Semantics Drive Noticing Rates in Realistic Change Blindness Scenarios

Muhammad Ramish
Centre for Applied Computing
University of Oulu
Oulu, Finland
Muhammad.Ramish@oulu.fi

Matti Pouke
Centre for Applied Computing
University of Oulu
Oulu, Finland
matti.pouke@oulu.fi

Steven LaValle
Centre for Applied Computing
University of Oulu
Oulu, Finland
Steven.LaValle@oulu.fi

Timo Ojala
Centre for Applied Computing
University of Oulu
Oulu, Finland
timo.ojala@oulu.fi

Evan Center
Center for Machine Vision and Signal Analysis
University of Oulu
Oulu, Finland
evan.center@oulu.fi

Abstract

Change blindness is the inability to perceive visual changes when the transient signals associated with the change are interrupted or absent. Previous research has found that both low-level perceptual features and higher-level semantic information can guide visual attention. However, in immersive environments, their relative contributions to change detection remain unclear. Here we investigate the effects of low-level and semantic similarity on change detection in virtual reality (VR) while controlling for bottom-up salience. Using a between-subjects design, we conducted an experiment in which $n = 64$ participants performed a visual search task consisting of two phases: (1) a target search phase, and (2) a change detection phase. Changes were induced using a blink-contingent paradigm, where the manipulated objects teleported between two fixed positions during eye blinks. A Cox proportional hazards model was used to analyze detection performance, which revealed that the semantic similarity between the target and the manipulated object was a stronger predictor of detection than low-level similarity. These findings argue that, contrary to many results using simple displays, higher-level semantic relationships play a dominant role in guiding attention during change detection in more realistic environments.

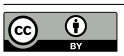
CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*; **Virtual reality**.

Keywords

change blindness, inattentional blindness, virtual reality, visual attention, visual search, visual perception, semantics, eye-tracking

ACM Reference Format:
Muhammad Ramish, Matti Pouke, Steven LaValle, Timo Ojala, and Evan Center. 2026. Semantics Drive Noticing Rates in Realistic Change Blindness Scenarios. In *ACM Symposium on Applied Perception 2026 (SAP '26), August 26–27, 2026, Rennes, France*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3821409.3821469



SAP '26, Rennes, France

ACM ISBN 979-8-4007-2799-3/26/08
https://doi.org/10.1145/3821409.3821469

1 Introduction

The human visual system is highly effective at constructing a coherent representation of complex environments. However, this representation is incomplete, as observers are unable to detect significant changes that occur repeatedly in the visual environment [Rensink et al. 1997]. This phenomenon is known as *change blindness*, which is defined as the inability to perceive visual changes when the transient signals associated with the change are interrupted or absent [Rensink 2002; Rensink et al. 1997]. Change blindness has been widely studied in controlled experimental models such as the flicker paradigm, where transient interruptions prevent the detection of the change [Rensink 2002]. Especially in conditions where changes have to be observed by comparing the representations of the internal scene rather than immediate sensory input, these findings emphasize important limitations in visual attention and memory [Alvarez and Cavanagh 2004; Blackmore et al. 1995; Rensink 2000, 2002; Rensink et al. 1997; Simons and Levin 1998].

Previous studies indicate that both perceptual and cognitive factors can influence change detection. As low-level perceptual characteristics (i.e., color, contrast, size, and orientation) affect the allocation of one's attention within a visual scene, higher-level semantic characteristics, such as the relevance of an object to the overall meaning of the scene, are also important for determining whether a change is noticed [Spotorno and Faure 2011; Spotorno et al. 2013]. Previous research indicates that changes that involve highly salient or semantically significant objects are detected more efficiently [Rensink et al. 1997; Spotorno and Faure 2011].

However, the relative contribution of low-level or semantic factors to change detection remains unclear. Although some studies report that low-level features and/or perceptual salience have a stronger influence [Williams and Simons 2000; Wood and Simons 2017b; Yeh and Yang 2008; Zelinsky 2003], or report an explicit lack of a semantic influence [Ding et al. 2024], others favor semantic consistency and scene understanding [Parasuraman and Martin

2001; Spotorno and Faure 2011; Spotorno et al. 2013]. As these factors are not usually independently controlled, the effects can be confounded. Consequently, inconsistencies in definition and measurement practices have led to contradictory findings across studies [Spotorno and Faure 2011; Spotorno et al. 2013].

When transient signals are interrupted or removed, change detection depends on the viewer's ability to retain and compare representations over time [Rensink 2002]. However, visual memory is selective and only allows the encoding of a subset of elements in the scene [Alvarez and Cavanagh 2004]. In cases involving objects that are not central to the scene or lack visual salience, this limitation contributes to the failure to detect changes [Rensink 2002].

Virtual reality (VR) has emerged as a capable platform for studying perception under more naturalistic conditions. VR allows users to observe their surroundings through natural sensorimotor contingencies which can also elicit a sensation of visiting a location similar to a real-world experience [Slater 2009]. This makes VR a promising alternative for traditional perception experiments since it allows controlled manipulation of complex three-dimensional (3-D) environments similar to real-world environments, with the potential for increased ecological validity [Martin et al. 2023]. Previous work, though more limited in implementation than the current experiment, shows that change blindness persists in immersive environments [Schöne et al. 2021; Steinicke et al. 2011]. Furthermore, change blindness has been leveraged in VR to allow novel interaction techniques. For instance, users' failure to detect changes in object position has been exploited by haptic remapping, creating the illusion of interacting with multiple virtual objects using a single physical proxy [Lohse et al. 2019]. There is a lack of studies that focus on investigating the influence of perceptual and semantic factors on change detection in VR environments. It is critical to understand this distinction for both theoretical models of perception and practical applications in VR system design.

Traditional approaches to induce change blindness, such as flicker-based methods or artificial visual interruptions, may not reflect real-world scenarios [Rensink 2002; Rensink et al. 1997]. Other approaches have explored inducing changes during natural interruptions, where scene modifications occur during saccades or eye blinks [Bridgeman et al. 1975; Langbehn et al. 2016]. Such methods enable more ecologically valid ways to study change blindness, and they circumvent the "one-shot" nature of previous change blindness paradigms that take place outside of the lab [Jensen et al. 2011; Simons and Levin 1998] as blinks can become akin to trials.

