

**Apparatus**

Participants were seated in front of a 22-inch computer screen (1,680 × 1,050 pixels) at a distance of 700 mm and instructed to position their head in a chin rest. The eye tracker system SMI iView RED sampled data from the right eye at 500 Hz and recorded with iView X 2.8 following a five-point calibration. Fixation detection was done with IDF Event Detector 9 (SMI, Teltow) using a peak velocity threshold of 30°/s and a minimum fixation duration of 80 ms.

**Materials**

Study materials of Experiment 1 consisted of four training exemplars and 44 additional test items (test candidates). All items contained information on four attributes (Figure 3): previous work experience (with the attribute values automotive industry, building industry, finance, and food service), language skills (with the attribute values French, Mandarin, Russian, and Spanish), professional training (with the attribute values human

---

[2] Because of a technical error, demographic data were not recorded for two participants.

resources, marketing, Web design, sales), and possession of computer skills (with the attribute values HTML, SQL, LaTeX, and GIMP). Test items varied in their *similarity* to the exemplars, ranging from two to four matches in attribute values with one exemplar (two matches: 24 items; three matches: 20 items; four matches: 4 items). For instance, a test item that matched on two attributes with one training candidate from one category matched on two other attributes with one training candidate from the other category. Accordingly, items with two matches were always ambiguous. Another 20 items shared three values with one training candidate and one value with another training candidate. That is, they were either more similar to an invited or a rejected training candidate (10 items each). Additionally, the four training items were included in the test set. Test materials were fully balanced. That is, each attribute value was tested similarly often and each training candidate was equally often the training candidate with the highest number of matches (see Appendix A for a full list of items).

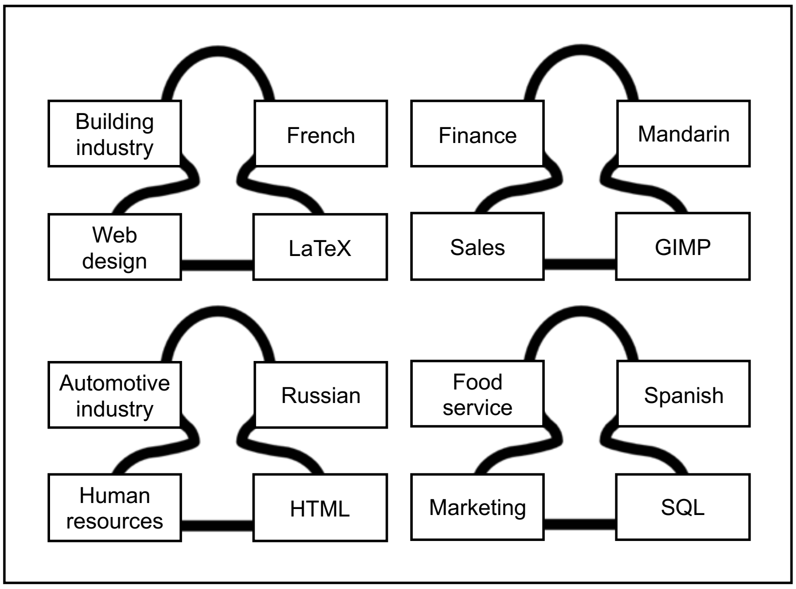


*Figure 3.* Exemplars consisting of information about four training candidates (two invited and two rejected exemplars). Each exemplar was presented in one of the four screen quadrants. In this example, candidates in the top row were invited and in the bottom row rejected. Rectangles contain attribute values (see text). Note that the size of the exemplar on the screen is increased to increase readability.

Each exemplar was presented in one of the four screen quadrants. The distance from the center of the screen to the center of each of the four exemplars was 9.62° of visual angle (576 pixels). Attribute values were presented as black text in rectangles with white borders and a light gray background. Each rectangle had a size of 2.84° × 1.17° of visual angle (170 × 70 pixels). The center of each of the four rectangles containing the information describing one exemplar had a distance of 2.21° of visual angle (132 pixels) from the center of each quadrant. Visual materials were presented in four balanced orders, varying the positions of the exemplars on the screen and the order of attribute values between participants. During the subsequent test phase, participants saw only the empty rectangles on the screen. In the condition with auditory stimulus presentation during the test phase, stimuli were read aloud from left to right and from top to bottom, following the visual presentation of exemplars during criterion learning. In the exemplar and criterion learning phases, feedback was provided with a high- and low-pitched tone.

**Procedure**

At the beginning of the experiment, the eye tracker was calibrated to check if eye movements could be recorded to a sufficient quality (< 1.5° of visual angle).

***Exemplar Learning***

Participants in the condition with exemplar learning (henceforth called ex. visual, see Figure 2, top row) were instructed to first memorize the attribute information about the four exemplars presented in rectangles on the screen. To do this, they first saw all the information on the screen and could study it until they pressed a button to start a training phase in which they could test their knowledge about the four training exemplars. During these exemplar learning trials, only empty rectangles were visible. Attribute values were presented auditorily and in random order. In each trial, participants heard the value of an attribute (e.g., "French") and had to indicate the rectangle to which it belonged by clicking with the mouse on the rectangle (e.g., the top-right rectangle of the top-left candidate; see Figure 3). They received visual and auditory feedback on their response.[3] One training block consisted of a test of all 16 pieces of information (4 exemplars × 4 attribute values). After four training blocks, all the attribute information became visible again. Learning continued until all the information was remembered correctly in two consecutive blocks or after 20 blocks with the last 2 blocks resulting in an accuracy of at least 95% correct. Participants in the Ex. Visual condition took on average 15.6 min (*SD* = 3.3) and on average 6.8 blocks (*SD* = 1.3) to complete the task.

***Criterion Learning***

In the criterion learning phase, participants learned how each exemplar had been evaluated by making decisions and receiving feedback. In each criterion learning trial, participants first indicated that they were ready for the next trial. Next, they looked at a fixation cross in the center of the screen for 1.5 s. Then, participants saw the attribute values of one training candidate while the rectangles of all the training candidates remained empty (see Figure 2, second column). After 5 s, they were prompted to indicate their choice by pressing the left mouse button to invite the candidate or the right button to reject. There was no time restriction during criterion learning. Afterward, they received visual and auditory feedback. A training block ended after each candidate had been judged once. Participants in the Ex. Visual condition, who had already received training on the attribute values during exemplar learning, worked through five training blocks. Participants in all other conditions worked through 10 training blocks. These participants additionally received the verbal instruction that they should also memorize the attribute values as they might need this information for the test phase in which they would have to decide about new job candidates. Criterion learning lasted on average 7.3 min (*SD* = 0.4) in the Ex. Visual condition and 14.5

[3] Detailed descriptions of the feedback procedure can be found in the online supplemental materials at https://osf.io/zpc2s/.

min (*SD* = 4.0) in all other conditions. Over all conditions, after an average of 2.4 blocks (*SD* = 1.1), participants had learned to correctly classify all candidates.

### *Test Phase*

At the beginning of each trial, participants indicated if they were ready to start the next trial. This was followed by a fixation cross that appeared in the center of the screen for 1.5 s. Next, participants saw the screen with the empty rectangles of the four candidates (see Figure 2, third and fourth columns). They were presented with attribute values of a test candidate and instructed to decide whether to invite the candidate. The presentation modality during test trials varied between conditions. In the Ex. Visual and Visual conditions, test items were presented in the center of the screen and were visible until the participant responded. In the Auditory condition, the center of the screen was empty. Participants heard the attribute values of the new test item in sequential order. In Experiment 1B, new test items were visible on the screen but disappeared either after 3 s (3-s Visual) or after 1.5 s (henceforth called 3-s Visual and 1.5-s Visual conditions). Presentation times were pretested in a pilot experiment and chosen to be lower than typical response times in these tasks but long enough to be readable. After the attribute information for the test items had disappeared, participants saw only the empty rectangles. Participants judged the 48 test items once. Over all conditions, the test phase lasted on average 7.9 min (*SD* = 1.1). Eye movements were recorded throughout the test phase.

### *Location Memory Test*

At the end of the experiment, participants were asked to remember the attribute values of the training candidates. Therefore, they saw the screen with the empty rectangles of the four training candidates (Figure 2, last column). Attribute values were presented auditorily and in random order. After hearing an attribute value, participants had to click on the rectangle where they thought the attribute value was presented during criterion learning (and during exemplar learning for participants in the Ex. Visual condition). The memory test ended after participants responded to all 16 attribute values once and without feedback.

**Results**

The aim of Experiment 1A was twofold: to test (a) if learning exemplar information by heart versus incidentally during criterion learning leads to more LAN in terms of the strength of the phenomenon, and (b) if LAN is reduced in its occurrence when new test items are presented visually in comparison to auditorily. Experiment 1B aimed to test if reducing the duration of visual information presentation can increase the occurrence of LAN. We furthermore aimed at directly linking LAN to categorization behavior.

In a first step, we analyzed how well participants performed in the categorization task, as our manipulations may affect not only LAN but also performance in the task itself or the location memory performance. We assumed that memory training could lead to better memories about exemplars in comparison to incidental learning, which could result in higher categorization accuracy and location memory performance. Whereas test items that were identical to trained exemplars may be classified more rapidly in the visual conditions, because of the sequential nature of auditorily presented materials, response time may be prolonged in the Auditory condition, which in turn could influence the probability of LAN occurring. Analyzing performance measures is thus important to understand LAN during JDM.

***Preparatory Data Analyses and Rationale for the Analyses***

Seven participants performed at chance level or worse during the test phase and were therefore excluded from further analyses. For the remaining 125 participants and trials, five trials (0.08% of trials) were excluded because response times were longer than 40 s. Note, mean response times were much lower: $M$ = 5.4 s ($SD$ = 1.6). In 21 trials (0.35% of trials), no gaze data were recorded (e.g., due to participants closing their eyes or looking off the screen); these trials could therefore not be analyzed.

For the gaze analyses, we drew four rectangular areas of interest (AOIs) around each of the four exemplar locations and the center of the screen where the test items were presented in the visual conditions (see Figure 2). Each exemplar AOI had a size of 7.69° × 4.01° of visual angle (460 × 240 pixels). The size of the exemplar AOI exceeded the outer borders of each of the four rectangles describing one exemplar by 0.25° of visual angle (15 pixels).

We used a mixed-model approach to analyze the data. Binomially distributed dependent variables (e.g., categorization accuracy) were analyzed with a generalized linear mixed model (GLMM) analysis with a logistic link function and Laplace approximation of parameter values. *P* values were estimated with the likelihood ratio test and sum-of-squares contrast coding. If the predicted variable was continuous (e.g., response times), we used linear mixed modeling. The latter models were fitted using residual maximum likelihood estimation. Fixed effects were evaluated via Satterthwaite approximation of degrees of freedom and sum-of-squares contrast coding. We aimed at implementing the maximal random effects structure justified by the design but also had to take model complexity into account (Barr, Levy, Scheepers, & Tily, 2013). With the exception of the analysis of LAN occurrence, which contains only by-subject random intercepts, all other analyses included by-subject and by-item random intercepts. Analyses were performed with R (R Core Team, 2018) and the following packages: lme4 (Bates, Maechler, Bolker, & Walker, 2015), afex (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2020), and estimated marginal means with emmeans (Lenth, 2020).

