frontiers | Frontiers in Cognition

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fcogn.2023.1125145

OPEN ACCESS

EDITED BY
Stefano Lasaponara,
Faculty of Medicine and Psychology, Sapienza University of Rome, Italy

REVIEWED BY
Davide Nardo,
School of Clinical Medicine, University of Cambridge, United Kingdom
Marco Costanzi,
Libera Università Maria SS. Assunta, Italy

*CORRESPONDENCE
Javier Ortiz-Tudela
✉ ortiztudela@psych.uni-frankfurt.de

SPECIALTY SECTION
This article was submitted to Attention, a section of the journal Frontiers in Cognition

RECEIVED 15 December 2022
ACCEPTED 20 January 2023
PUBLISHED 28 February 2023

CITATION
Ortiz-Tudela J, Jiménez L and Lupiáñez J (2023) Scene-object semantic incongruity across stages of processing: From detection to identification and episodic encoding. *Front. Cognit.* 2:1125145. doi: 10.3389/fcogn.2023.1125145

Scene-object semantic incongruity across stages of processing: From detection to identification and episodic encoding

Javier Ortiz-Tudela[1,2]*, Luis Jiménez[3] and Juan Lupiáñez[1]

[1]Centro de Investigación Mente, Cerebro y Comportamiento (CIMCYC), Universidad de Granada, Granada, Spain, [2]Department of Psychology, Goethe University Frankfurt, Frankfurt Am Main, Germany, [3]Facultad de Psicología, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

Visual processes are assumed to be affected by scene-object semantics throughout the stream of processing, from the earliest processes of conscious object detection to the later stages of object identification and memory encoding. However, very few studies have jointly explored these processes in a unified setting. In this study, we build upon a change detection task to assess the influence of semantic congruity between scenes and objects across three processing stages, as indexed through measures of conscious detection, object identification, and delayed recognition. Across four experiments, we show that semantically incongruent targets are easier to detect than their congruent counterparts, but that the latter are better identified and recognized in a surprise memory test. In addition, we used eye-tracking measures, in conjunction with these three behavioral indexes, to further understand the locus of the advantage observed in each case. The results indicate that (i) competition with other congruent objects modulates the effects of congruity on target detection, but it does not affect identification nor recognition memory, (ii) the *detection cost* of scene-congruent targets is mediated by earlier fixations on incongruent targets, (iii) neither fixation times, dwell times, nor pupil dilatation are related to the effects obtained in identification and recognition; and (iv) even though congruent targets are both better identified and remembered, the *recognition benefit* does not depend on the identification demands. The transversal approach taken in this study represents a challenging but exciting perspective that holds the potential to build bridges over the seemingly different but related fields of conscious detection, semantic identification, and episodic memory.

KEYWORDS

semantic congruity, detection, identification, episodic encoding, recognition memory, stream of processing

Introduction

The amount of information with which our cognitive system is continuously faced is overwhelming. Of all the information that gets through our senses, only a small portion reaches a state in which we actually become aware of it. In turn, an even smaller fraction of that information is stored into memory and can eventually be remembered. Understanding what sorts of transformations that information undergoes across the stream of processing is thus a very important, but often neglected, aspect of the study of human cognition. Analyzing the course of the same information across different processing stages can provide new insights into the underlying mechanisms and processes at play throughout this course.

One of the key modulators at several stages of that multiple-filter operation is semantic information. For instance, previous knowledge about the world may bias the information that gets access to our conscious awareness, by anticipating the most likely stimuli given a set of priors (Rao and Ballard, 1999; Summerfield et al., 2006). Similarly, the semantic features of a scene can also determine which objects will actually be attended, even beyond the biases imposed by other lower-level perceptual features (Peelen and Kastner, 2014; Santangelo et al., 2015; Henderson and Hayes, 2017). Moreover, previous knowledge can help us to interpret and give meaning to seemingly meaningless stimuli (Mooney, 1957; Gorlin et al., 2012) and it can even adjust which information gets stored into memory and which does not (Henson and Gagnepain, 2010; Van Kesteren et al., 2012). In this study, we will use prior semantic knowledge of real-world visual scenes to jointly characterize three key stages in the processing of information: detection, identification, and episodic encoding.