The current study aims to address these gaps by investigating the effects of perceptual and semantic similarity between a target and a manipulated item on change blindness in a VR environment using a blink-contingent paradigm. Participants are tasked with searching for a target, which is meant to bias their attentional template towards the features of that target [Duncan and Humphreys 1989; Wolfe 1994]. By comparing the rates at which they notice a manipulated item that shares either perceptual or semantic features with the target, we may then infer whether the features present in their attentional template are more perceptual or semantic in nature. In a preregistered study with 64 participants, we find in our case that semantic information outweighs low-level information in driving change detection, and we go on to make the case that the conflicting results in previous work on this topic are likely determined the realism of the stimuli and environments used.

2 Related Work

Though this study was focused on the determinants of change blindness in realistic scenarios, the methods invoke a mix of aspects from visual search, inattentional blindness, and change blindness, and thus each is discussed here in turn.

2.1 Visual Search and Attention Guidance

Visual attention is inherently selective, and scene representations are, therefore, not uniform. Previous work has shown that observers selectively encode objects based on attentional priority, which determines the parts of a scene that should be consolidated sufficiently to support later detection [Rensink 2002; Spotorno and Faure 2011]. For complex scenes, where several objects compete for limited attentional and mnemonic resources, this is particularly important [Martin et al. 2023; Spotorno and Faure 2011]. Hence, attentional selection, visual memory, and comparison processes collectively determine what information is encoded and later available for change detection.

Bottom-up saliency has been highlighted by early models of visual attention. In this mechanism, attention is guided by low-level features such as luminance, color, contrast, and orientation, toward visually distinctive regions [Itti et al. 1998]. This view has been formalized by computational saliency models, which generate saliency maps of the visual scene that help predict where observers are probable to fixate. Through these models, a significant part of attention allocation can be explained, specifically in simplified displays or when an item visually contrasts with its surroundings. However, a large body of work argues that bottom-up salience does not control attention in naturalistic environments alone. Top-down factors, such as task goals, expectations, and prior knowledge, strongly drive attentional selection [Wolfe 1994]. According to the Guided Search Framework, it is proposed that attention is driven by both bottom-up feature information and top-down relevance signals, enabling observers to match items based on their current search template [Wolfe 1994]. In realistic contexts, this means that the focus of attention is not only decided by visually prominent items, but also by how the observer interprets the scene and their task.

Attention is further guided by semantic information in real-world scenes. Studies have shown that observers use scene meaning and object-scene relationships to efficiently guide visual search rather than scanning scenes as collections of isolated features [Henderson and Hayes 2017; Le-Hoa Võ and Wolfe 2015]. As a result, attention is preferentially allocated to objects that are semantically relevant to the scene or the current task. At the same time, semantically inconsistent or unexpected objects can also attract attention because they violate scene expectations [Bainbridge et al. 2017, 2021; Cornelissen and Võ 2017; Plano and Williams 2025]. Eye-tracking studies support these findings by demonstrating that fixations are biased toward informative and semantically meaningful regions rather than solely toward visually salient regions [Cornelissen and Võ 2017; Henderson and Hayes 2017]. Thus, in

complex scenes, semantic guidance can sometimes override bottom-up salience effects, particularly in realistic environments [Stirk and Underwood 2007].

2.2 Change Blindness and Inattentional Blindness

Change blindness refers to the failure to detect visible changes in a scene when the local transient that normally signals the change is masked, interrupted, or otherwise unavailable [Rensink 2002; Rensink et al. 1997]. Foundational studies demonstrate that visual experience does not automatically produce a complete and continuously updated representation of the scene. For instance, in the flicker paradigm, when observers are shown two alternating scenes separated by a blank gray screen, they are often unable to notice even large changes [Rensink 2002]. Similarly, in the interruption paradigm, natural visual interruptions such as blinks or saccades mask the motion signal that normally signals a change [Bridgeman et al. 1975; Rensink et al. 1997]. This finding has since been extended to both laboratory and more naturalistic settings, including immersive environments [Martin et al. 2023; Steinicke et al. 2011].

A primary mechanism for this phenomenon is thought to be related to selective attention. Large and obvious changes in the environment go unnoticed because one's attention is not directed to the region where the change occurs [Rensink 2002; Rensink et al. 1997]. When an informative local transient is absent, the attention of the observer must first be directed to the relevant region and then compare the current input with a stored representation of the previous scene [Mitroff et al. 2004; Rensink 2002]. Therefore, for successful change detection, information must not only be present in the scene but also be selected, encoded, and made available for later comparison.

This perspective places visual memory at the center of change blindness research. Multiple studies state that the limited capacity and precision of visual working memory are closely related to failures in change detection [Brady et al. 2013; Martin et al. 2023; Rensink 2002]. Martin et al. argue that the visual system only keeps a restricted subset of objects and properties, with trade-offs between the number of retained elements and the fidelity of their representations [Martin et al. 2023]. Moreover, studies related to scene and object memory have found that change detection often fails if the preserved information about previously viewed objects is incomplete, weakly encoded, or unavailable at the moment of comparison [Mitroff et al. 2004; Yeh and Yang 2008]. Moreover, previous research also shows that change blindness cannot be limited to a mere absence of representation. Studies have found that change detection and later memory performance are dissociated, implying that more information is retained by the observers than what is disclosed by their immediate detection response [Hollingworth 2003; Mitroff et al. 2002; Yeh and Yang 2008]. Therefore, failure to retrieve or compare pre-change and post-change representations is an important factor in change blindness.

A closely related phenomenon to change blindness is *inattentional blindness*. This phenomenon is defined as the failure to notice unexpected but visible stimuli when attention is engaged elsewhere [Simons and Chabris 1999]. The "invisible gorilla" experiment is one of the classic demonstrations of inattentional blindness, where participants often miss a person in a gorilla suit when counting basketball passes [Simons and Chabris 1999]. In addition, the "door study" is another example in which, in a real-world interaction, a person is replaced during a brief interruption without being noticed [Simons and Levin 1998]. The findings of these studies indicate that conscious perception is more dependent on attentional allocation than on the physical visibility of stimuli.

Even for large and visible changes, disrupting the visual transient (e.g, flicker, blinks, saccades, mudsplashes, etc.) can significantly influence change detection [O'Regan et al. 1999]. Furthermore, change detection is not a strictly all-or-none process. A recent study found that observers sometimes "sense" that a change has happened in the scene but fail to identify or locate it explicitly, showing a form of partial awareness [Scrivener et al. 2020]. Likewise, large-scale studies in inattentional blindness research have found that participants who fail to notice a stimulus can still reveal an above-chance sensitivity to its features, which suggests residual processing outside of the conscious report [Nartker et al. 2021]. Hence, these findings indicate that failure to detect a change may reflect gaps in access or comparison processes rather than total absence of representation.