### *Categorization and Location Memory Performance*

**Categorization Accuracy.** To analyze how well participants performed in the categorization task (correct response = 1, wrong response = 0), we excluded ambiguous items that were equally similar to an invited and a rejected exemplar. We included fixed effects for the five conditions (Ex. Visual, Auditory, Visual, 3-s Visual, 1.5-s Visual), for the similarity between test candidates and exemplars (three and four matches), and for their interaction.

Categorization accuracy varied between conditions, $\chi^2(4) = 13.17$, $p = .01$, and between levels of similarity, $\chi^2(1) = 9.96$, $p = .002$. Factors did not interact: $\chi^2(4) = 2.31$, $p = .68$. Post hoc analyses comparing the Ex. Visual condition against all other incidental learning conditions revealed that participants were more accurate in the categorization task when they worked through the exemplar learning phase, $z = 2.55$, $p = .011$, in line with the idea that more memory training leads to higher activations of similar exemplars in memory. Furthermore, participants were better when information stayed visible on the screen, as in the Visual condition, compared to when information was removed, as in the Auditory, 3-s Visual, and 1.5-s Visual conditions (comparison of the Visual condition against the group of the Auditory, 3-s Visual, and 1.5-s Visual conditions: $z = 1.99$, $p = .046$; see Table 1). Removing information from the screen can lead to participants forgetting attribute information, which increases error rates.

Table 1

*Means (and Standard Deviations) of Performance and Looking-at-Nothing (LAN) Behavior for Experiment 1A and B in Test Trials and the Location Memory Test*

| Variable | Experiment 1A condition | | | Experiment 1B condition | |
|---|---|---|---|---|---|
| | Ex. Visual ($N$ = 25) | Auditory ($N$ = 26) | Visual ($N$ = 25) | 3-s Visual ($N$ = 25) | 1.5-s Visual ($N$ = 24) |
| Test trials | | | | | |
| Categorization accuracy | 94 (9) | 82 (11) | 88 (10) | 84 (12) | 84 (12) |
| Categorization response time | 5.68 (1.72) | 7.19 (1.01) | 5.11 (1.26) | 4.51 (1.48) | 4.38 (1.12) |
| LAN occurrence | 0.09 (0.09) | 0.71 (0.32) | 0.05 (0.09) | 0.22 (0.23) | 0.37 (0.29) |
| LAN occurrence ($N$) | 21 | 25 | 17 | 21 | 21 |
| LAN strength | 0.34 (0.32) | 0.39 (0.14) | 0.38 (0.38) | 0.26 (0.15) | 0.31 (0.14) |
| Location memory test accuracy | 99 (3) | 71 (22) | 70 (27) | 66 (29) | 78 (21) |

*Note.* Mean percentage correct responses were calculated leaving out ambiguous items. Mean response times and LAN measures are based on all items. LAN occurrence describes mean proportion of LAN trials. The number of participants showing LAN (LAN occurrence $N$) indicates the number of participants who looked at least once in one experimental trial at one of the exemplar locations. LAN strength describes the participants' mean proportion of fixations on the relevant exemplar for trials in which LAN occurred and random selection of one area of interest for ambiguous items; chance level is .25. Location accuracy describes mean percentage correct during the location memory test on the exemplar level. The Auditory and 1.5-s Visual conditions had the strongest observed LAN behavior. Ex. Visual = Visual with exemplar learning.

**Categorization Response Times.** Categorization response times were measured from the onset of the screen containing attribute information in the visual conditions and the onset of the auditory attribute presentation in the Auditory condition until participants gave their response. For the comparison of categorization response times, we analyzed data of all item types (invited, rejected, ambiguous). We used the same effects structure as for the analyses of categorization accuracy. We found a main effect of condition, $F(4, 134.3) = 17.00$, $p < .001$, with the longest response times in the Auditory condition in which information on attribute values was presented sequentially (see Table 1). This was confirmed by significant pairwise contrasts between the Auditory and all other (visual) conditions (all *p*s < .001). Additionally, response times were longer in conditions in which information stayed visible on the screen (Ex. Visual and Visual) in comparison to conditions in which the visible information was removed (3-s Visual and 1.5-s Visual), $z = 2.28$, $p = .02$. Response times were further determined by the similarity between test candidates and exemplars. We found a main effect of similarity, $F(2, 44.99) = 74.36$, $p < .001$, with the slowest response times for ambiguous items (i.e., two matches; $M_{\text{sim2}} = 5{,}861$ ms, $SD = 2{,}029$), followed by items with three matches ($M_{\text{sim3}} = 5{,}029$ ms, $SD = 1{,}563$), and the fastest responses for items resembling the exemplars ($M_{\text{sim4}} = 4{,}411$ ms, $SD = 1{,}742$, all *p*s < .001). The factors condition and similarity interacted, $F(8, 5794.27) = 5.90$, $p < .001$. That is, the observed differences in response times between conditions were most pronounced for test candidates that were identical to the training exemplars.

**Location Memory Test.** We compared the number of correct retrievals in the location memory test at the end of the experiment on the level of exemplars. That is, if a participant clicked on *any one* of the four rectangles belonging to one exemplar that was associated with the attribute value tested in that trial, the response was counted as correct. As there was only one observation per participant, we compared the number of correct retrievals on the exemplar level between conditions with an analysis of variance (ANOVA). Participants significantly differed in location memory performance between conditions, $F(4, 120) = 8.53$, $MSE = 12.84$, $p < .001$ (see Table 1). Participants best remembered the locations of the attribute values in the Ex. Visual condition that included the exemplar learning phase, in comparison to all other conditions (all *ps* < .02). No other contrasts reached significance (all *p*s > .32).

**Covariate Analysis for Accuracy.** Adding response time and location memory test performance as predictors to the analyses of categorization accuracy[4] revealed significant main effects for the four predicting variables: condition, $\chi^2(4) = 14.69$, $p = .005$; similarity, $\chi^2(1) = 9.76$, $p = .002$; categorization response time, $\chi^2(1) = 26.65$, $p < .001$; and location memory test performance, $\chi^2(1) = 49.61$, $p < .001$,[5] and a significant interaction of condition and response time, $\chi^2(4) = 14.69$, $p = .005$. That is, in addition to the effects of condition and similarity, participants performed better in the categorization task when they had better location memories about exemplars (linear contrast: $z = 7.58$, $p < .001$) and when they gave their responses more quickly (linear contrast: $z = -5.08$, $p < .001$).

After observing these differences between the conditions in categorization behavior and location memory performance, we subsequently controlled for categorization accuracy, response times, and location memory performance by adding them as covariates to all following analyses. This improved model fits of models on LAN occurrence and strength, but not LAN and categorization decisions. Still, running models without covariates shows the same pattern of results.

[4] In comparison to the analysis of categorization accuracy, here we added the two continuous covariates: response time and location memory test performance. We used the logarithm of response times to meet the criteria of normal distribution of predictor variables. In addition, response times and location memory test performance were centered to zero.

[5] Please note that given that the location memory task occurred after the categorization task, performance in the location memory task may have been influenced by the performance in the categorization task and thus it is not possible to make claims regarding the directionality of the effect.

***LAN Occurrence***

To test if LAN varies with different presentation modalities, we assessed if a participant looked at least once at one of the four exemplar locations for each trial (coded as 1 if yes, 0 if no). Then, we calculated the proportion of trials in which LAN occurred per person. Figure 4 shows clear differences between the experimental conditions.

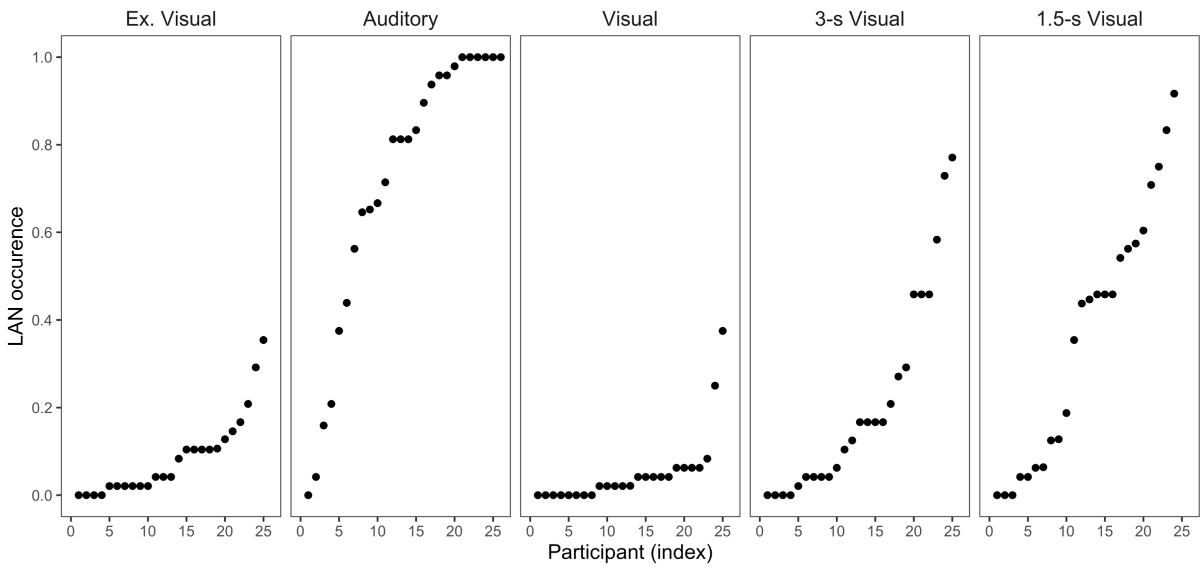


*Figure 4.* Proportion of trials per participant in which looking-at-nothing (LAN) occurred in each of the five presentation conditions. Each point shows the data of one participant.

Corresponding to the previous GLMM analyses, condition and similarity were added as main effects as well as their interaction. As covariates, we included response time and location memory performance as well as the interaction of response time with condition. We found a significant difference in LAN occurrence between the conditions, $\chi^2(4) = 81.50$, $p <$ .001. The most LAN occurred in the condition where information was presented auditorily, followed by the condition in which visually presented information was removed after 1.5 s (pairwise comparison of the Auditory and 1.5-s Visual conditions, $z = 2.69$, $p = .06$). These two conditions—Auditory and 1.5-s Visual—differed significantly from all other conditions (all *p*s < .05). There was no difference in LAN occurrence between the conditions in which participants received exemplar training and those in which information was presented only 10 times (pairwise comparison between Ex. Visual and Visual, $z = 0.41$, $p = .99$). LAN occurrence steadily increased from keeping information visible on screen, to removing information after 3 s (pairwise comparison between Visual and 3-s Visual, $z = -3.24$, $p = .01$), to removing information after 1.5 s (pairwise comparison of 3-s Visual and 1.5-s Visual, $z = -2.76$, $p = .05$).

Participants were slightly more likely to look at the exemplar locations when a test item had identical attribute values to one of the training exemplars, revealing a main effect of similarity, $\chi^2(2) = 6.21$, $p = .04$. However, none of the pairwise comparisons between the levels of the factor similarity reached significance (all *p*s > .09). There was no interaction of similarity with condition, $\chi^2(2) = 15.71$, $p = .05$.