Object detection

The unspecific report of the detection of a visual stimulus can be studied by means of many different paradigms. Most of them require participants to press a given key in response to the detection of a target stimulus independently of features such as its location, color, or identity. These seemingly unimportant features are often used as independent variables that either speed up or slow down detection times and can even facilitate or impair detection accuracy, leading to positive and negative effects like priming (Kroll and Potter, 1984), change blindness (Simons and Rensink, 2005) or inhibition of return (Posner et al., 1985), which are often interpreted as the result of a *detection cost* (Lupiáñez et al., 2013).

The semantic features of an image are also thought to bias detection responses during scene processing. Hollingworth and Henderson (2000) showed that the detection of a changing target improves when the to-be-detected object is embedded in a semantically incongruent context (Hollingworth and Henderson, 2000). Moreover, LaPointe and Milliken (2016) showed that incongruent objects had shorter first *fixation latencies.* This variable represents the lag of time from the moment the trial starts until the object is fixated for the first time and it has been often used as a measure of pre-attentional processes influencing attentional capture.

Object identification

Even though detection and identification of an object appear to be two seamless stages of perception, LaPointe et al. (2013) showed that semantic information can be used to dissociate both processes, as they were affected in opposite ways by semantic congruity (LaPointe et al., 2013). They used a change detection task in which the identity of the to-be-detected object either matched or mismatched the gist of the surrounding scene, and they asked participants to detect and subsequently identify the changing object. Their results replicated the previously reported congruity *detection cost*, but they showed a simultaneous benefit for congruent targets on the identification task. This congruity *identification benefit* thus refers to facilitated access to the semantic features of a target when it is presented in the context of other semantically related objects. This finding is in line with research on prior knowledge and expectations, which shows that object identification is improved when the visual input matches what the observer is expecting (Eger et al., 2007; Esterman and Yantis, 2010). Importantly, at least one previous study has looked at *on-target dwell time* (i.e., the sum of time spent fixating the target region) as a proxy for total target processing time in the contest of the *identification benefit* (LaPointe and Milliken, 2016). This study found no differences in dwell time between congruent and incongruent objects thus supporting the notion that this benefit does not reflect merely increased processing time.

Long-term storage and retrieval

Both, the *detection cost* and the *identification benefit* are immediate measurable consequences of embedding an object in a semantic context. However, surrounding semantic information can have also long-term consequences by impacting how the object is encoded into memory. As a consequence, the ability to distinguish a previously seen object from one never seen before (i.e., a recognition memory), will be modulated by the semantic context in which the object was presented. For instance, a congruent background can facilitate later access to a given object by easing its integration into existing schemas (Gronau and Shachar, 2015; Kaiser et al., 2015; Ortiz-Tudela et al., 2016; Brod and Shing, 2019; Wynn et al., 2019). Conversely, an incongruent background can also render memorable a given object by signaling it as salient or unexpected (Henson and Gagnepain, 2010; Van Kesteren et al., 2012). This seemingly incompatible finding is currently the focus of active research (Ortiz-Tudela et al., 2018b; Greve et al., 2019; Quent et al., 2021) and the consideration of the role of the adjacent process can provide important insights into the debate.

Previous research using gaze measures to study memory phenomena (Võ et al., 2008; Otero et al., 2011; Kafkas and Montaldi, 2012) has largely relied on *pupil dilation* which is the variation in the diameter of the pupil, and has often been used as a measure of cognitive effort devoted to the task. These studies consistently observe larger pupil dilation at retrieval for successfully remembered items. This effect is generally assumed to be a consequence of either increased mental effort that leads to better memory or of a subjective feeling of familiarity with the correctly identified items; either of these interpretations must be ascribed to processes taking place at the moment of retrieval. In our study, we placed our focus on semantic congruency effects during encoding (i.e., during visual processing of the stimuli) and how this relates to eventual memory performance.