Although extensive work has been done in this research area, most studies have relied on static or highly controlled paradigms, whereas those that take place "in the wild" are difficult to control for confounds and difficult to run in general, lending to their lower prevalence in the change blindness literature. Therefore, many details remain unclear regarding how attentional guidance and change detection operate in realistic and dynamic environments.

2.3 Low-level vs. Semantic Influences

An important question is built on the mechanisms of attentional guidance described above, which is how different types of similarity effect change detection performance. Specifically, low-level and semantic relations between the participant's search template and a manipulated item can affect change detection through different processes.

Objects are similar at the low-level when they visually resemble each other based on features such as color, shape, texture, or overall appearance. For example, in the space of all manmade artifacts, a toothbrush is perceptually similar to a paintbrush. In paradigms where one object is replaced with another (unlike ours, which changes the location of a single item), changes between perceptually similar objects may be more difficult to notice due to reduced discriminability. The observer may struggle to differentiate between the pre-change and post-change states even when attention is directed toward the relevant location in such cases [Mitroff et al. 2004; Rensink 2002]. Therefore, the probability of change blindness increases, especially when detection depends solely on comparison between views and transient signals are interrupted [Rensink 2002].

Whereas, semantic similarity shows relationships that are based on meaning, category, or functional association. For example, a tube of toothpaste is semantically related to a toothbrush. Semantically related objects can be more easily confused or substituted within memory representations, despite differing in low-level features. Particularly in terms of memory encoding and comparison, this suggests that semantic similarity may mainly affect later stages of

change detection [Chong and Beck 2024; Spotorno et al. 2013; Yeh and Yang 2008]. For instance, if two objects are semantically similar (i.e. sharing similar conceptual representations), observers may not be able to detect a change due to the new object being interpreted as consistent with the previously stored representation.

Despite the findings discussed above, limited research has directly compared the influence of low-level and semantic relationships on change detection within the same experimental framework. This limitation is particularly relevant in immersive environments, where richer scene context may amplify the role of semantic information [Martin et al. 2023]. The current study addresses this gap by investigating the influence of low-level and semantic similarity on change detection performance in a controlled VR environment while controlling for bottom-up salience, with the hypothesis that change detection will be driven largely by semantic associations to targets.

3 Methods

All procedures were approved by the Ethics Committee of Human Sciences of the University of Oulu. The study's sample size, exclusion and replacement criteria, materials, design, and analyses were preregistered at OSF (anonymous link).

3.1 Experimental Design and Procedure

The study used a between-subjects design with a single independent variable: the type of object similarity to a target item. In this design, only one similarity condition within one object set was presented to each participant to prevent learning effects, strategy development, and carryover effects arising from repeated exposure to multiple similarity types. Throughout this paper, the term *manipulated object* refers to the object whose position was altered between two predefined locations during participant blinks. Meanwhile, the target object (the target of the visual search task) remained stationary throughout the experiment and never changed location. In the *low-level* similarity condition, the manipulated object shared visual features with the target object, such as shape, color, or texture. However, the two objects differed in terms of their semantic category. Meanwhile, in the *semantic similarity* condition, the manipulated object was conceptually related to the target object (i.e., belonging to the same semantic class), however, the two objects were visually distinct from each other. Figure 1 displays the target object and manipulated object combinations used in both similarity conditions. Objects within a set (i.e., a row in Figure 1) were similarly sized.

Participants were alternately assigned to one of the two conditions. When participants arrived, they were briefed about the task and instructed to complete the eye-tracking calibration to facilitate reliable gaze tracking. The participants then completed two short practice search tasks that did not include change detection mechanics. The purpose of these shorter tasks was to get the participants accustomed to the VR environment and interaction mechanics and set up the pretext that this was merely a visual search experiment. Moreover, the practice tasks allowed us to estimate baseline blink behavior. The practice targets were intentionally designed to require active visual search rather than immediate detection, thereby providing sufficient observation time for estimating individual blink rates. An important aspect of the design was that participants remained naive to the true purpose of the experiment during their first exposure to the main environment. The practice tasks contained no manipulated objects and were presented as ordinary visual search tasks. Revealing the blink-contingent nature of the changes beforehand could have substantially altered participants' search strategies and reduced the effectiveness of the paradigm, which is also why a between-subjects design was deemed necessary over a within-subjects design.

Figure 1: Target objects and manipulated object pairs used in the experiment (not depicted to scale) [Ramish 2026].

In the main task, participants were asked to search for a target using the basic level labels shown in 1. They were not shown any images or given any further descriptions.. Participants were allowed to search for the target object for up to 120 seconds. After identifying the target object, participants confirmed the detection by pressing the trigger button on the Valve Index controller, triggering the start of the change detection phase. The purpose of the target search phase was to establish a search template in visual working memory while promoting active exploration of the environment.

After successful detection of the target object, the experiment immediately transitioned to the change detection phase. Participants had to detect the manipulated object that was teleporting between two fixed positions in the environment. The change always took place during blinks (i.e., when both eyes were detected as closed by the eye-tracking system). To avert rapid successive changes and inaccurate blink counts, a cooldown period of 500 ms was implemented. Furthermore, the manipulated object only teleported if it was in the participant's field of view when they blinked (i.e., inside the camera frustum in Unity). If the participant blinked when the manipulated object was outside the visible region, no change was triggered, therefore ensuring that all changes were induced under conditions where they could be observed. Synchronizing the changes with blinks allowed us to induce change blindness in a controlled yet natural manner, as participants were unable to directly perceive the motion transient.

All object sets were assigned the same predefined target positions and moving object positions to remove any potential biases of room layout cues and location-based detection difficulty. These positions

and the standing position of the participant (marked "VR Rig" in Figure 2) are depicted in Figure 2. Each participant was exposed to only one of the object sets with the manipulated item deriving only from either the low-level or semantic category.

The change detection phase had a maximum duration of 155 seconds. During this phase, a sequence of three prompts was presented to the participant through the virtual user-interface (UI), progressively informing them and guiding them toward detecting the change (mirroring the progressively more narrow line of questions used in [Simons and Chabris 1999]). Each prompt was overlayed on the UI for 10 seconds, after which a maximum duration of 45 seconds was allowed for the participant to detect the change. If the participant could not detect the change, the stage proceeded to the next prompt. In case of successful detection during any of the stages (i.e., after prompts 1, 2, or 3), the experiment terminated immediately and a completion message was displayed. If no detection occurred within during any of the stages, the experiment automatically terminated, and the participant was treated as a non-detection (right-censored) case. Participants received progressively more specific prompts: (1) "Did you see anything unusual happening? Look for it!", (2) "Something is disappearing from one location and appearing in another. Look for it!", and (3) "Something is disappearing from one location and appearing in another when you BLINK! Look for it!"