Participants were more likely to show LAN when response times were increased, $\chi^2(1) = 373.21$, $p < .001$ (linear contrast: $z = 17.89$, $p < .001$), interaction of response time and condition, $\chi^2(4) = 80.64$, $p < .001$. That is, the more time a participant took, the higher their chance of hitting one of the exemplar AOIs with their eyes. There was, however, no effect of location memory performance on LAN occurrence, $\chi^2(1) = 0.15$, $p = .70$. Adding categorization accuracy[6] instead of location memory performance to the model revealed the same results pattern.[7]

### *LAN Strength*

For trials in which participants showed LAN, we tested if LAN reflects the similarity between test candidates and exemplars. Therefore, we analyzed LAN strength, defined as fixation proportions based on the number of fixations on the matching exemplar (henceforth called “relevant exemplar”) divided by the summed number of fixations on all four exemplar

[6] Categorization accuracy for ambiguous items was set to 0.5.
[7] Adding exemplar learning duration to the model of LAN occurrence did not change our main results.

locations. If two exemplars were similarly relevant in a given trial (ambiguous items had a similarity of 2), we randomly selected the fixation proportion for one of the two possible exemplars. Figure 5 shows mean proportions of fixation on the relevant exemplar by similarity and condition. Overall, participants looked more at the exemplar locations that shared more attribute values. However, due to a reduced number of participants and trials in which LAN occurred, this pattern is blurred in the Ex. Visual, Visual, and 3-s Visual conditions. Therefore, all following statistical analyses were performed only for the Auditory and 1.5-s Visual conditions.

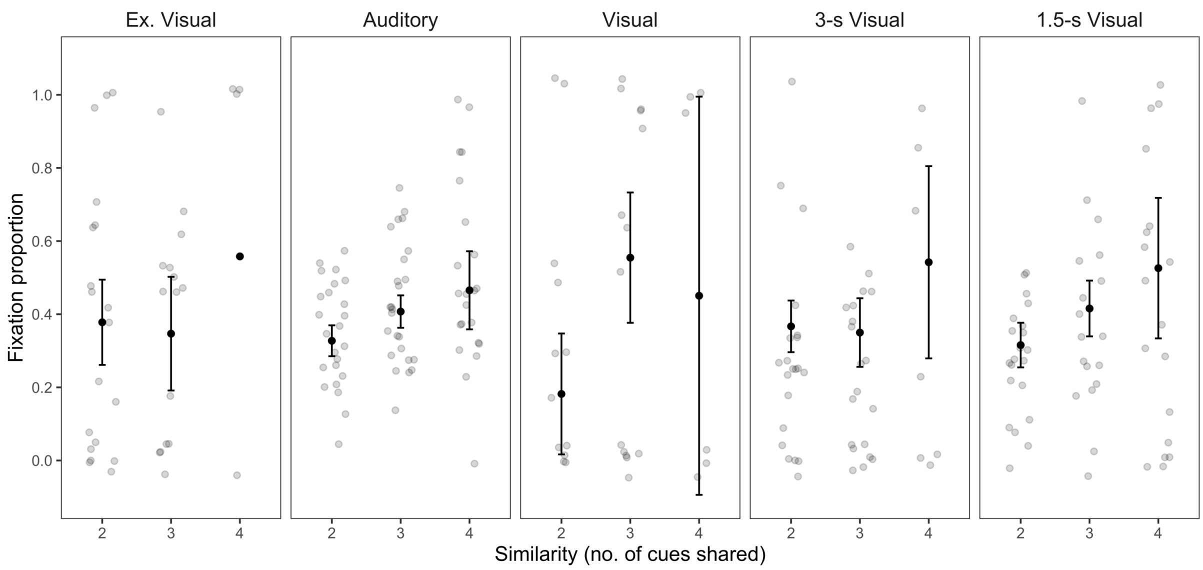


*Figure 5.* Mean proportions of fixation on the relevant exemplars for test items that shared two, three, or four attribute values with the relevant exemplar for the presentation conditions of Experiment 1. Standard errors show within-subject 95% confidence intervals (Morey, 2008). Gray jittered dots in the background show the individual participants' means.

To test LAN strength, we used the same fixed effects structure as for the analysis of LAN occurrence. We found a main effect of similarity, $\chi^2(2) = 16.58$, $p < .001$. There was neither an effect of condition, $\chi^2(1) = 0.85$, $p = .36$, nor an interaction of condition and similarity, $\chi^2(2) = 0.23$, $p = .89$. That is, in both conditions, LAN strength reflected the activation of exemplars from memory.

Concerning the covariates, there was no effect of location memory performance, $\chi^2(1) = 0.03$, $p = .86$. Longer response times lead to significantly reduced LAN strength, $\chi^2(1) = 10.52$, $p = .001$ (linear contrast: $z = -3.24$, $p = .001$), but there was no interaction between condition and response time, $\chi^2(1) = 1.82$, $p = .18$. Thus, whereas longer response times increased the chance of hitting one of the exemplar locations with the eyes (see results on LAN occurrence), they reduced the chance of looking proportionally more at the most similar exemplar. One reason for this could be that in trials with longer response times, activations for the most similar exemplars differ less from the activations of the other exemplars (due to the probabilistic nature of the process). This should reduce LAN strength but also categorization accuracy. Indeed, categorizations were more accurate with shorter response times (see covariate analysis for accuracy). In the same vein, adding categorization accuracy instead of location memory performance to the model revealed that participants showed higher LAN strength the higher their categorization accuracy, $\chi^2(1) = 29.48$, $p < .001$, while the results pattern of all other factors in the model was the same.

In the previous analysis, we could not study how LAN strength was affected by the difference between the memory training conditions (i.e., exemplar vs. criterion learning), because in the exemplar learning condition, too few participants showed LAN. To test for possible differences, we conducted a reanalysis of data from Scholz et al. (2015) comparing LAN strength in the Auditory condition of this study that had only a criterion learning phase with two test phases of Scholz et al. (2015) that included an exemplar learning phase in addition to criterion learning and also had an auditory presentation format. Test materials and the learning procedure were the same as in this study. Participants first learned attribute values of three attributes about four (Experiment 1, $N_{\text{Exp.1}} = 26$[8]) or eight (Experiment 2, $N_{\text{Exp.2}} = 28$) exemplars, where the attribute direction (i.e., the sign of an attribute–criterion relation, Bröder et al., 2010; von Helversen, Karlsson, Mata, & Wilke, 2013; von Helversen & Rieskamp, 2009) was unknown, presented at four spatial locations on a screen. During test trials they were auditorily presented with each training candidate once. Figure 6 shows proportions of fixation on the relevant exemplar under the instruction to use exemplars to make the decision. Visual inspection of mean values and confidence intervals in Figure 6 indicates that there is no difference between the conditions.

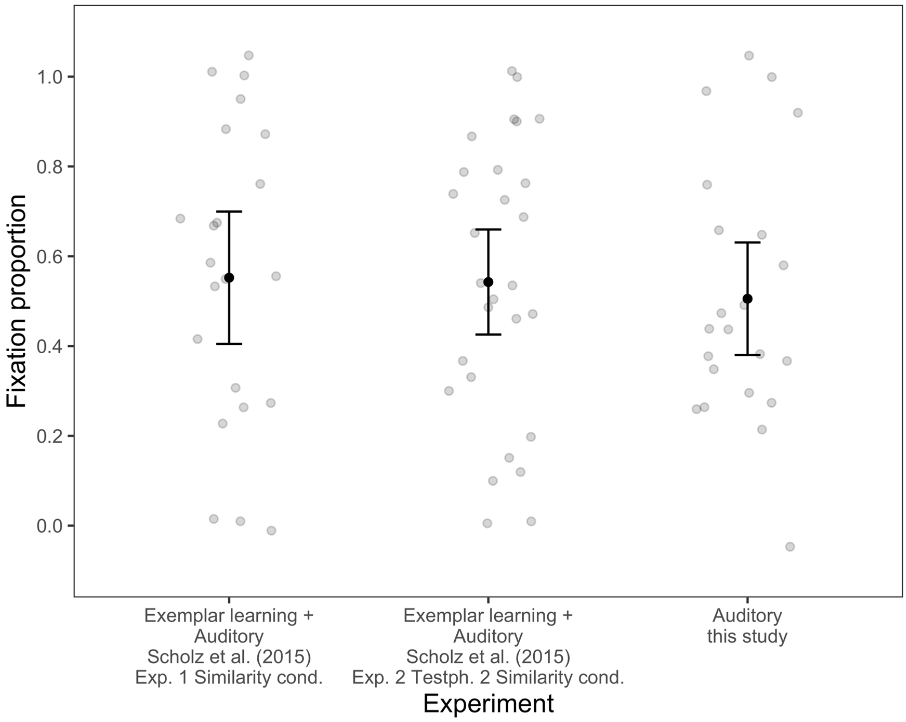


*Figure 6.* Mean proportions of fixation on relevant exemplars for training candidates in the test sets of two experiments reported in Scholz et al. (2015) and the Auditory condition of Experiment 1. Standard errors show between-subject 95% confidence intervals. Gray jittered dots in the background show the individual participants' mean.

[8] Note that due to some participants not showing LAN in some trials of Experiment 1 in Scholz et al. (2015), the number of participants is reduced to 22 for the presented analysis.

***LAN and Categorization Decisions***

Previous research has shown that LAN can bias judgments in the direction of the exemplar that was looked at most (Rosner & von Helversen, 2019). To test if such gaze biases also occurred for categorization decisions in this study, for trials in which LAN occurred, we ran a GLMM for categorization decisions (1 = invite, 0 = reject) with condition (all five presentation conditions) and proportions of fixation on the invited candidates as fixed effects as well as covariates' response times and location memory performance. Proportions of fixation on invited candidates indeed significantly predicted participants' decisions, $\chi^2(1) = 28.92$, $p < .001$. The more participants looked at invited exemplars, the larger the chance they would invite the test candidate. We could not find a significant difference between the conditions, $\chi^2(4) = 5.06$, $p = .28$. None of the covariates significantly predicted categorization decisions. Adding categorization accuracy instead of location memory performance revealed the same result pattern.

**Discussion**

Experiment 1 tested if a memory training on exemplar information leads to a higher LAN strength by increasing memory-driven activations on the shared priority map. Furthermore, we tested if LAN occurrence is reduced through an increase in activations on the shared priority map due to a visual presentation of verbal information about test candidates.

We found that the presentation modality strongly influenced the occurrence of LAN. LAN occurrence was strongest in the condition in which test items were presented auditorily, in line with previous findings on LAN in JDM (Rosner & von Helversen, 2019; Scholz et al., 2015). When verbal information about test candidates stayed visible on the screen until participants gave their response, no LAN was observed. Removing visual information from the screen led to increases in LAN. The shorter the presentation duration of the visual information on the screen, the more LAN was observed. We discuss these results in more detail in the General Discussion.

In conditions in which we had sufficient data to measure LAN strength, gaze behavior reflected the similarity between test items and exemplars. However, the results on whether memory strength increases LAN strength were inconclusive. Because in the condition with exemplar memory training not much LAN occurred, we compared the data from the Auditory condition with a previous data set that included exemplar memory training (Scholz et al., 2015). The analyses did not show any differences in LAN strength with and without exemplar memory training. We also did not find an effect of location memory on LAN strength. However, more correct classifications were related to more looks at the associated spatial locations, which could indicate an effect of better memories on LAN strength. One reason why we did not find much of an effect of memory training could be that memory representation may have been sufficiently strong to elicit LAN even without exemplar training. However, the results of the reanalysis need to be interpreted with caution given that they rely on a comparison of two different experiments and thus participants were not randomly assigned to conditions. This leaves open the question of whether the strength of memory-driven activations influences LAN. To investigate this question further and to understand better how LAN is related to memory processes when stimuli are presented visually, we conducted a second experiment. We used a visual color categorization task with two inseparable attribute dimensions, in which the use of exemplar memory has often been demonstrated (Nosofsky, 1988; Nosofsky & Palmeri, 1997; Nosofsky & Stanton, 2005). Moreover, categorization with these stimulus materials is likely to be more difficult, which allowed us to investigate the impact of memory accuracy when memory representations are less accurate. In addition, we employed an explicit imagery instruction and instructed participants to deliberately use the exemplar information for their decisions, to increase memory-driven activations. Imagining the most similar exemplar may increase memory activations, which should become visible in higher LAN strength.