The present study

Because much of the abovementioned research has focused exclusively on one or a subset of these three different stages, it remains largely unknown whether they rely on independent mechanisms. We argue that a simultaneous study of these different phenomena might provide a more realistic picture of the hierarchical nature of this continuous stream that would have the potential to reveal existing interactions and dependencies between them. Thus, in this study, we intend to better explore how the semantic relatedness between an object and its scene context may affect different stages in the perceptual processing of the object, and ultimately determine its encoding in memory. We designed four experiments with a change detection task in which we manipulated the semantic congruity of the targets with the gist of the scenes in which they were embedded and assessed which of these changing targets were more efficiently detected, identified, and recognized. In Experiments 1A and 1B we compared two presentation procedures and two types of scenes differing in the number of objects presented on the scenes by assessing the indices of detection, identification, and recognition. In Experiment 2, we removed the identification task and replicated the setup for detection and recognition, to assess whether the effects obtained in recognition were independent of explicit identification demands. Finally, Experiment 3 typified the gaze patterns associated with each of these three processes, analyzing separately the amount of time elapsed from the start of the trial to the first fixation on the target, the amount of time spent fixating the target region, and the average pupil dilatation measured on each trial. Because each of these measures has been taken to reflect different cognitive functions such as attentional capture (first fixation), total processing time (dwell time) or cognitive effort (pupil dilatation), we surmise that this study might reveal important information on the impact of semantic relatedness at each of these three processing stages and illustrates a potentially useful approach to the study of how semantic congruity may affect the full stream of processing.

Experiment 1

Whether the semantic effects described in the introduction (i.e., *detection cost*, *identification benefit*, and *recognition benefit*) are a consequence of priming or of object competition mechanisms is still unsolved. Stein and Peelen (2015) recreated a situation in which detection took place with no competition from other objects (i.e., the target was presented alone in the context of visual noise). Their study included a cue which could either match or mismatch the category of an object suppressed under CFS conditions (Tsuchiya and Koch, 2005). With this paradigm, participants benefited from congruent cues. In these conditions, and in the absence of potential competitors, mechanisms such as priming (Kroll and Potter, 1984) or top-down inferences over ambiguous stimuli (Bar, 2003; Gorlin et al., 2012) are most likely responsible for guiding behavior. In contrast, in the conditions imposed by change detection paradigms, can be considered as the opposite situation: responding to cluttered images heavily relies on object competition since the participants' goal is to selectively detect a changing target among many distracters. Under this conditions, the presence of many different but semantically related objects hinders the detection of the specific (changing) target (Hollingworth and Henderson, 2000; LaPointe et al., 2013; LaPointe and Milliken, 2016; Ortiz-Tudela et al., 2016, 2018a). In Experiment 1 of the present study, we attempted at recreating an intermediate situation, using LaPointe et al.'s task, but reducing the presence of distracters, to prevent competition. We presented participants with two types of natural scenes: cluttered scenes, in which the images included many non-target objects together with the target one, and sparse scenes, in which only the target object was presented against a background image.

If semantic effects take place as a consequence of priming-like or top-down inferential mechanisms, they ought to be present in both types of scenes, since the propagation of semantic properties from the scenes to the individual objects can equally occur in both conditions. Conversely, if the aforementioned effects arise as a consequence of stimulus competition, they should appear selectively in cluttered trials, where there are many objects that compete with each other. More specifically: we hypothesized that, in the present experiment, the *detection cost* ought to be present only for cluttered trials. In opposition, the *identification benefit*, which arguably relies on spreading activation from the context image to the object (Palmer, 1975; Davenport and Potter, 2004; Eger et al., 2007), ought to be present in both cluttered and sparse trial types. Lastly, given that the *recognition benefit* has been previously hypothesized to be driven by schema-integration processes (Ortiz-Tudela et al., 2016), and those rely solely on the availability of contextual schema and not on the presence of other objects, we hypothesized that the *recognition benefit* should also be observed for both stimulus types.