Note that the manipulated object was already undergoing blink-triggered teleportation during the target search phase. However, participants were not informed that any object was changing and were instructed only to search for the target object. Thus, although the manipulated object could have been noticed incidentally during target search, it was not yet explicitly task-relevant. If a participant had already noticed the manipulated object during this phase, it could be reported immediately after the target object was found and the change detection phase began.

3.2 Participants

The study included an initial pilot phase that was followed by the main study. The pilot study included 16 participants (N=16; 10 male, 6 female) and was conducted to assess the feasibility of the task, validate the blink-contingent change induction mechanism, and obtain preliminary estimates of the effect size. The sample size for the main study was estimated from the results of the 16 participant pilot study, using a priori power analysis within the Cox proportional hazards framework and the Schoenfeld approximation [Schoenfeld 1981]. According to the estimation, a sample size of 64 participants was planned using a between-subjects design that provided equal allocation under the two similarity conditions, which would provide sufficient statistical power (0.80) to detect a medium effect size (hazard ratio = 2.5) using an alpha level of .05 (two-sided).

Participants were recruited through the university's online recruitment system. Before participation, all participants provided their written informed consent and were screened for sensitivity to motion sickness and nausea due to the VR setup. Participants were at least 18 years of age, had normal or corrected-to-normal vision (using corrective eye-wear in the HMD if needed), had no known neurological impairments, could stand comfortably for 20 minutes and rotate their head, and did not participate in the pilot study. Participants were excluded if they had less than 90% valid gaze data throughout the experiment, if their gaze data had a continuous gap of over 2 seconds, if any other log fields had missing or corrupted data, if they failed to complete the experiment due to technical malfunction or early self-termination, or if they pressed the trigger button to select targets excessively (i.e., "button spamming" was defined as a trigger count Z-score greater than 2.5).

Participants were included in the censored data for analysis involving detection time and outcome if they made no detection attempts (i.e., zero trigger presses) during the complete task. These cases indicate that the participant either did not engage with the detection task or did not complete it. Moreover, if their detection time was $\geq$ 155 seconds (the maximum allowed time), they were also treated as censored cases.

3.3 Apparatus and Environment

The experiment was carried out in a VR setup showing an indoor bedroom scene (Fig 2), designed in Unity (version 2022.3.47f1) [Unity Technologies 2026]. The virtual environment contained multiple everyday objects in a cluttered arrangement. The experiment setup supported the controlled measurement of behavioral variables for the visual search task. Participants were allowed to explore the environment through natural head movements from a fixed position.

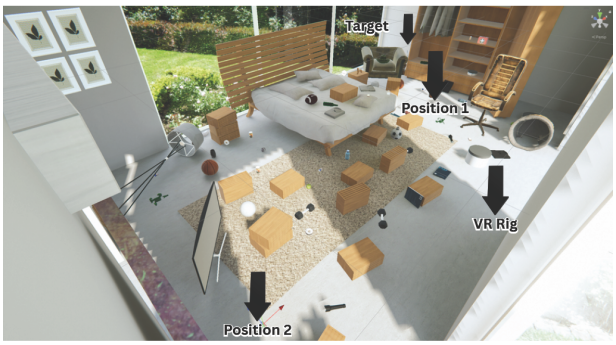


Figure 2: The VR setup for the experiment, designed and implemented in Unity (version 2022.3.47f1). The positions of the participant (VR Rig), the target, and the manipulated object are marked [Unity Technologies 2026] [Ramish 2026].

The head-mounted display (HMD) used in the experiment was a Varjo XR-4, at a high resolution (3840 × 3744 pixels) with visual rendering and integrated eye-tracking capabilities, which continuously recorded gaze behavior and detected blinks during the experiment [Varjo Technologies 2026]. Whereas a blink takes roughly 100 to 150 ms on average (with "blink suppression" lasting 200 to 250 ms) [Volkmann et al. 1980], an object can be teleported in the space of one frame lasting only $\approx$ 11 ms in our pipeline, which allowed for covert repositioning of objects akin to the flicker paradigm. A Valve Index controller was used for interaction with the environment as a pointing device to facilitate participants in indicating the detected changes through a trigger press [Valve Corporation 2023].

The object pairs were matched on low-level features such as size, color, luminance, contrast and were positioned in consistent spatial locations across conditions, to control for bottom-up salience. To ensure that these object pairs did not differ systematically in their ability to attract attention, a computational salience analysis inspired by the Itti-Koch-Niebur model was conducted during stimulus preparation [Itti et al. 1998]. First, a 2-D screenshot was captured for each object scene from the participant's viewpoint. A salience map was then generated for each screenshot which was based on the intensity conspicuity channel, highlighting regions that stand out in brightness compared to their surroundings, across both fine and coarse image details. A binary region-of-interest (ROI) mask was then created for each manipulated object to estimate the object-level salience. Each ROI mask included an 'object area' which was marked in white using a rectangular region that covered its pixel extent, leaving the rest of the image marked in black. For each manipulated object, the mean salience value within this mask was used as the main salience score. Salience balance was evaluated at both pairwise and class levels. The salience values of the low-level and semantic manipulated objects were compared within each class to minimize differences. In addition, mean salience values at the class level were compared against the global average for all manipulated objects to avoid systematic bias towards either condition. Objects were iteratively edited in Blender until a satisfactory range of salience values was achieved.

The salience validation confirmed that the manipulated object set was broadly balanced in terms of intensity-based salience. Across all objects (target and manipulated), salience values remained within the predefined tolerance range of ±0.1 relative to the global manipulated object average. Critically, individual low-level and semantic manipulated object pairs differed no more than ±0.1 and the average of these differences between conditions was effectively zero. Detailed salience statistics are provided in Appendix Table 1. Although this procedure reduced systematic differences in bottom-up salience, minor residual differences cannot be completely eliminated and should be considered when interpreting the results.

3.4 Measures and Analyses

The main dependent variable was change detection performance, used as a combined time-to-event outcome that represents both *detection time* and *detection success*. Detection time was defined as the period between the start of the change detection phase (i.e., immediately after the first prompt) and the detection of the manipulated object. As discussed, non-detection cases were treated as right-censored observations which allowed the use of a survival analysis framework. *Weighted average blink rate* was included as a participant-level covariate in the Cox proportional hazards regression model to control for differences in baseline blinking behavior, as participants who naturally blink more often might obtain more chances to spot the manipulated item, given that the item only moved during blinks. This variable was computed as the sum of blinks divided by the sum of task durations across both practice tasks (see Appendix).