## Experiment 2

Participants' task was to classify color stimuli into one of two categories. The stimuli all had the same hue but varied in their brightness and saturation (e.g., Nosofsky & Palmeri, 1997). In a criterion learning phase, participants learned which of six colors (training exemplars) belonged to which category via outcome feedback. During the test phase, they repeatedly classified the exemplar items and new items (the factor item type varied within participants). Half the participants received the instruction to use the knowledge gained during criterion learning (intuitive condition). Half were instructed to imagine the exemplars with their inner eye and to choose the category with exemplars of the highest similarity (explicit condition). Thus, task instructions varied between participants. They were tested on their memories about the exemplar locations before the test phase and at the end of the experiment. Additionally, at the end of the experiment, we tested their color memories in an old–new discrimination test. The experiment lasted on average 24.87 min (*SD* = 3.28). The study design and methods were approved by the ethics committee of the University of Zurich. All materials and data are available at https://osf.io/zpc2s/.

### Participants

To find a medium effect (e.g., $\eta_p^2 \approx .07$) of the instruction condition on LAN occurrence, a total sample of 60 participants was needed to reach a power of 80% (repeated measures ANOVA, between factors, 2 groups, number of measurements = 10 – corresponding to 10 test items, see Faul et al., 2007). Overall, 61 (15 male, $M_{age}$ = 30.5 years, range 18–68 years) participants at the University of Zurich took part in the study. Participants received course credit or financial compensation (8 CHF). All participants had normal or corrected-to-normal vision. Mean track accuracy in the test trials was very high at *M* = 0.8° of visual angle. All participants signed informed consent forms.

### Apparatus

The same eye tracker and setup was used as in Experiment 1.

**Materials**

The stimuli consisted of a set of 16 colors. Ten of them were identical to colors used by Nosofsky and Stanton (2005). Stimuli consisted of a constant red hue of the Munsell color scheme (7.5R). They varied in their brightness and saturation. Items were chosen based on a simulation with the generalized context model (GCM; Nosofsky, 2011, see Appendix B for details). Six colors formed the training set, with three belonging to Category A and three to Category B. To correctly classify the training items, both attribute values—saturation and brightness—had to be considered. The test set consisted of the six training exemplars and four ambiguous items that lay exactly in the middle of the two categories. Additionally, six colors were generated for the old–new discrimination test at the end of the experiment. Figure 7 shows a schematic illustration of the Munsell color configuration used in Experiment 2. Appendix C provides the red, green, and blue (RGB) values corresponding to each color.

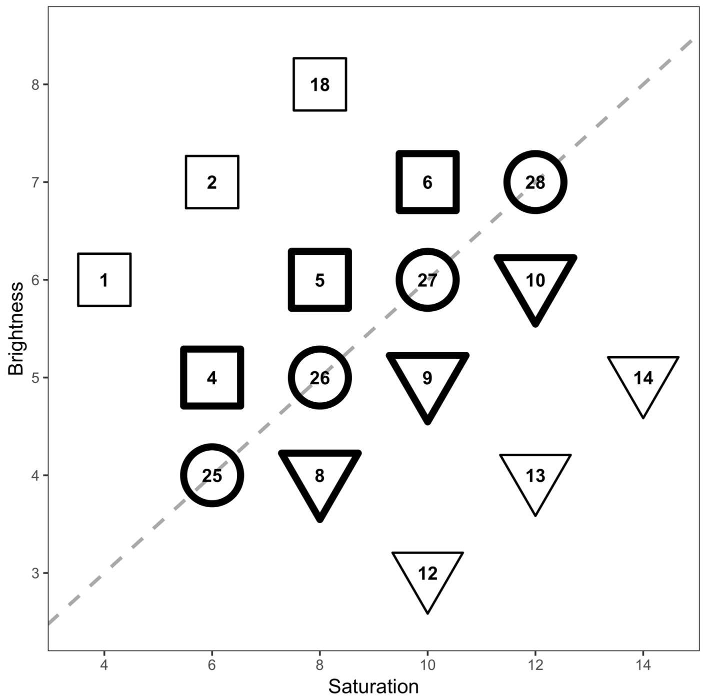


*Figure 7.* Schematic illustration of the Munsell color configuration used in Experiment 2. Squares represent training exemplars of one category and triangles exemplars of the other. Circles represent ambiguous items that lay on the category border (dashed gray line). Thick

frames mark items used in the test phase and the old–new discrimination test. Thin-framed items were used solely in the old–new discrimination test.

Exemplars were arranged in a circle with a size of 3.77° of visual angle (140 pixels) at a distance of 11.28° of visual angle (420 pixels) from the center of the screen (see Figure 8). Test items were of the same size as exemplars and presented in the center of the screen.

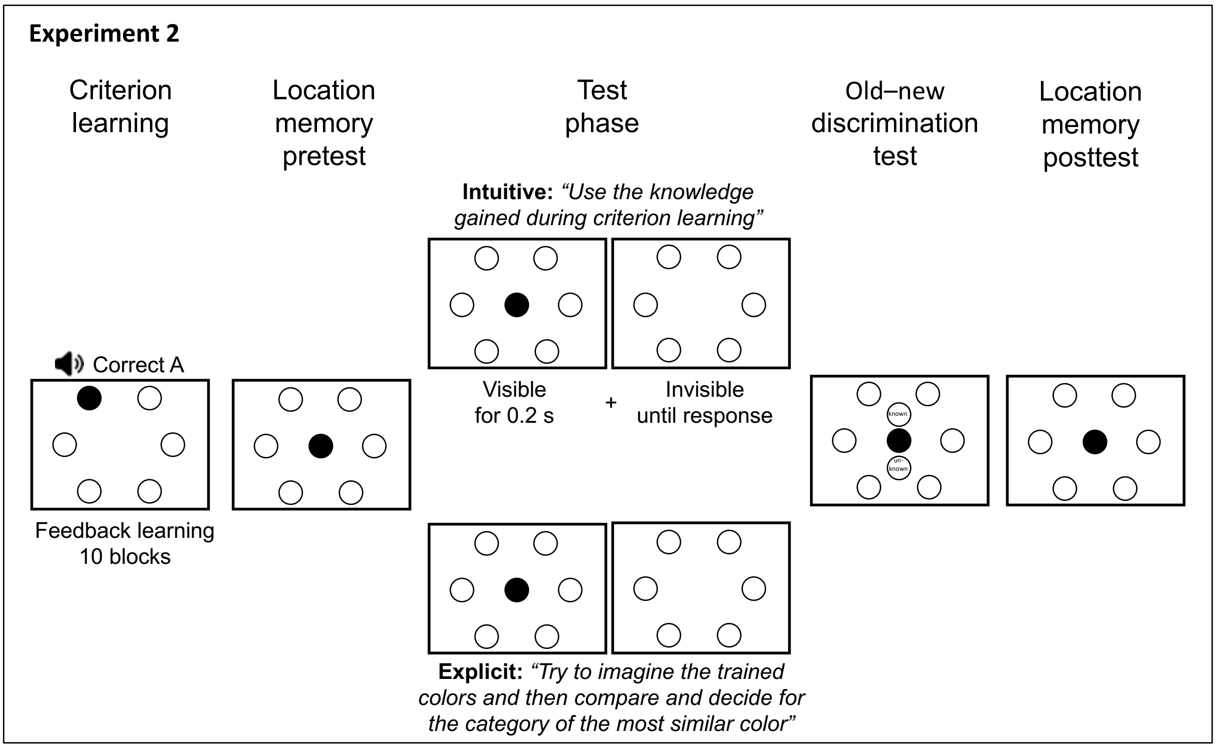


*Figure 8.* Experimental procedure. All participants worked through the criterion training phase, the test phase, and the location and old–new discrimination memory tests. Conditions differed in the instructions that participants received before and during the test phase. Black circles contained the exemplar or test item. The circles with text under old–new discrimination read known (top) and unknown (bottom).

## Procedure

At the beginning of the experiment, the eye tracker was calibrated to check if eye movements could be recorded to a sufficient quality (< 1.5° of visual angle).

### *Criterion Learning*

During criterion learning, participants were instructed to classify each of the six training exemplars into one of two categories (A or B). Therefore, they first saw one of the exemplars in one of the six circles (Figure 8). They were instructed to press, at their own pace, the left mouse button if they thought the exemplar belonged to Category A and the right mouse button for Category B. Auditory feedback on the correctness of decisions was given over headphones and the correct category label was repeated. After the feedback, the training exemplar disappeared and a new training exemplar was shown in a different spatial location. One block consisted of the random presentation of all six exemplars. Criterion learning lasted for 10 blocks. For each participant, the training exemplars always appeared in the same spatial positions. We controlled for exemplar location in 12 orders with the constraint that at most two exemplars of the same category were presented next to each other. Furthermore, category labels for exemplars were reversed for half the participants. For instance, for half the participants exemplar 4 (see Figure 7) belonged to Category A, and to Category B for the other half. Criterion learning lasted on average 8.98 min (*SD* = 0.96).

### *Location Memory Pretest*

Next, we assessed participants' memories of the trained exemplar locations. For this, one exemplar was presented in the center of the screen and the participants' task was to click on the spatial location that had contained the exemplar during criterion learning. Each exemplar location was tested once. Participants received no feedback on the correctness of their decisions. The location memory pretest lasted on average 0.96 min (*SD* = 0.15).

***Test Phase***

Each trial of the test phase began by participants being asked if they were ready to start the next trial. After participants pressed the left mouse key, a fixation cross appeared in the center of the screen for 1.5 s, followed by the test item that was visible for 200 ms. After that, the screen contained only the frames of the six training exemplars (Figure 8). Participants indicated their decisions by clicking on either the right or the left mouse button. One block in the test phase consisted of the presentation of all six training exemplars and the four ambiguous color items in randomized order. The test phase consisted of 10 blocks (i.e., 100 test trials in total).

The critical manipulation of Experiment 2 was introduced prior to the beginning of the test phase. Participants in the intuitive condition were instructed: "In order to make your decisions on the category membership of the test items, please use your previous knowledge about the trained colors." Participants in the explicit condition were instructed: "After seeing the color of the test item, imagine which training exemplar was the most similar. Try to remember the previously introduced colors and then compare and decide for the category of the most similar color." This instruction was repeated after one third and two thirds of the test phase for participants in the explicit condition. Participants in the intuitive condition were informed only about the progress of the test phase. Eye movements were recorded throughout the test phase. The eye tracker was recalibrated during two breaks, one after one third and one after two thirds of the trials. The test phase lasted on average 8.63 min (*SD* = 1.01).

***Old–New Discrimination Test***

The procedure was similar to that of the location memory pretest. However, this time, participants saw the six exemplar items intermixed with six completely new items in randomized order. Their task was to indicate if they had previously seen the tested color. Therefore, they clicked on either the "known" or the "unknown" button located at a distance of 5.38° of visual angle (200 pixels) above and below the center of the screen (Figure 8). Each item occurred only once. The old–new discrimination test lasted on average 1.32 min (*SD* = 0.21).