Finally, because including qualitatively different sets of images in a task might entail not only the differential processing of those images but an overall change in participants' task set and strategies, we conducted two separate but complementary experiments. In Experiment 1A, the order of presentation of the two stimulus types was randomized so that it was impossible to anticipate the nature of the upcoming trial and to be specifically prepared for it in advance. In Experiment 1B, stimuli from the same set of images (i.e., cluttered vs. sparse) were grouped into blocks, so that all the trials from one group were presented together; this blocked setup allows participants to adjust their strategy to the corresponding block so that the optimal task set can be prepared before the onset of every trial.

Material and methods

Participants

Twenty students (18 female; mean age: 21.84; SD: 6.30) from the Universidad de Granada participated in Experiment 1A; another 20 students (18 female; mean age: 20.45; SD: 5.65), extracted from the same pool, participated in Experiment 1B. All of them volunteered in exchange for course credit and signed an informed consent approved by the local ethics committee. The sample size was determined based on previous studies using a similar paradigm (LaPointe et al., 2013; Ortiz-Tudela et al., 2016, 2018a) and sensitivity analysis was conducted to estimate the smallest detectable effect size. This analysis revealed that, with the available sample size, we would be able to detect effect sizes of at least $d = 0.58$, with 80% power and an alpha level of 0.05 (one-tailed matched samples t-test). All experiments in this paper, which are part of a larger research project approved by the Universidad de Granada Ethical Committee (175/CEIH/2017), were conducted according to the ethical standards of the 1964 Declaration of Helsinki (last update: Seoul, 2008).

Stimuli

All of the stimuli included in this and subsequent experiments in this study were either borrowed from previous publications (LaPointe et al., 2013; LaPointe and Milliken, 2016; Ortiz-Tudela et al., 2016, 2018a) or specifically built to match the needs of our experiment (see also below). All the stimuli consisted of scene-object combinations and both, scenes and objects, depicted real-world content (e.g., the image of a forest with a deer as an object). All the scene images were 850 × 565 pixels and the original object images were 500 × 500 pixels in size. All the objects were digitally resized and embedded in the scenes using Adobe Photoshop CS6. Each object was paired with two images, one congruent and one incongruent (Supplementary Table S1). Although the size of the objects was adjusted for each individual scene, an attempt was made to keep the size relatively similar across the two versions. We provide probability maps of the area covered by the objects in both congruency conditions as well as a statistical analysis of the differences in size between conditions and a correlation of each object's size across conditions (Supplementary Figure S1). The analysis confirmed the lack of differences in object size between conditions (BF01 = 4.327) and a strong within-object correlation of the small differences (Pearson's $R = 0.846$, $p < 0.001$). In addition, we also computed pixel-wise saliency (Supplementary Figure S2) and luminance (Supplementary Figure S3) metrics and run a Bayesian t-test between congruency conditions. The results also supported the lack of differences in either of the measures (BF01 = 5.968 and BF01 = 7.951, respectively).

Procedure

Each participant completed three sequential phases: the first one consisted of a change detection task. This phase was followed by 10 min of mathematical operations that served as a distracter task. Finally, memory of the target objects from the change detection task was assessed *via* a surprise recognition test. The duration of the entire session was ∼45 min.

The overall structure of the session was identical for Experiment 1A and 1B with the sole exception of the order of presentation of the cluttered vs. sparse trial types of the change detection task (i.e., randomized for Experiment 1A and blocked for Experiment 1B). In Experiment 1B randomization was applied within each block so that the sequence of trials within that block was different for each participant; the order of the blocks was counterbalanced across participants.