Before the main analysis, a time compression procedure was applied that removed fixed-duration intervals corresponding to the presentations of prompts (i.e., 10 seconds each) from the timeline. Therefore, the original maximum duration of 155 seconds was reduced to 135 seconds, ensuring that the detection times only reflected active search behavior and not time spent waiting in a prompt menu in which changes were not detectable. As a combined measure of detection probability and speed, we estimated a Cox proportional hazards regression model [Cox 1972] wherein detection at any given moment was predicted by a participant's similarity condition and their weighted average blink rate (see Appendix). This model would serve as the primary test of our hypothesis that manipulated objects that were semantically similar to targets would be better detected than manipulated objects that were perceptually similar. Wald tests were used to evaluate the statistical significance of the model coefficients [Wald 1943], and the log-rank test was used to assess differences between survival distributions across conditions [Mantel 1966]. To evaluate the proportional hazards assumption, scaled Schoenfeld residuals were computed and formally tested [Schoenfeld 1981] (see Appendix).

In addition, exploratory analyses were conducted to study the variability between stimuli by integrating object identity at the item level. Mixed-effects extensions of the Cox model were used to account for random variability across stimuli.

4 Results

A total of 64 participants, equally distributed between conditions (i.e., low-level similarity: n=32, semantic similarity: n=32), were included in the main analysis. Five participants were excluded and replaced prior to the main analysis because their total controller trigger presses exceeded the predefined outlier threshold (Z-score > 2.5). Three exclusions occurred in the low-level similarity condition and two in the semantic similarity condition. The final sample consisted of 46 male and 18 female participants, ranging from 18 to 40 years, and a median age of approximately 26. In addition, most of the participants reported that they had never used VR at all or had only used it once or twice.

4.1 Confirmatory Results

In terms of overall detection performance, 81% of participants in the semantic similarity condition successfully detected the change compared to 58% in the low-level similarity condition. In addition to higher detection success, we found that participants in the semantic similarity condition detected the change more quickly than participants in the low-level condition. Figure 3 plots the detection speed for both categories.

Our primary confirmatory analysis combined both detection rate and time information using Cox proportional hazards models. Kaplan-Meier survial curves depicting the time-to-detection between conditions are shown in Figure 4. Of the 63 participants included in the survival analysis, 19 were treated as right-censored observations because the manipulated object (or in one case, the target object) was not detected within the allotted time. This included 13 participants in the low-level similarity condition and 6 participants in the semantic similarity condition. Consequently, the survival analyses were based on 44 observed detection events and 19 censored observations. The figure displays that participants in the semantic similarity condition exhibited a faster decline in survival

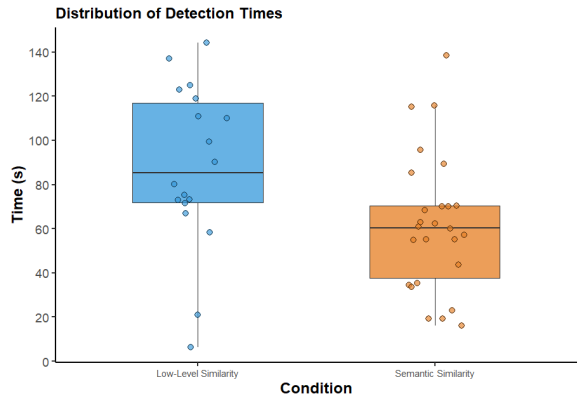


Figure 3: Distribution of detection times across conditions, shown only for those who successfully detected the change [Ramish 2026].

probability (i.e., they noticed the manipulated item sooner). This finding further reinforces earlier detection of the manipulated object in the semantic similarity condition compared to the low-level similarity condition. The distinction between the curves appears very relatively early and stays consistent over time. A significant difference was confirmed between the survival distributions of the two conditions using a log-rank test ($p = 0.004$), which shows improved detection performance for semantically related manipulated objects.

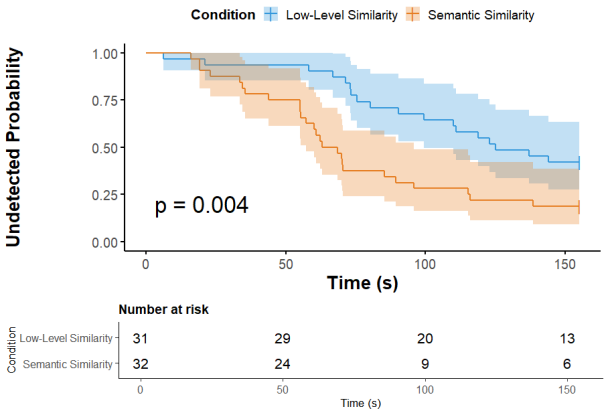


Figure 4: Kaplan–Meier survival curves showing the probability of remaining undetected over time for each condition. Shaded regions represent confidence intervals [Ramish 2026].

The Cox proportional hazards model showed that there is a significant effect of the similarity condition on the change detection performance ($HR = 5.57$, $95\%CI[1.42, 21.78]$, $p = 0.014$), as shown in Appendix Table 2. At any given time point, the odds that participants in the semantic similarity condition could detect the change were approximately 5.6 times those of participants in the low-level similarity condition. Furthermore, the table also shows that the weighted average blink rate did not significantly affect detection performance ($HR = 0.28$, $p = 0.377$), suggesting that differences in blinking behavior at the participant-level could not explain the probability or speed of detection.

4.2 Confound Analyses

An item-level Cox model was estimated without considering the similarity condition to examine whether detection performance varied between stimuli. The vase was used as the reference category; therefore, all hazard ratios represent comparisons relative to this object. The trumpet exhibited a significantly higher detection hazard than the reference object (HR = 9.78, p = .001), as did the tortoise (HR = 5.57, p = .014). These findings indicate that some stimulus-specific differences were present. However, the primary effect of similarity condition remained after accounting for item-level variability, suggesting that the observed effect was not driven by a single highly noticeable object. Detailed model coefficients are provided in Appendix Table 3.

We estimated variation due to participants and stimuli using a Cox model with random effects for those factors, in addition to fixed effects for condition and blink rate. There was greater variability at the participant level ($\sigma^2 = 0.37$) compared to the object level ($\sigma^2 = 0.06$), revealing that differences at the individual level contributed more to the variability of the detection than differences between stimuli.

5 Discussion

From multiple perspectives, the findings consistently support the notion that semantic similarity significantly improves change detection performance in an immersive environment. Participants in the semantic similarity condition were more likely and faster to detect the change than those in the low-level similarity condition. This effect was persistent across survival analyses and was not explained by individual differences in blink rate or effects of specific stimuli.