### *Location Memory Posttest*

The procedure was the same as for the location memory pretest. The test lasted on average 0.46 min (*SD* = 0.06).

## Results

The aim of Experiment 2 was to test if an explicit imagery instruction leads to higher LAN strength than a more intuitive exemplar instruction. Furthermore, we aimed to replicate the result of Experiment 1 in a visual categorization task, where exemplar-based categorization behavior has often been observed. Like in Experiment 1, we first analyzed participants' performance in the categorization task.

### *Preparatory Data Analyses and Rationale for the Analyses*

Mean response time in the test phase was 1.5 s (*SD* = 1.8). The response times of one participant exceeded the mean response times of all participants by more than 5 times the standard deviation of the response-time distribution. Therefore, this participant was excluded from further analyses. In addition, 25 trials (0.42% of trials) were excluded because response times were larger than 10 s (again, more than 5 times larger than the standard deviation). In 109 trials (1.82% of trials), no gaze data were recorded, and these trials could therefore not be analyzed. Using the same exclusion criteria as in Experiment 1 (exemplar classification accuracy at chance level or worse) would have led to the exclusion of 14 participants from the analyses of Experiment 2. However, given that the second experiment was overall more difficult (see results on categorization accuracy and memory test performance), such a strict performance criterion did not seem appropriate and thus we included these participants in the analyses. However, removing them leads to the same overall results pattern.

We analyzed gaze data of looks toward the exemplar locations for a circular area 3 times the size of the exemplar location (11.30° of visual angle, 420 pixel). The mixed-model analyses followed the same rationale as in Experiment 1. The analyses of LAN occurrence and LAN in relation to categorization behavior contained by-subject random intercepts. All other analyses included by-subject random intercepts and by-subject random slopes for exemplars. Further details and results of the analyses can be found in the online supplemental materials provided on OSF (https://osf.io/zpc2s/).

***Categorization and Location Memory Performance***

Participants in the two instruction conditions did not differ in task performance (see Table 2). To test for differences, we ran separate mixed models for each dependent measure including a fixed effect for instruction. Instruction conditions showed no meaningful differences in categorization accuracy, $\chi^2(1) = 0.48$, $p = .49$, categorization response time, $F(1, 58.03) = 0.37$, $p = .55$, location memory performance, $F(1, 58.01) = 0$, $p = .96$, or discrimination accuracy, $F(1, 57.98) = 2.88$, $p = .10$. Note, for the analyses of location memory performance, we summarized location memory pre- and posttest performance into one location memory score. The score indicates the number of correct localizations (2 = an exemplar was correctly localized in the pre- and posttest, 1 = correctly localized in either test, 0 = never localized correctly). Thus, the higher the score, the better participants remembered the exemplar locations ($M_{\text{Int}} = 0.7$, $SD_{\text{Int}} = 0.48$, $M_{\text{Exp}} = 0.7$, $SD_{\text{Exp}} = 0.31$, where int = intuitive and exp = explicit).

Table 2

*Means (and Standard Deviations) of Performance and Looking-at-Nothing (LAN) Behavior in Experiment 2 in Test Trials and Memory Tests*

| Variable | Condition | |
|---|---|---|
| | Intuitive ($N$ = 30) | Explicit ($N$ = 30) |
| Test trials | | |
| Categorization accuracy | 65 (19) | 67 (17) |
| Categorization response time (s) | 1.30 (0.59) | 1.50 (0.82) |
| LAN occurrence | 0.22 (0.31) | 0.23 (0.28) |
| LAN occurrence ($N$) | 28 | 30 |
| LAN strength | 0.21 (0.17) | 0.27 (0.25) |
| Memory tests | | |
| Location accuracy pretest | 38 (25) | 33 (32) |
| Location accuracy posttest | 33 (19) | 38 (20) |
| Discrimination accuracy | 69 (13) | 67 (18) |

*Note.* Mean percentage correct responses was calculated for exemplar items leaving out ambiguous items. Mean response times and LAN measures are based on all items. LAN occurrence describes mean proportion of LAN trials. The number of participants showing LAN (LAN occurrence *N*) indicates the number of participants who looked at least once in one experimental trial at one of the exemplar locations. LAN strength describes the participants' mean proportions of fixation on the relevant exemplar for trials in which LAN occurred and random selection of one area of interest for ambiguous items; chance level is .17. Location and discrimination accuracy describe mean percentage correct in the memory tests.

### *Covariate Analyses for Accuracy*

In addition to the type of instruction, we added response time, location memory performance, discrimination accuracy, and block[9] as covariates to the analyses of categorization accuracy. Like in the analysis of categorization accuracy without covariates, there was no effect of instruction, $\chi^2(1) = 0.55$, $p = .46$. Concerning covariates, this analysis revealed significant influences of response time, $\chi^2(1) = 16.15$, $p < .001$, and location memory, $\chi^2(1) = 27.25$, $p < .001$. That is, in line with the results in Experiment 1, the longer participants took to decide, the worse their decisions (linear contrast: $z = -4.03$, $p < .001$), and

[9] As test items were repeated in blocks of trials, we controlled for this by adding block as a covariate.

the better their location memory, the better their decisions (linear contrast: $z = 5.22$, $p < .001$). We added categorization accuracy, response times, location memory performance, discrimination accuracy, and block as covariates to the subsequent analyses of LAN, which improves model fits for analyses of LAN occurrence and strength, but not LAN and categorization decisions. However, our main results stay the same when running models without covariates.

### *LAN Occurrence*

LAN occurrence was calculated in the same way as in Experiment 1, that is, as the proportion of trials in which participants looked at least once at one of the exemplar locations. As the presentation modality was the same in both instruction conditions, LAN occurrence should be comparably strong between the conditions. Indeed, LAN occurrence in Experiment 2 was comparably high to that in Experiment 1B in which visually presented verbal information was also removed from the screen (see Tables 1 and 2). A mixed-model analysis revealed no difference between the instruction conditions, $\chi^2(1) = 0.11$, $p = .74$ (Figure 9).

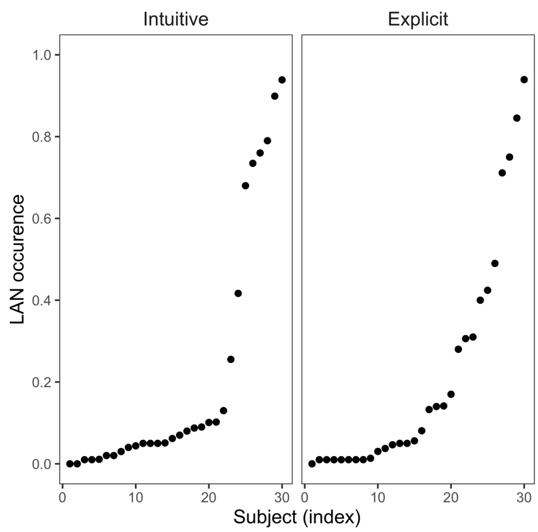


*Figure 9.* Proportion of trials per participant in which looking-at-nothing (LAN) occurred in the two instruction conditions of Experiment 2. Each point shows the data of one participant.

***LAN Strength***

The mixed model analysis of LAN strength, including fixed effects for instruction condition and item type, as well as the covariates categorization accuracy, response time, and block, found a main effect of instruction. Participants fixated proportionally more on the most similar exemplar location when they received an explicit instruction in comparison to an intuitive instruction, $\chi^2(1) = 15.52$, $p < .001$ (see Figure 10). There was a main effect of item type, with the effect being more pronounced for ambiguous items than for exemplar items, $\chi^2(2) = 10.27$, $p < .001$. The interaction of instruction and item type did not reach significance, $\chi^2(1) = 3.25$, $p = .07$. However, the main effect of item type appears not to be very stable. When we added the interaction of response time and item type to the model, the main effect of item type was no longer significant, $\chi^2(1) = 3.09$, $p = .08$. LAN strength was more pronounced for correct classifications, main effect of categorization accuracy: $\chi^2(2) = 133.67$, $p < .001$, and when decisions were made more quickly, main effect of categorization response time: $\chi^2(1) = 6.72$, $p = .01$ (linear contrast response time: $z = -2.60$, $p = .009$).

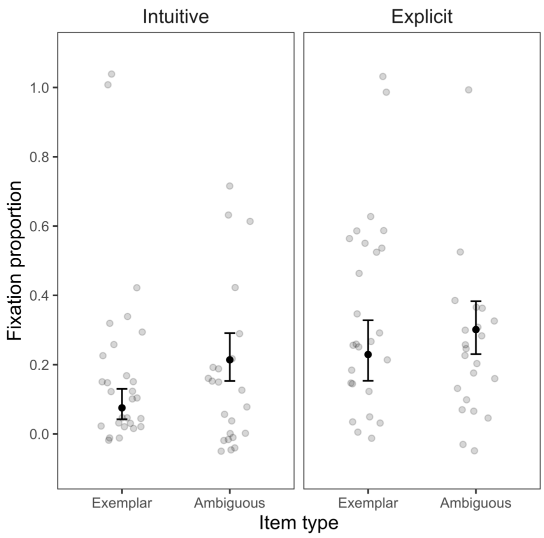


*Figure 10.* Estimated mean proportions of fixation on the relevant exemplar for exemplar and ambiguous items in the two instruction conditions of Experiment 2. Standard errors show

estimated within-subject 95% confidence intervals. Gray jittered dots in the background show the individual participants' means.

Analyzing only the exemplar items allowed us to assess the influence of location memory performance and discrimination accuracy. A mixed-model analysis excluding the factor item type and including location memory performance as well as discrimination accuracy revealed a significant influence of location memory performance, $\chi^2(1) = 45.97$, $p < .001$, but no effect of discrimination accuracy, $\chi^2(1) = 0.55$, $p = .46$. That is, the better participants remembered the exemplar locations, the higher the LAN strength was (linear contrast location memory: $z = 6.77$, $p < .001$). All other results stayed the same.

### *LAN and Categorization Decisions*

Like in Experiment 1, we tested if participants were more likely to choose the category they looked at most. We ran a mixed model for categorization decisions (A = 1, B = 0) with instruction, item type, and their interaction, as well as proportion of fixation on Category A as fixed effects. As covariates, we added categorization accuracy, response time, and block. We found a main effect of fixation proportion. The more participants looked at Category A, the more likely they were to respond with A, $\chi^2(1) = 268.55$, $p < .001$, replicating the results of Experiment 1.

## Discussion

The goal of the second experiment was to test if an explicit instruction to imagine relevant information from memory would lead to higher LAN strength than when participants' instruction for how to retrieve exemplar information was more implicit (intuitive condition). As expected, people showed more LAN to locations associated with stored information when they received an explicit imagery instruction.

As in Experiment 1, long response times increased the chance of a participant hitting one of the exemplar locations by looking around the screen, but they decreased the proportion of fixations on the most similar exemplar locations. With long response times, it is possible that several exemplars are being retrieved from memory in order to come up with a classification decision (Nosofsky & Palmeri, 1997). On the one hand, this can increase the chances of looking at at least one of the exemplar locations, increasing LAN occurrences. On the other hand, if different exemplars are activated in memory, this reduces the chances of looking at the most similar exemplar location, which leads to reduced LAN strength. In line with this, as in the first study, longer response times went along with worse classification decisions. The more correctly participants decided, and the better their memories for the exemplar locations, the higher the chances of looking at the most similar exemplar locations. In sum, depending on the assumed retrieval process and the way LAN was measured, LAN behavior can be either increased or decreased with long response times.