Change detection task

Each trial consisted of a rapid alternation of two versions of the same image, each displayed for 250 ms. The two versions represented scenes which were identical to each other except for the presence or absence of a key object. Participants were required to press the space bar on a QWERTY keyboard as soon as they noticed any detail that was different between the two versions of the scene. To prevent the changing object from popping out, an intervening blank screen was displayed for 250 ms between the two presentations. This intervening screen rendered the standard flickering appearance of the paradigm (Rensink et al., 1997). Critically, we manipulated the congruity between the to-be-detected object and the background scene. On half of the trials, the target identity matched the gist of the scene (i.e., congruent trials) and on the other half, it corresponded to an object that was not expected or frequent in that context (i.e., incongruent trials). After the detection response, or after a maximum of nine alternation cycles, the sequence stopped and a new screen prompted participants to identify the changing object with a few words (e.g., black dog) or by locating it on the screen (e.g., bottom-left) if identification was not possible (Figure 1). To assure participants' engagement in the task, 10% of no-change trials were included (i.e., catch trials). Participants were not informed of the presence of these no-change trials since previous studies have shown that being aware of the presence of those trials can change participants' response bias (Ortiz-Tudela et al., 2016). A total of 90 object-image combinations were used.

More importantly for our purposes, we included two sets of trials. The cluttered set was built so that the target object (i.e., the changing one) was one among many other presented objects. Conversely, in the sparse set scenes, the target object was presented in isolation against an open background image (Figure 2). For the cluttered set complex natural scenes were selected such as a busy city street, a park with children and trees or a big city skyline; for the sparse set, rather *empty* scenes were selected such as a wide prairie, a desert, or an open sky. Cluttered and sparse set scene trials were intermixed within the same block of trials in Experiment 1A and in different blocks of trials in Experiment 1B.

Distracter task

Participants completed paper and pencil math operations for a maximum time of 10 min. None of the participants completed the entire set of proposed operations. The exact operations used are available at https://github.com/ortiztud/three_indices.

Recognition memory test

All the target objects from the change detection task, together with 90 new objects, were used in the memory test. Each object was presented alone (i.e., stripped from any scene context) at the center of the screen and covering ∼10° of visual angle. Participants performed an old vs. new judgment without any time restriction. Correct responses to old objects were coded as hits and incorrect responses to old objects were coded as False Alarms (FAs).

Results

Experiment 1A

Participants ($N = 5$) who reported a change in more than 40% of catch trials were excluded from the analyses. The three dependent variables of interest were analyzed separately using

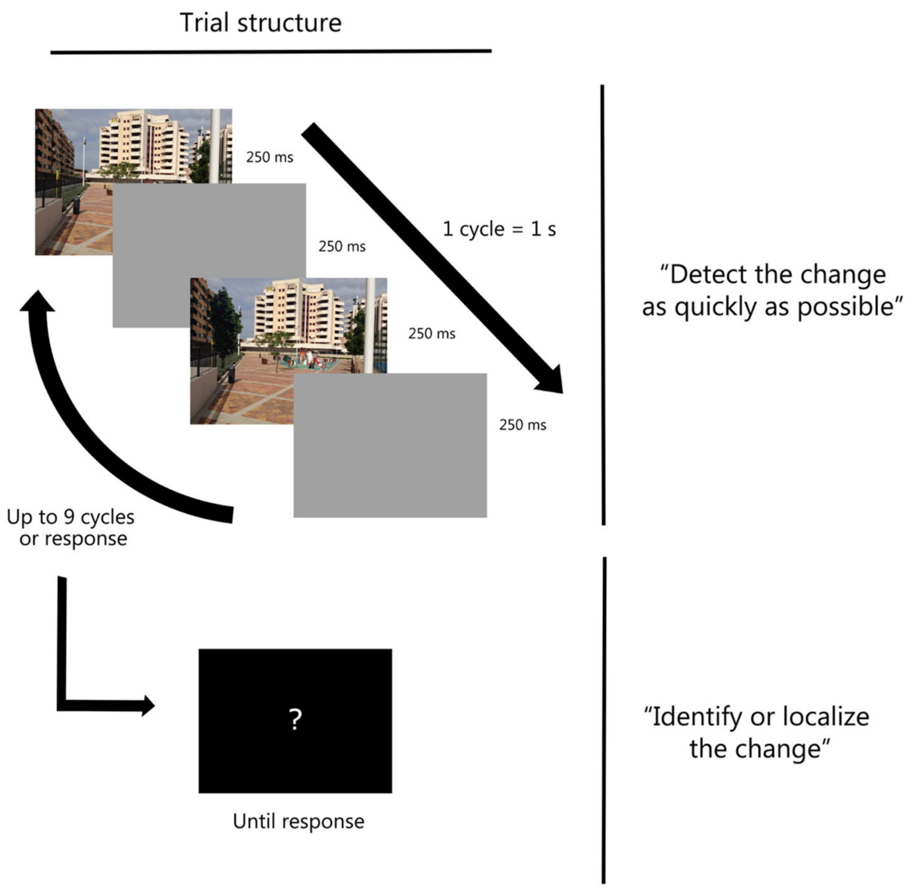
Trial structure
250 ms
250 ms
250 ms
250 ms
1 cycle = 1 s
"Detect the change as quickly as possible"
Up to 9 cycles or response
?
Until response
"Identify or localize the change"