Our findings imply that the importance of semantic and low-level factors on change blindness may depend on the nature of the visual environment. In much of the classic change blindness literature, the studies have used simplified or static displays where low-level features are more important in guiding attention and detection [Rensink 2002; Rensink et al. 1997]. In such cases, rich contextual information is missing, and therefore low-level similarity may dominate. However, since this study used a realistic VR environment with objects positioned within a coherent and meaningful scene, semantic relationships become dominant in guiding attention and detection [Bahle et al. 2025; Cornelissen and Võ 2017]. This leads to a broader interpretation that attentional guidance may increasingly depend more on semantics than low-level features in more realistic and semantically structured environments.

This trend is noticeable across the inattentional blindness and change blindness literatures. In grayscale 2-D displays of simple shapes, low-level features are known to drive change detection [Most et al. 2001]. Even more recent investigations have targeted semantically related concepts such as the role of memory [Wood and Simons 2017a], where some manipulated items were studied beforehand, and familiarity [Ding et al. 2024], where unexpected items were familiar or unfamiliar flags and logos, and neither found

impacts on noticing rates. However, all of these studies used isolated objects over uniform backgrounds, and were thus completely devoid of the contextual information that coincides with meaningful objects in the real world. In comparison, increasing realism by using comic strips [Spotorno and Faure 2011; Spotorno et al. 2013] or real-world scenes [Chong and Beck 2024] produces an effect of semantics that is on par with or greater than that of low-level features. By increasing the realism to a much greater level still by leveraging VR, we now observe a dominance of semantic factors over perceptual ones.

Although visual search was more of a a tool in our change blindess paradigm than the main focus of the study, it cannot be ignored that a similar pattern of effects exists here as well, and we are not the first to make this argument [Stirk and Underwood 2007]. When using simple stimuli [Godwin et al. 2014], or even real objects but in grayscale and overlaid within an unnatural context [Yeh and Peelen 2022; Zelinsky 2003], perceptual features tend to guide search. However, when realistic stimuli are used in a contextually congruent manner, such as in airport security scans [Biggs et al. 2015] or natural images [Cornelissen and Võ 2017; Stirk and Underwood 2007], semantic information becomes more impactful or even overrides low-level perceptual information. To the degree that context-congruent semantic information is present, participants will use it [Bahle et al. 2025].

At a functional level, this semantic activation effect can be interpreted in terms of both attentional allocation and memory processes. Target objects were specified using textual labels (e.g., "toothbrush") rather than visual exemplars or feature-based descriptions (e.g., "purple and white toothbrush"), which may have encouraged participants to form semantic rather than low-level visual search templates. This design increases ecological validity but may also promote greater reliance on semantic representations during search. During the target-search phase, participants likely formed semantic representations that activated related concepts and biased attention toward semantically similar objects [Cohen et al. 2017; Kramer et al. 2023]. Because changes occurred during blinks, change detection relied on visual working memory to compare scene representations [Hollingworth 2003; Mitroff et al. 2004]. In realistic scenes, semantic similarity may therefore facilitate both attentional guidance and memory-based comparison, increasing the likelihood of detecting meaningful changes.

The findings of this study have direct implications for the design and understanding of immersive virtual environments. Previous studies have highlighted that techniques such as redirection and haptic remapping can exploit change blindness in VR [Lohse et al. 2019]. This study shows that the semantic role of objects within the scene also has an influence on the effectiveness of these techniques, rather than just low-level similarity. Moreover, if changes are induced to semantically meaningful or task-relevant objects, the chances of detection increase, even when low-level differences are minimized.

Our results show that it may be important to consider the semantic structure of the environment when designing attention-aware or adaptive extended-reality (XR) systems. Users interpret realistic environments as coherent and meaningful contexts, rather than collections of isolated features. Therefore, systems designed to guide or manipulate attention should not rely solely on perceptual salience, but should also incorporate semantic context and user goals.

5.1 Limitations and Future Directions

Our results indicate the importance of studying visual attention and change detection in more ecologically valid environments. In future work, this paradigm can be extended to include a broader range of object categories, object exemplars, semantic relationships, and more dynamic interactive environments. Furthermore, future work can combine behavioral measures with neurophysiological signals [Beck et al. 2001; Koivisto and Revonsuo 2003] in immersive settings to provide deeper insights into the interaction of attentional and memory processes during change detection. More objective or scalable measures of similarity could also be explored. For instance, large language models (LLMs) or embedding-based approaches could be used to quantify semantic similarity. Similarly, approaches in computer vision could be used to train models on visual features of objects to estimate low-level similarity. More recent bottom-up salience models beyond the Itti-Koch framework could also provide different types of control for perceptual factors.

Furthermore, target items were cued at what could be considered the basic object classification level (e.g., "apple") [Mervis et al. 1981]. It is likely that objects cued at the superordinate level (e.g., "fruit") or the subordinate level (e.g., "Fuji apple") could bias the attentional template of the participant [Maxfield and Zelinsky 2012] further towards the semantic level or back towards the perceptual level, respectively. Future work could examine the degree to which the particular target item description affects noticing rates in immersive environments. In addition, due to the time and resource constraints associated with adding a third between-subjects condition, the present study did not include a baseline condition in which the manipulated object was unrelated to the target object. Such a condition could help to further clarify the contributions of similarity-based attentional guidance. Finally, although our environment is more realistic than a 2-D presentation, it does not perfectly mirror reality, and there always remains the possibility that imperfections, such as illumination artifacts, could have influenced the results.

6 Conclusion

This paper presented a blink-contingent paradigm for examining the relative influences of low-level and semantic similarity on change detection in a VR environment while controlling for visual salience. The results showed that semantic similarity was a stronger predictor of both detection success and detection speed than low-level perceptual similarity. Even after controlling for bottom-up salience and individual differences in blinking behavior, participants were more likely and faster to detect changes when manipulated objects were semantically related to a recently located target object. These findings suggest that high-level semantic relationships play a greater role than overlapping low-level visual features in guiding attention and supporting change detection in realistic, context-rich environments. Future work should investigate a broader range of object categories, semantic relationships, and interactive environments to further understand how attention, memory, and scene understanding interact during change detection.

Acknowledgments

This work was supported by the European Research Council (project ILLUSIVE 101020977), the Research Council of Finland (projects BANG! 363637 and PERCEPT 322637), and the University of Oulu and Research Council of Finland (PROFI7 352788).

References

George A Alvarez and Patrick Cavanagh. 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science* 15, 2 (2004), 106–111.

Brett Bahle, Kurt Winsler, John E Kiat, and Steven J Luck. 2025. Combined conceptual and perceptual control of visual attention in search for real-world objects. *Attention, Perception, & Psychophysics* (2025), 1–19.