## General Discussion

Whereas there is strong agreement among researchers that memory plays a major role in the decision-making process (e.g., Weber et al., 1995), the exact nature of the influence is not well understood. A promising path for investigating memory processes in JDM is to study eye movements (e.g., Orquin & Mueller-Loose, 2013) in general, and the LAN behavior more specifically (e.g., Renkewitz & Jahn, 2012). This study focused on the interaction of memory-driven and stimulus-driven influences on LAN in two typical JDM paradigms, where people often use a memory-based decision strategy (i.e., the retrieval of past instances stored as exemplars in memory; e.g., Nosofsky, 2011). Our assumptions on eye-movement behavior were derived from the idea of a shared priority map (e.g., Hedge et al., 2015), wherein activations resulting both from memory retrieval and from processing information from the visual environment influence eye movements, as they both create activations in one spatially organized representation in memory. The location in the representation that receives the highest neural activation determines the location of the next fixation in the visual world (Figure 1).

Concerning memory-driven influences on LAN, we observed that LAN reflected the similarity between test items and exemplars stored in memory. Furthermore, the more people looked at the exemplar locations of one category, the more likely they were to categorize an item in the same category, suggesting that they indeed used similarity-based exemplar memory to make their decisions. In regard to increasing memory-driven activations by strengthening memory traces, our results were less clear. In Experiment 1, we found no evidence that first learning exemplar information by heart led to observable difference in LAN strength in comparison to 10 rounds of criterion learning. However, in both experiments, we found evidence that categorization accuracy was related to more LAN strength. Furthermore, an explicit instruction to imagine the most similar exemplar from memory increased LAN strength.

Concerning the influence of presentation modality, we found that presenting test items visually in comparison to auditorily increased eye movements to visual information presented on the screen and reduced LAN. However, when visually presented information was removed after people perceived it, LAN occurred and reflected memory-driven activation. In the following we discuss the different influences investigated in more detail and then consider the implications of our results for using LAN to understand memory processes in decision making.

**Effects of Memory Training on LAN Behavior**

In Experiment 1 we did not find differences in LAN strength when participants had only a criterion learning phase or an additional exemplar learning phase in which they learned the attributes of each exemplar by heart. However, the more correct participants' classifications were, the more they looked at the associated screen locations. Furthermore, the results of Experiment 2 with a more difficult categorization task (due to less distinct attributes) showed that the better each individual remembered the locations at which exemplars were presented, the more likely they were to look at the associated screen locations. First, these results suggest that exhaustive training is not needed to observe LAN during JDM. This is in line with previous LAN studies that used encoding-retrieval paradigms without extensive learning phases (Johansson et al., 2012; Krefeld-Schwalb & Rosner, 2020; Kumcu & Thompson, 2020; Scholz et al., 2016). Second, the results suggest that if variance is large enough, the quality of the memory representation can matter. Thus, overall, the results are in line with the idea that the more strongly information and its location information is activated in memory, the more likely it becomes that the exemplar will be retrieved (e.g., Dougherty et al., 1999) and the higher the quality of LAN. This is in line with the assumptions in the shared priority framework that activated knowledge increases attention in memory, which in turn increases the likelihood of looking at spatial locations associated with retrieved information (Hedge & Leonards, 2013; Hedge et al., 2015; Theeuwes et al., 2009). But note, materials in this study consisted of rather difficult exemplars. If exemplars would be over-learned or easy, this may result in less LAN, possibly due to rather automatic, short responses (Jones et al., 2018; Scholz et al., 2011; Wantz et al., 2016). Future research is needed to describe under what conditions LAN may be strongest and thereby most informative about exemplar retrieval from memory.

**Effects of Task Instruction on LAN Behavior**

In Experiment 2, the explicit instruction to imagine the most similar exemplar affected the quality of the LAN behavior. People showed more LAN to locations associated with the retrieved information in the explicit imagery condition. This is in line with research demonstrating that eye movements can be guided by manipulating the viewing strategy (e.g., Brandstatt & Voss, 2014; Chan, Kamino, Binns, & Ryan, 2011; Foulsham & Kingstone, 2013) or as in this study by task instructions (Shih et al., 2012). The results of this study can be explained with the shared priority map framework by the imagery instruction increasing the activation of the most similar exemplar held in memory (Dougherty et al., 1999), because imagining exemplars made them more vivid (Kosslyn, 1994), which led to a higher accessibility of exemplar information and increased chances of reactivating location information stored along with the exemplars in memory.

One potential alternative explanation is that we found the differences between the instruction conditions because in the intuitive condition in Experiment 2, participants did not rely on an exemplar-based process. Yet, often exemplar-based processes are assumed to be automatic and other research has found exemplar-based processing with highly similar stimulus materials (Nosofsky & Palmeri, 1997; Nosofsky & Stanton, 2005). Furthermore, also in the intuitive condition, eye movements were more likely to be directed at the most similar exemplar, suggesting that in both conditions, LAN reflected exemplar retrieval from memory.

**Effects of Presentation Format on LAN Behavior**

The presentation modality strongly influenced LAN behavior. LAN was most observable when test items were presented auditorily, as in previous research on LAN during exemplar-based JDM (Rosner & von Helversen, 2019; Scholz et al., 2015). When verbal information about test candidates was presented visibly and the information stayed visible on the screen until participants gave their response, almost no LAN behavior occurred, although behavioral measures indicated that the same decision-making process took place. We explain this finding in light of the framework of the shared priority map. Visually presented information about the job candidates that should be judged created an activation peak in the center region of the spatially organized representation held in memory, corresponding to the visuospatial location at which information was presented on the screen. This activation was stronger than the activations resulting from exemplar retrieval. Thus, although people likely retrieved exemplars from memory, eye movements were mainly guided by the activations resulting from the visually salient and relevant stimuli and the effect of exemplar similarity on LAN was no longer observed. A similar argument was used by Jones et al. (2018). They assumed that an activation threshold in memory has to be exceeded to show LAN behavior. If the activation of retrieval-relevant information is sufficiently strong (e.g., when information is highly activated in memory; see Scholz et al., 2011) but location memory—that is, knowledge about where information was presented—is too weak, LAN will not be shown.

In the Auditory condition of Experiment 1, almost no visual information was presented (as is typical for the blank screen paradigm; Altmann, 2004). Thus, eye movements reflected activations of exemplars from memory. In addition, we assume that this effect was further increased through the sequential presentation of auditory stimulus materials, reflecting an effect of processing spoken information on LAN (Altmann & Kamide, 2009; Huettig et al., 2011). That is, in this case, LAN may reflect both auditory language processing and decision making. Language processing, however, might not explain the observed LAN behavior completely, as it does not occur when people draw less on memory retrieval to make their judgments or decisions, for instance, when using a rule-based strategy (see Scholz et al., 2015).

The effect of language processing on LAN may be reduced when verbal information about job candidates is presented visually, but is removed as soon as participants have read the information. In Experiment 1B, we show that removing visually presented verbal information indeed increased LAN occurrence compared to a condition in which the visual information remained visible. Indeed, the shorter the presentation duration of the visual information on the screen, the more LAN was observed. In Experiment 2, with a similar procedure, there were a reasonable number of trials in which LAN was observed. In light of the framework of the shared priority map, this finding can be explained as resulting from the removal of the strong activations in the center of the map. As soon as information is removed from the screen, eye movements seem to reflect memory activations resulting from exemplar retrieval.

**Top-Down and Bottom-Up Processing**

Throughout this study, we assumed that increased memory activations (due to similarity, task instructions, or training) lead to increases in the strength of LAN and that activations resulting from visual information presentation lead to decreases in the occurrence of LAN. An interesting question is to what extent these different sources of activation map to the widely used dichotomy of bottom-up and top-down processing that is prevalent in both the attention (Awh et al., 2012; Theeuwes, 2010) and decision-making (Orquin et al., 2013; Orquin & Mueller Loose, 2013) literature. Certainly, a one-to-one mapping of memory-driven activations stemming from top-down or goal-directed processes and stimulus-driven activations stemming from the visually presented pieces of information is not possible.

Indeed, activations resulting from visual stimulus presentation are likely to stem from both types of processing. On the one hand, the visual presentation of items should increase activations in the shared priority map in a bottom-up manner through its physical salience. On the other hand, the verbal information presented on screen was task relevant and thus should also increase attention to it in a top-down manner in order to enable participants to accomplish the decision task.

The questions of whether and to what extent memory-driven activations in a shared priority map are top down is complex. Under explicit task instructions, the use of similarity to the previously seen exemplars and the resulting memory-driven activations through the exemplars are likely goal directed, suggesting top-down processing. However, for instance in the implicit instruction condition, memory-driven activation may be automatic. Thus, while activations due to exemplar retrieval in this condition may reflect top-down processing it is also possible that they are a by-product of the goal to come up with a classification decision. In the latter case, the training phase could have biased participants’ attention to spatial locations that had contained the exemplars during the initial training phase. Such an effect of the selection history (Awh et al., 2012) could have caused automatic memory-related activations in the shared priority map, comparable to effects of intertrial priming observed in visual search tasks (Kristjansson & Campana, 2010).

More research is clearly necessary to clarify the nature of the assumed activations. But, LAN and LAN in JDM tasks may provide an interesting test bed to tease apart different sources of activations and how they guide attention and eye movements.

**LAN as a Process Measure in JDM**

Our secondary goal with this study was to investigate the extent to which LAN behavior is informative about retrieval processes in memory occurring during JDM, that is, the extent to which LAN can be used as a process measure in JDM (e.g., Schulte-Mecklenbeck et al., 2017). In 22% to 37% of trials in which LAN was observed, the behavior systematically reflected the similarity of exemplars stored in memory to test items presented on the screen. This was also the case when participants were repeatedly presented with exemplar items while receiving outcome feedback on the criterion values, instead of memorizing the exemplar information.

Furthermore, using LAN revealed an exemplar-gaze cascade effect. That is, when LAN occurred in this research, it predicted participants' decisions: The more participants looked at exemplars of one category, the more likely they were to choose that category. This is in line with one of the most robust findings on eye movements in JDM, which can be observed when all decision-relevant information stays visible on screen: the gaze cascade effect (Glaholt & Reingold, 2009; Shimojo, Simion, Shimojo, & Scheier, 2003).

This suggests that even when LAN occurrence is reduced with visual presentations, studying LAN is a promising way to better understand memory processes in JDM. An important constraint, however, is that in particular with visual presentation formats, conclusions can only be drawn reliably on the aggregate level, because inferences about the behavior of single participants comes at the risk that a participant does not show LAN in a sufficient number of trials and thus eye movements are not informative about the assumed retrieval processes.

To get the most out of LAN for studying memory processes in JDM, we recommend that when visually presenting new decision options, they should be removed from the screen as soon as participants have encoded the attribute information. This can be achieved by being informed about expected reading or viewing times or by individual adjustments of the presentation durations (García-Pérez, 1998). Furthermore, providing outcome feedback during criterion learning is sufficient to elicit exemplar retrieval and LAN. However, individual learning performance should be assessed and controlled for. As outcomes of the decision-making process including categorization decisions were comparable between conditions, a more explicit instruction on imagining past experiences during decision making could thus increase LAN in cases where there is good reason to assume that the instruction will not alter the decision process.