FIGURE 1
Trial structure for the change detection task in Experiments 1A, 1B, and 3. Participants sequentially performed a detection task followed by an identification task (see Ortiz-Tudela et al., 2016, for a simiar procedure).

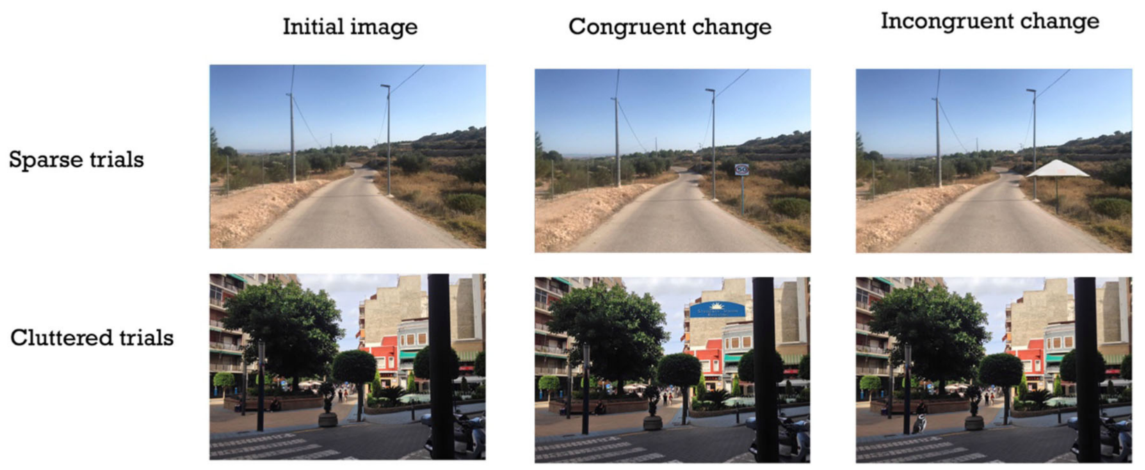

FIGURE 2
Example of stimuli used in Experiment 1A and 1B. Scenes in the cluttered set were taken from Ortiz-Tudela et al. (2016); for the sparse set, scenes with none or just a few non-target objects were selected.

2 × 2 repeated measures ANOVAs with scene-object congruity (congruent vs. incongruent) and trial type (cluttered vs. sparse) as within-subjects factors.