Wilma A. Bainbridge et al. 2017. Disrupted object-scene semantics boost scene recall but diminish object recall. *Journal of Vision* 17, 3 (2017), 16. doi:10.1167/17.3.16

Wilma A Bainbridge, Wan Y Kwok, and Chris I Baker. 2021. Disrupted object-scene semantics boost scene recall but diminish object recall in drawings from memory. *Memory & Cognition* 49, 8 (2021), 1568–1582.

Diane M Beck, Geraint Rees, Christopher D Frith, and Nilli Lavie. 2001. Neural correlates of change detection and change blindness. *Nature neuroscience* 4, 6 (2001), 645–650.

Adam T Biggs, Stephen H Adamo, Emma Wu Dowd, and Stephen R Mitroff. 2015. Examining perceptual and conceptual set biases in multiple-target visual search. *Attention, Perception, & Psychophysics* 77, 3 (2015), 844–855.

Susan J Blackmore, Gavin Brelstaff, Kay Nelson, and Tom Trościanko. 1995. Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception* 24, 9 (1995), 1075–1081.

Timothy F. Brady, Talia Konkle, Jonathan Gill, Aude Oliva, and George A. Alvarez. 2013. Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science* 24, 6 (2013), 981–990. doi:10.1177/0956797612465439

Bruce Bridgeman, Derek Hendry, and Lawrence Stark. 1975. Failure to detect displacement of the visual world during saccadic eye movements. *Vision research* 15, 6 (1975), 719–722.

Ling Lee Chong and Diane M Beck. 2024. Real-world Statistical Regularity Impacts Inattentional Blindness. *Consciousness and Cognition* 125 (2024), 103768.

Michael A Cohen, George A Alvarez, Ken Nakayama, and Talia Konkle. 2017. Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of neurophysiology* 117, 1 (2017), 388–402.

Tim HW Cornelissen and Melissa L-H Võ. 2017. Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics* 79, 1 (2017), 154–168.

D. R. Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.

Yifan Ding, Daniel J Simons, Connor M Hults, and Rishi Raja. 2024. Are Familiar Objects More Likely to Be Noticed in an Inattentional Blindness Task? *Journal of Cognition* 7, 1 (2024), 28.

John Duncan and Glyn W Humphreys. 1989. Visual search and stimulus similarity. *Psychological review* 96, 3 (1989), 433.

Hayward J Godwin, Michael C Hout, and Tamaryn Menneer. 2014. Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review* 21, 3 (2014), 689–695.

John M. Henderson and Taylor R. Hayes. 2017. Regarding scenes. *Current Directions in Psychological Science* 26, 4 (2017), 321–327. Used for semantic guidance and scene perception.

Andrew Hollingworth. 2003. Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 388–403. doi:10.1037/0096-1523.29.2.388

Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259. doi:10.1109/34.730558

Melinda S Jensen, Richard Yao, Whitney N Street, and Daniel J Simons. 2011. Change blindness and inattentional blindness. *Wiley Interdisciplinary Reviews: Cognitive Science* 2, 5 (2011), 529–546.

Mika Koivisto and Antti Revonsuo. 2003. An ERP study of change detection, change blindness, and visual awareness. *Psychophysiology* 40, 3 (2003), 423–429.

Max A Kramer, Martin N Hebart, Chris I Baker, and Wilma A Bainbridge. 2023. The features underlying the memorability of objects. *Science advances* 9, 17 (2023), eadd2981.

Eike Langbehn, Gerd Bruder, and Frank Steinicke. 2016. Subliminal reorientation and repositioning in virtual reality during eye blinks. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. 213–213. doi:10.1145/2983310.2985765

Melissa Le-Hoa Võ and Jeremy M Wolfe. 2015. The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences* 1339, 1 (2015), 72–81.

Andreas L. Lohse et al. 2019. Leveraging change blindness for haptic remapping in virtual environments. In *IEEE Workshop on Everyday Virtual Reality (WEVR)*. 1–5. doi:10.1109/WEVR.2019.8809605

N. Mantel. 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* 50, 3 (1966), 163–170.

Daniel Martin, Xin Sun, Diego Gutierrez, and Belen Masia. 2023. A study of change blindness in immersive environments. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2446–2455. doi:10.1109/TVCG.2023.3247085

Justin T Maxfield and Gregory J Zelinsky. 2012. Searching through the hierarchy: How level of target categorization affects visual search. *Visual cognition* 20, 10 (2012), 1153–1163.

Carolyn B Mervis, Eleanor Rosch, et al. 1981. Categorization of natural objects. *Annual review of psychology* 32, 1 (1981), 89–115.

Stephen R Mitroff, Daniel J Simons, and Steven L Franconeri. 2002. The siren song of implicit change detection. *Journal of Experimental Psychology: Human Perception and Performance* 28, 4 (2002), 798.

Stephen R. Mitroff, Daniel J. Simons, and Daniel T. Levin. 2004. Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics* 66, 8 (2004), 1268–1281. doi:10.3758/BF03195000

Steven B Most, Daniel J Simons, Brian J Scholl, Rachel Jimenez, Erin Clifford, and Christopher F Chabris. 2001. How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological science* 12, 1 (2001), 9–17.

Andrew J. Nartker et al. 2021. Sensitivity to visual features in inattentional blindness. *Journal of Vision* 21, 9 (2021), 1–15. doi:10.1167/jov.21.9.11

J. Kevin O'Regan, Ronald A. Rensink, and James J. Clark. 1999. Change-blindness as a result of "mudsplashes". *Nature* 398 (1999), 34.

Raja Parasuraman and Alex Martin. 2001. Interaction of semantic and perceptual processes in repetition blindness. *Visual cognition* 8, 1 (2001), 103–118.

Andrew L Plano and Carrick C Williams. 2025. Does object-to-scene binding depend on object and scene consistency? *Attention, Perception, & Psychophysics* 87, 3 (2025), 899–908.

Muhammad Ramish. 2026. *Investigating the influence of low-level and semantic similarity on change blindness in VR.* Master's thesis. M. Ramish.

Ronald A Rensink. 2000. The dynamic representation of scenes. *Visual cognition* 7, 1-3 (2000), 17–42.