## Conclusions

Process measures such as eye tracking provide valuable insights into JDM, even if relevant information must be retrieved from memory. By applying recent findings on how different sources of activation influence eye movements to the domain of JDM, we gained deep insights into when eye movements inform us about retrieval processes from memory and when they are mainly driven by features of the stimulus environment. The results highlight the potential of using eye movements based on the LAN behavior to study memory-based processes in JDM.

## Context paragraph

Agnes Rosner's research focuses on the looking-at-nothing (LAN) behavior. Together with Bettina von Helversen, she uses it to study how memory about past instances influences current judgments and decisions. Both authors have found that to continue with this line of research, a better theoretical understanding is needed about the boundary conditions under which LAN behavior reflects memory processes in judgment and decision making. This was the starting point of the experimental series presented in this manuscript. Michael Schaffner later joined the author team by conducting the second experiment and contributing to the theoretical understanding of LAN during judgment and decision making.

**References**

Albrecht, R., Hoffmann, J. A., Pleskac, T. J., Rieskamp, J., & von Helversen, B. (2020). Competitive retrieval strategy causes multimodal response distributions in multiple-cue judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(6), 1064–1090. https://doi.org/10.1037/xlm0000772

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition*, *93*, B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71. https://doi.org/10.1016/j.cognition.2008.12.005

Awh, E., Belopolsky, A., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bone, M. B., St-Laurent, M., Dang, C., McQuiggan, D. A., Ryan, J. D., & Buchsbaum, B. R. (2019). Eye movement reinstatement and neural reactivation during mental imagery. *Cerebral Cortex*, *29*(3), 1075–1089. https://doi.org/10.1093/cercor/bhy014

Brandstatt, K. L., & Voss, J. L. (2014). Age-related impairments in active learning and strategic visual exploration. *Frontiers in Aging Neuroscience, 6*(2), 19. https://doi.org/10.3389/fnagi.2014.00019

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38. https://doi.org/10.1162/jocn.1997.9.1.27

Bröder, A., Newell, B. R., & Platzer, C. (2010). Cue integration vs. exemplar-based reasoning in multi-attribute decisions from memory: A matter of cue representation. *Judgment and Decision Making*, *5*(5), 326–338.

Chan, J. P. K., Kamino, D., Binns, M. a, & Ryan, J. D. (2011). Can changes in eye movement scanning alter the age-related deficit in recognition memory? *Frontiers in Psychology, 2*(5), 1-11. https://doi.org/10.3389/fpsyg.2011.00092

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

de Groot, F., Huettig, F., & Olivers, C. N. L. (2016). Revisiting the looking at nothing phenomenon: Visual and semantic biases in memory search. *Visual Cognition*, *24*(3), 226–245. https://doi.org/10.1080/13506285.2016.1221013

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. https://doi.org/10.1037/0033-295X.106.1.180

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140. https://doi.org/10.1037/0096-3445.127.2.107

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General*, *115*(2), 155–174. https://doi.org/10.1037/0096-3445.115.2.155

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior*

*Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Foerster, R. M. (2018). "Looking-at-nothing" during sequential sensorimotor actions: Long-term memory-based eye scanning of remembered target locations. *Vision Research*, *144*, 29–37. https://doi.org/10.1016/j.visres.2018.01.005

Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: an experimental test of scanpath theory. *Journal of Experimental Psychology. General, 142*(1), 41-56. https://doi.org/10.1037/a0028227

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–1881. https://doi.org/10.1016/S0042-6989(97)00340-4

Glaholt, M. G., & Reingold, E. M. (2009). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*(8), 1228–1243. https://doi.org/10.1080/13506280802362962

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*(2–3), 197–230. https://doi.org/10.1016/S0010-0277(97)00044-9

Hedge, C., & Leonards, U. (2013). Using eye movements to explore switch costs in working memory. *Journal of Vision*, *13*(4), 1–19. https://doi.org/10.1167/13.4.18

Hedge, C., Oberauer, K., & Leonards, U. (2015). Selection in spatial working memory is independent of perceptual selective attention, but they interact in a shared spatial priority map. *Attention, Perception, and Psychophysics*, *77*(8), 2653–2668. https://doi.org/10.3758/s13414-015-0976-4

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences, 21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior*

*Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological Science*, *24*(6), 869–879. https://doi.org/10.1177/0956797612463581

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*(6), 2242–2261. https://doi.org/10.1037/a0037989

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2016). Similar task features shape judgment and categorization processes. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1193–1217. https://doi.org/10.1037/xlm0000241

Hoffmann, J. A., von Helversen, B., Weilbächer, R. A., & Rieskamp, J. (2018). Tracing the path of forgetting in rule abstraction and exemplar retrieval. *Quarterly Journal of Experimental Psychology*, *71*(11), 2261–2281. https://doi.org/10.1177/1747021817739861

Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition*, *108*(2), 533–542. https://doi.org/10.1016/j.cognition.2008.02.011

Huettig, F., Mishra, R. K., & Olivers, C. N. L. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00394

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*.

https://doi.org/10.1016/j.actpsy.2010.07.013

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Jahn, G., & Braatz, J. (2014). Memory indexing of sequential symptom processing in diagnostic reasoning. *Cognitive Psychology*, *68*, 59–97. https://doi.org/10.1016/j.cogpsych.2013.11.002

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*(4), 482–553. https://doi.org/10.1016/S0010-0285(02)00505-4

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1289–1314. https://doi.org/10.1037/a0026585

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079. https://doi.org/10.1207/s15516709cog0000_86

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*(1), 236–242. https://doi.org/10.1177/0956797613498260

Johansson, R., & Johansson, M. (2020). Gaze position regulates memory accessibility during competitive memory retrieval. *Cognition*, *197*. Advance online publication. https://doi.org/10.1016/j.cognition.2019.104169

Johansson, R., Oren, F., & Holmqvist, K. (2018). Gaze patterns reveal how situation models and text representations contribute to episodic text memory. *Cognition*, *175*, 53–68. https://doi.org/10.1016/j.cognition.2018.02.016

Jones, M. W., Kuipers, J. R., Nugent, S., Miley, A., & Oppenheim, G. (2018). Episodic traces and statistical regularities: Paired associate learning in typical and dyslexic readers. *Cognition*, *177*, 214–225. https://doi.org/10.1016/j.cognition.2018.04.010

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(5), 924–941. https://doi.org/10.1037/0278-7393.29.5.924

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298. https://doi.org/10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A. C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156. https://doi.org/10.1037/0096-3445.132.1.133

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making*, *3*(3), 244–260.

Kinjo, H., Fooken, J., & Spering, M. (2020). Do eye movements enhance visual memory retrieval? *Vision Research*, *176*, 80–90. https://doi.org/10.1016/j.visres.2020.07.013

Klichowicz, A., Strehlau, S., Baumann, M. R. K., Krems, J. F., & Rosner, A. (2020). Tracing current explanations in memory: A process analysis based on eye-tracking. *Quarterly Journal of Experimental Psychology, 73,* 1703–1717. https://doi.org/10.1177/1747021820922509

Kosslyn, S. M. (1994). *Image and brain*. MIT Press. https://doi.org/10.7551/mitpress/3653.001.0001

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483. https://doi.org/10.1016/j.visres.2010.12.014

Krefeld-Schwalb, A., & Rosner, A. (2020). A new way to guide consumer’s choice: Retro-cueing alters the availability of product information in memory. *Journal of Business*

*Research*, *111*, 135–147. https://doi.org/10.1016/j.jbusres.2019.08.012

Kristjansson, A., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics, 72,* 5–18.

Kumcu, A., & Thompson, R. L. (2020). Less imageable words lead to more looks to blank locations during memory retrieval. *Psychological Research*, *84*, 667–684. https://doi.org/10.1007/s00426-018-1084-6

Laeng, B., Bloem, I. M., D'Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, *131*(2), 263–283. https://doi.org/10.1016/j.cognition.2014.01.003

Laeng, B., & Teodorescu, D. S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, *26*(2), 207–231. https://doi.org/10.1016/S0364-0213(01)00065-9

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.5 [Computer software]. https://CRAN.R-project.org/package=emmeans

Martarelli, C. S., & Mast, F. W. (2011). Preschool children's eye-movements during pictorial recall. *British Journal of Developmental Psychology*, *29*(3), 425–436. https://doi.org/10.1348/026151010X495844

Martarelli, C. S., & Mast, F. W. (2013). Eye movements during long-term pictorial recall. *Psychological Research*, *77*(3), 303–309. https://doi.org/10.1007/s00426-012-0439-7

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

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli:

A choice model analysis. *Perception & Psychophysics, 38*(5), 415–432.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/cbo9780511921322.002

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review*, *7*(3), 375–402.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Stanton, R. D. (2005). Speeded classification in a probabilistic category structure: Contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 608–629. https://doi.org/10.1037/0096-1523.31.3.608

Olsson, A. C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear multiple-cue judgment task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(6), 1371–1384. https://doi.org/10.1037/0278-7393.32.6.1371

Orquin, J. L., Bagger, M. P., & Mueller Loose, S. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*(6), 700–716.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*, 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Pärnamets, P., Johansson, R., Gidlöf, K., & Wallin, A. (2016). How information availability interacts with visual attention during judgment and decision tasks. *Journal of Behavioral Decision Making*, *29*(2–3), 218–231. https://doi.org/10.1002/bdm.1902

Persson, M., & Rieskamp, J. (2009). Inferences from memory: Strategy- and exemplar-based judgment models compared. *Acta Psychologica*, *130*(1), 25–37. https://doi.org/10.1016/j.actpsy.2008.09.010

Platzer, C., Bröder, A., & Heck, D. W. (2014). Deciding with the eye: How the visually manipulated accessibility of information in memory influences decision behavior. *Memory and Cognition*, *42*(4), 595–608. https://doi.org/10.3758/s13421-013-0380-z

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences*, *28*(1), 1–14. https://doi.org/10.1017/S0140525X05000014

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001

Renkewitz, F., & Jahn, G. (2012). Memory indexing: A novel method for tracing memory processes in complex cognitive tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(6), 1622–1639. https://doi.org/10.1037/a0028073

Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing: A response to Ferreira et al.: Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *13*(6), 235–236. https://doi.org/10.1016/j.tics.2009.02.006

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*(1), 46–62. https://doi.org/10.1037/0096-

3445.133.1.46

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207–236. https://doi.org/10.1037/0096-3445.135.2.207

Rosner, A., & von Helversen, B. (2019). Memory shapes judgments: Tracing how memory biases judgments by inducing the retrieval of exemplars. *Cognition*, *190*, 165–169. https://doi.org/10.1016/j.cognition.2019.05.004

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, *133*(1), 63–82. https://doi.org/10.1037/0096-3445.133.1.63

Scholz, A., Klichowicz, A., & Krems, J. F. (2018). Covert shifts of attention can account for the functional role of "eye movements to nothing." *Memory and Cognition*, *46*(2), 230–243. https://doi.org/10.3758/s13421-017-0760-x

Scholz, A., Krems, J. F., & Jahn, G. (2017). Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning. *Psychonomic Bulletin and Review*, *24*(5), 1398–1412. https://doi.org/10.3758/s13423-017-1294-8

Scholz, A., Mehlhorn, K., Bocklisch, F., & Krems, J. F. (2011). Looking at nothing diminishes with practice. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1070–1075). Cognitive Science Society.

Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological Research*, *80*(1), 149–158. https://doi.org/10.1007/s00426-014-0639-4

Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition*, *136*, 228–246. https://doi.org/10.1016/j.cognition.2014.11.019

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, *26*, 442–450. https://doi.org/10.1177/0963721417708229

Shih, S.-I., Meadmore, K. L., & Liversedge, S. P. (2012). Aging, eye movements, and object-location memory. *PloS One, 7*(3), 1-7. https://doi.org/10.1371/journal.pone.0033485

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). afex: Analysis of factorial experiments. R package version 0.26-0 [Computer software]. https://CRAN.R-project.org/package=afex

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*(4), 235–241. https://doi.org/10.1007/s004260100059

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, *132*(2), 106–114. https://doi.org/10.1016/j.actpsy.2009.01.005

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity information provide benefits in inference problems? The role of strategy selection and knowledge of cue importance. *Acta Psychologica*, *144*(1), 73–82.

https://doi.org/10.1016/j.actpsy.2013.05.007

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*(1), 73–96. https://doi.org/10.1037/0096-3445.137.1.73

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(4), 867–889. https://doi.org/10.1037/a0015501

Wantz, A. L., Martarelli, C. S., & Mast, F. W. (2016). When looking back to nothing goes back to nothing. *Cognitive Processing*, *17*(1), 105–114. https://doi.org/10.1007/s10339-015-0741-6

Weber, E. U., Goldstein, W. M., & Barlas, S. (1995). And let us not forget memory: The role of memory processes and techniques in the study of judgment and choice. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation* (*Advances in research and theory*, vol. 32, pp. 33–81). Academic Press. https://doi.org/10.1016/S0079-7421(08)60307-2

Wynn, J. S., Ryan, J. D., & Buchsbaum, B. R. (2020). Eye movements support behavioral pattern completion. *PNAS, 53*(9), 1689-1699. https://doi.org/https://doi.org/10.1073/pnas.1917586117

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, *3*(21), 1–19. https://doi.org/10.3390/vision3020021

## Appendix A

### Item Sets Used in Experiment 1A and B

*Table A1*

| Set | Attribute 1: Language skills | Attribute 2: Computer skills | Attribute 3: Work experience | Attribute 4: Professional training | Similarity | Decision |
|---|---|---|---|---|---|---|
| *1* | *French* | *LaTeX* | *Building industry* | *Web design* | *4* | *Invite* |
| *2* | *Mandarin* | *GIMP* | *Finance* | *Sales* | *4* | *Invite* |
| *3* | *Russian* | *HTML* | *Automotive industry* | *Human resources* | *4* | *Reject* |
| *4* | *Spanish* | *SQL* | *Food service* | *Marketing* | *4* | *Reject* |
| 5 | Spanish | HTML | Automotive industry | Human resources | 3 | Reject |
| 6 | Russian | SQL | Food service | Marketing | 3 | Reject |
| 7 | French | HTML | Automotive industry | Human resources | 3 | Reject |
| 8 | Russian | LaTeX | Automotive industry | Human resources | 3 | Reject |
| 9 | Russian | HTML | Finance | Human resources | 3 | Reject |
| 10 | Russian | HTML | Automotive industry | Web design | 3 | Reject |
| 11 | Spanish | SQL | Food service | Sales | 3 | Reject |
| 12 | Spanish | GIMP | Food service | Marketing | 3 | Reject |
| 13 | Mandarin | SQL | Food service | Marketing | 3 | Reject |
| 14 | Spanish | SQL | Building industry | Marketing | 3 | Reject |
| 15 | Mandarin | GIMP | Automotive industry | Human resources | 2 | Ambiguous |
| 16 | French | LaTeX | Automotive industry | Human resources | 2 | Ambiguous |
| 17 | Mandarin | HTML | Finance | Human resources | 2 | Ambiguous |
| 18 | Russian | GIMP | Finance | Human resources | 2 | Ambiguous |
| 19 | French | HTML | Building industry | Human resources | 2 | Ambiguous |
| 20 | Russian | LaTeX | Building industry | Human resources | 2 | Ambiguous |
| 21 | French | HTML | Automotive industry | Web design | 2 | Ambiguous |
| 22 | Russian | LaTeX | Automotive industry | Web design | 2 | Ambiguous |
| 23 | Spanish | LaTeX | Food service | Web design | 2 | Ambiguous |
| 24 | French | SQL | Food service | Web design | 2 | Ambiguous |
| 25 | Russian | HTML | Building industry | Web design | 2 | Ambiguous |
| 26 | Spanish | SQL | Building industry | Web design | 2 | Ambiguous |
| 27 | Mandarin | HTML | Automotive industry | Sales | 2 | Ambiguous |
| 28 | Russian | GIMP | Automotive industry | Sales | 2 | Ambiguous |
| 29 | Russian | HTML | Finance | Sales | 2 | Ambiguous |
| 30 | Spanish | SQL | Finance | Sales | 2 | Ambiguous |
| 31 | Spanish | GIMP | Food service | Sales | 2 | Ambiguous |
| 32 | Mandarin | SQL | Food service | Sales | 2 | Ambiguous |
| 33 | Spanish | GIMP | Finance | Marketing | 2 | Ambiguous |
| 34 | Mandarin | SQL | Finance | Marketing | 2 | Ambiguous |
| 35 | Mandarin | GIMP | Food service | Marketing | 2 | Ambiguous |
| 36 | French | LaTeX | Food service | Marketing | 2 | Ambiguous |
| 37 | Spanish | LaTeX | Building industry | Marketing | 2 | Ambiguous |
| 38 | French | SQL | Building industry | Marketing | 2 | Ambiguous |

| Set | Attribute 1: Language skills | Attribute 2: Computer skills | Attribute 3: Work experience | Attribute 4: Professional training | Similarity | Decision |
|---|---|---|---|---|---|---|
| 39 | French | LaTeX | Building industry | Human resources | 3 | Invite |
| 40 | French | LaTeX | Automotive industry | Web design | 3 | Invite |
| 41 | French | HTML | Building industry | Web design | 3 | Invite |
| 42 | Spanish | LaTeX | Building industry | Web design | 3 | Invite |
| 43 | Russian | GIMP | Finance | Sales | 3 | Invite |
| 44 | Mandarin | SQL | Finance | Sales | 3 | Invite |
| 45 | Mandarin | GIMP | Food service | Sales | 3 | Invite |
| 46 | Mandarin | GIMP | Finance | Marketing | 3 | Invite |
| 47 | French | LaTeX | Finance | Web design | 3 | Invite |
| 48 | Mandarin | GIMP | Building industry | Sales | 3 | Invite |

*Note.* Sets 1 to 4 (in italics) were training exemplars.

# Appendix B

## Selection of Items Used in Experiment 2 Based on a Simulation With the Generalized Context Model

Stimuli were chosen on the basis of a simulation with the generalized context model (GCM; for an overview see Nosofsky, 2011). Table B1 shows the simulation results. Parameter settings were chosen following the recommendations of Nosofsky and colleagues (Nosofsky, 1985, 2011; Nosofsky & Johansen, 2000).

In the GCM, the similarity $s_{ij}$ between item *i* and exemplar *j* is a function of the distance $d_{ij}$ between the item *i* and exemplar *j* (see Equation B1). Parameter *p* defines the shape of the similarity function, and parameter *c* is a sensitivity parameter. The larger the value of *c*, the faster similarity will fall off with increasing distance of attribute values.

$$s_{ij} = \mathrm{e}^{-cd_{ij}^{p}} \tag{B1}$$

The distance $d_{ij}$ is calculated by using the weighted Minkowski power model. That is, in the GCM, every exemplar *j* and item *i* is represented as a point in an *M*-dimensional psychological space (Nosofsky, 2011). The values $x_{im}$ and $x_{jm}$ in Equation B2 are the values of exemplar *j* and item *i* on dimension *m*. The parameter *r* determines the distance metric, and $w_m$ is the attention weight for dimension *m*. The attention weight reflects the percentage of the attention to each dimension. The sum of all attention weights equals 1.

$$d_{ij} = \left[\, \sum_{m=1}^{M} \left( w_m \left| x_{im} - x_{jm} \right|^{r} \right) \right]^{\frac{1}{r}} \tag{B2}$$

The probability of choosing one category is calculated by summarizing the weighted similarities of item *i* to all exemplars from category $\mathfrak{J}$ and dividing them by the sum of the weighted similarities to all exemplars from all categories (see Equation B3). The number of training exemplars is denoted *n*, and $K_n$ denotes the available categories. $V_{j\mathfrak{J}}$ represents the memory strength of exemplar *j* in category $\mathfrak{J}$. The response-scaling parameter γ defines the determinism in classifications. Response bias for category $\mathfrak{J}$ is denoted $b_{\mathfrak{J}}$.

$$P(\Im|i) = \frac{b_{\Im}\left[\sum_{j=1}^{n} v_{j\Im} s_{ij}\right]^{\gamma}}{\sum_{K=1}^{K_N} b_K \left[\sum_{k=1}^{n} v_{kK} s_{ik}\right]^{\gamma}} \quad \text{(B3)}$$

Table B1

*Simulation Results With the Generalized Context Model*

| Exemplar/item no. | Category | Probability of choosing category B | |
|---|---|---|---|
| | | Sensitivity = 1 | Sensitivity = 2 |
| 4 | A | .27 | .12 |
| 5 | A | .27 | .12 |
| 6 | A | .27 | .12 |
| 8 | B | .73 | .88 |
| 9 | B | .73 | .88 |
| 10 | B | .73 | .88 |
| 25 | Ambiguous | .50 | .50 |
| 26 | Ambiguous | .50 | .50 |
| 27 | Ambiguous | .50 | .50 |
| 28 | Ambiguous | .50 | .50 |

*Note.* Parameter settings: shape parameter $p$ = 2 (2 is recommended for highly confusable stimuli; see Nosofsky, 1985), response biases $b_k$= .5, attention weights $w_m$ = .5, response scaling $\gamma$ = 1, distance metric $r$ = 2 (2 is for Euclidean metric and is recommended for integral dimensions; see Nosofsky & Johansen, 2000). Training exemplars: 4, 5, 6 and 8, 9, 10; test items: 4, 5, 6, 8, 9, 10, and 25, 26, 27, 28. Saturation was multiplied by 0.5 to account for differences in step-size values in the Munsell color scheme.

## Appendix C

### Munsell Keys (Hue, Brightness, Saturation) and Red, Green, Blue (RGB) Values of Stimuli in Experiment 2

Table C1

| Exemplar/item no. | Munsell key | RGB value |
|---|---|---|
| 1 | 7.5R-6-4 | 185, 136, 136 |
| 2 | 7.5R-7-6 | 227, 156, 152 |
| 4 | 7.5R-5-6 | 172, 103, 99 |
| 5 | 7.5R-6-8 | 213, 123, 116 |
| 6 | 7.5R-7-10 | 255, 141, 130 |
| 8 | 7.5R-4-8 | 157, 69, 65 |
| 9 | 7.5R-5-10 | 198, 87, 79 |
| 10 | 7.5R-6-12 | 239, 107, 93 |
| 12 | 7.5R-3-10 | 139, 28, 33 |
| 13 | 7.5R-4-12 | 178, 49, 46 |
| 14 | 7.5R-5-14 | 220, 66, 58 |
| 18 | 7.5R-8-8 | 255, 174, 164 |
| 25 | 7.5R-4-6 | 145, 77, 74 |
| 26 | 7.5R-5-8 | 186, 96, 89 |
| 27 | 7.5R-6-10 | 227, 115, 104 |
| 28 | 7.5R-7-12 | 255, 133, 117 |