Detection

Performance on the detection task was evaluated by combining detection times with the proportion of correct responses in an overall detection index (proportion of correct responses/detection times; Ortiz-Tudela et al., 2018b). The analysis of the detection index revealed a significant trial type by congruity interaction, $F_{(1,14)} = 5.954$, $p = 0.029$, $\eta p^2 = 0.40$, showing that on the cluttered set responding to congruent targets was less efficient than responding to incongruent targets, $F_{(1,14)} = -3.41$, $p = 0.004$, $\eta p^2 = 0.43$, but there were no differences in the sparse set, $F_{(1,14)} = -1.32$, $p = 0.208$, $\eta p^2 = 0.01$.

Identification

Only correctly detected objects for each participant were included in the following analyses. The results of the analysis of the proportion of correctly identified objects appropriately replicated previous findings of higher identification scores for congruent objects, $F_{(1,14)} = 10.981$, $p = 0.005$, $\eta p^2 = 0.47$. Importantly, the trial type by scene-object congruity interaction was not significant in this measure, $F < 1$, suggesting that the *identification benefit* was present in both trial types, $F_{(1,14)} = 2.49$, $p = 0.026$, $\eta p^2 = 0.36$ and $F_{(1,14)} = 3.24$, $p = 0.006$, $\eta p^2 = 0.43$ for cluttered and sparse respectively.

Recognition

Trials that were correctly detected and correctly identified were passed along to the recognition analyses. Overall *d'* was 1.27 and beta 1.84. Since it was not possible to assess independent FA rates for congruent and incongruent trials, overall hit rates were used as a measure of memory performance. The analysis did not show a significant effect of trial type, $F_{(1,14)} = 3.082$, $p = 0.101$, $\eta p^2 = 0.15$, even though we measured numerically higher recognition scores for objects in the sparse set (0.78) compared to those in the cluttered set (0.75). The numerical pattern also showed higher memory rates for congruent than for incongruent objects, at least for the cluttered scenes (see Table 1), but neither this difference nor the two-way congruity x trial type interaction were close to statistical significance, *F*s < 1.

Experiment 1B

The same approach as in Experiment 1A was adopted for the analyses of Experiment 1B. Data from three participants were excluded from the analysis for poor performance in the detection task.

Detection

The analysis of detection efficiency replicated those of Experiment 1A. The trial type by congruity interaction was close to significance for the detection index, $F_{(1,16)} = 3.977$, $p = 0.063$, $\eta p^2 = 0.20$. In other words, again more efficient responses were made on incongruent than on congruent trials on cluttered trials, $F_{(1,16)} = -3.89$, $p = 0.001$, $\eta p^2 = 0.43$, but no differences between congruent and incongruent target objects were obtained on sparse trials, both $F_{(1,16)} = -1.56$, $p = 0.139$, $\eta p^2 = 0.13$.

Identification

The pattern of the identification scores in Experiment 1B mimicked that of Experiment 1A. Consistent with an *identification benefit* effect, congruent target objects were better identified than incongruent objects, $F_{(1,16)} = 4.746$, $p = 0.045$, $\eta p^2 = 0.21$. There was no indication of an effect of trial type, or of interaction between stimulus type and congruity, $F < 1$.

Recognition

The memory pattern in Experiment 1B also resembles that of Experiment 1A. Overall *d'* was 1.35 and overall beta was 2.05. The main effect of trial type was close to significance, $F_{(1,16)} = 4.92$, $p = 0.05$, $\eta p^2 = 0.23$, with better memory for objects in the sparse trials (0.75) than in the cluttered ones (0.66). No significant effect of congruity nor an interaction between trial type and congruity were observed, both *F*s < 1.

Discussion

The aim of Experiments 1A and 1B was to test whether the semantic congruity effects reported in the literature on the detection, identification and delayed recognition of objects could rely on different combinations of semantic facilitation and object competition. To that end, we used a change detection paradigm, that reliably produces the expected indexes [i.e., a *detection cost*, *identification benefit*, and *recognition benefit*; (LaPointe et al., 2013; Ortiz-Tudela et al., 2016, 2018a)], and we compared two stimulus sets which either included the target among many distracter objects or presented the target embedded in a sparse background. Because we reasoned that participants' responses can be affected by the adoption of a specific mindset evoked by surrounding trials, Experiment 1A and 1B also explored the potential effect induced by presenting these two types of contexts either in a random order (Experiment 1A) or grouped into blocks (Experiment 1B).