Ronald A. Rensink. 2002. Change detection. *Annual Review of Psychology* 53, 1 (2002), 245–277. doi:10.1146/annurev.psych.53.100901.135125

Ronald A. Rensink, J. Kevin O'Regan, and James J. Clark. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science* 8, 5 (1997), 368–373. doi:10.1111/j.1467-9280.1997.tb00427.x

D. Schoenfeld. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68, 1 (1981), 316–319. doi:10.1093/biomet/68.1.316

Benjamin Schöne, Rebecca Sophia Sylvester, Elise Leila Radtke, and Thomas Gruber. 2021. Sustained inattentional blindness in virtual reality and under conventional laboratory conditions. *Virtual Reality* 25, 1 (2021), 209–216.

Christina L. Scrivener et al. 2020. Sensing and seeing associated with overlapping occipitoparietal activation. *NeuroImage* 222 (2020), 117208. doi:10.1016/j.neuroimage.2020.117208

Daniel J. Simons and Christopher F. Chabris. 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception* 28, 9 (1999), 1059–1074. doi:10.1068/p281059

Daniel J. Simons and Daniel T. Levin. 1998. Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review* 5, 4 (1998), 644–649. doi:10.3758/BF03208840

Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.

Sara Spotorno and Sylvane Faure. 2011. Change detection in complex scenes: Hemispheric contribution and the role of perceptual and semantic factors. *Perception* 40, 1 (2011), 5–22. doi:10.1068/p6630

Sara Spotorno, Benjamin W Tatler, and Sylvane Faure. 2013. Semantic consistency versus perceptual salience in visual scenes: Findings from change detection. *Acta psychologica* 142, 2 (2013), 168–176.

Frank Steinicke, Gerd Bruder, Klaus Hinrichs, and Pete Willemsen. 2011. Change blindness phenomena for virtual reality display systems. *IEEE Transactions on visualization and computer graphics* 17, 9 (2011), 1223–1233.

Jonathan A Stirk and Geoffrey Underwood. 2007. Low-level visual saliency does not predict change detection in natural scenes. *Journal of vision* 7, 10 (2007), 3–3.

Unity Technologies. 2026. Unity Engine: 2D and 3D. https://unity.com/products/unity-engine. Accessed: 03-12-2026.

Valve Corporation. 2023. Valve Index® Controllers. https://www.valvesoftware.com/en/index/controllers. Accessed: 03-12-2026.

Varjo Technologies. 2026. VR/XR Headset for Training and Simulation – Varjo XR-4 Series. https://varjo.com/products/xr-4. Accessed: 03-11-2026.

Frances C. Volkmann, Lorrin A. Riggs, and Robert K. Moore. 1980. Eyeblinks and visual suppression. *Science* 207, 4433 (1980), 900–902. doi:10.1126/science.7355270

A. Wald. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54, 3 (1943), 426–482.

doi:10.2307/1990256

Pepper Williams and Daniel J Simons. 2000. Detecting changes in novel, complex three-dimensional objects. *Visual Cognition* 7, 1-3 (2000), 297–322.

Jeremy M. Wolfe. 1994. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review* 1, 2 (1994), 202–238. doi:10.3758/BF03200774

Katherine Wood and Daniel J Simons. 2017a. Reconciling change blindness with long-term memory for objects. *Attention, Perception, & Psychophysics* 79, 2 (2017), 438–448.

Katherine Wood and Daniel J Simons. 2017b. Selective attention in inattentional blindness: Selection is specific but suppression is not. *Collabra: Psychology* 3, 1 (2017), 19.

Lu-Chun Yeh and Marius V Peelen. 2022. The time course of categorical and perceptual similarity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance* 48, 10 (2022), 1069.

Yei-Yu Yeh and Cheng-Ta Yang. 2008. Object memory and change detection: Dissociation as a function of visual and conceptual similarity. *Acta Psychologica* 127, 1 (2008), 114–128. doi:10.1016/j.actpsy.2007.04.003

Gregory J Zelinsky. 2003. Detecting changes between real-world objects using spatiochromatic filters. *Psychonomic Bulletin & Review* 10, 3 (2003), 533–555.

A Appendix

Here we provide further details regarding salience calculations, model specifications and analyses, and estimates from exploratory models.

A.1 Salience Calculations

Table 1: Salience validation results for manipulated objects. LM = low-level manipulated object mean salience; SM = semantic manipulated object mean salience; Δ (SM-LM) = within-class difference; Δ Global = deviation of class mean from the global manipulated object average (0.207).

Class	LM	SM	Class Mean	Δ (SM-LM)	Δ Global
Violin	0.255	0.161	0.208	-0.095	0.001
Apple	0.202	0.212	0.207	0.010	-0.000
Toothbrush	0.170	0.160	0.165	-0.010	-0.042
Iguana	0.220	0.278	0.249	0.058	0.042

A.2 Models and Analyses

Weighted average blink rate was computed as

$$\text{BlinkRate} = \frac{B_1 + B_2}{T_1 + T_2}$$

where B denotes blink counts and T denotes task duration for the two practice tasks.

The Cox proportional hazards model used in the confirmatory analyses took the form

$$h(t \mid X) = h_0(t) \exp(\beta_1 \text{Group} + \beta_2 \text{BlinkRate})$$

where Group denotes similarity condition and BlinkRate represents baseline blinking behavior.

Schoenfeld residuals were calculated as

$$r_j(t_i) = x_{ij} - \hat{E}[X_j | t_i]$$

and formally tested under

$$H_0 : \frac{d\beta_j(t)}{dt} = 0.$$

A.3 Exploratory Model Estimates

Table 2: Item-level Cox model predicting change detection time.

Predictor	HR	95% CI	p
Semantic vs Low-level	5.57	[1.42, 21.78]	**0.014**
Trumpet vs Vase	1.76	[0.60, 5.18]	0.307
Holiday orn. vs Vase	3.02	[0.74, 12.27]	0.123
Banana vs Vase	0.49	[0.16, 1.48]	0.203
Paintbrush vs Vase	1.49	[0.35, 6.32]	0.591
Toothpaste vs Vase	0.41	[0.12, 1.37]	0.148
Cactus vs Vase	1.30	[0.28, 6.04]	0.738
Blink rate	0.28	[0.02, 4.68]	0.377

Table 3: Item-only Cox model without similarity condition.

Predictor	HR	95% CI	p
Trumpet vs Vase	9.78	[2.40, 39.87]	**0.001**
Holiday orn. vs Vase	3.02	[0.74, 12.27]	0.123
Banana vs Vase	2.70	[0.67, 10.87]	0.162
Paintbrush vs Vase	1.49	[0.35, 6.32]	0.591
Toothpaste vs Vase	2.30	[0.54, 9.82]	0.262
Cactus vs Vase	1.30	[0.28, 6.04]	0.738
Tortoise vs Vase	5.57	[1.42, 21.78]	**0.014**
Blink rate	0.28	[0.02, 4.68]	0.377