The results of the two experiments showed that while the *identification benefit* is present when using both cluttered and sparse stimuli, the *detection cost* is only found in the presence of stimulus competition. This result suggests that the *detection cost* arises only when there is a number of coactive stimuli competing for attentional resources, whereas the benefits found for identification seem to depend on semantic facilitation which might arise either from the activation of a group of semantically related objects or from the overall meaning of the background scene (Eger et al., 2007; Esterman and Yantis, 2010). The absence of differences in detecting congruent and incongruent trials in the sparse set is consistent with the idea that sparse scene contexts represent an intermediate situation between Stein and Peelen's minimalistic setup (in which better detection followed a category-matching cue) and the cluttered arrangement of LaPointe et al.'s (2013) paradigm (in which a *detection cost* was obtained).

Lastly, and surprisingly, we were not able to measure a statistically significant *recognition benefit* in spite of having arranged conditions very similar to those presented in Ortiz-Tudela et al. (2016). This unexpected result can be due to the inclusion of the sparse trials within the list of items to be retrieved at the memory test. Indeed, performance in any memory test is highly dependent not only on the processes taking place at encoding but also on those taking place during consolidation and retrieval and those can be affected by the amount and

TABLE 1 Mean RT and percentage of accurate detection responses (in parenthesis) for object detection, and percentage of accurate responses for object identification and delayed recognition, for each of the four experiments.

Experiment	Object detection				Object identification				Object recognition			
	Cluttered		Sparse		Cluttered		Sparse		Cluttered		Sparse	
	C	I	C	I	C	I	C	I	C	I	C	I
1A	2,784 (0.7)	2,757 (0.84)	2,049 (0.94)	2,072 (0.95)	0.87	0.72	0.83	0.73	0.70	0.65	0.74	0.74
1B	2,423 (0.75)	2,539 (0.87)	1,891 (0.97)	1,783 (0.96)	0.83	0.76	0.84	0.78	0.67	0.65	0.74	0.76
2	2,393 (0.83)	1,883 (0.94)	–	–	–	–	–	–	0.41	0.35	–	–
3	2,351 (0.87)	1,882 (0.96)	–	–	0.84	0.73	–	–	0.69	0.64	–	–

C, congruent; I, incongruent.

nature of the elements to be held in memory. Thus, before jumping to speculative conclusions about the *recognition benefit*, we decided to further explore and characterize the processes in another experiment.

The purpose of Experiment 2 was, therefore, two-fold. First, replicating the *recognition benefit* by attempting to measure it only with the standard cluttered scenes (as used in previous studies). Second, to further characterize this memory process by dissociating the *recognition benefit* from the identification task.

Experiment 2

LaPointe et al. (2013) used the *detection cost* and the *identification benefit* to claim that a clear dissociation could be behaviorally established between the detection and identification processes. Ortiz-Tudela et al.'s (2018a) later report of the *recognition benefit* followed the same direction as the *identification benefit*. However, the dual-task conditions arranged in this latter study, in which participants were required to detect and then identify the changing object, made it impossible to separate the influence of each of these two tasks in the memory results. Thus, it is possible that the *recognition benefit* arises as a consequence of the offline elaboration required to respond to the identification question and not to the mechanisms at play while the processing of the scene was carried out.

Therefore, in Experiment 2 we eliminated the identification question altogether to avoid any effects of this post-response task on later recognition. In addition, in order to ensure the *detection cost* and to improve the chances of measuring the *recognition benefit* effect, we used only cluttered scenes as in previous reports (LaPointe et al., 2013; Ortiz-Tudela et al., 2016, 2018a; Spaak et al., 2020).

Material and methods

Participants

To guarantee enough power to replicate previous results, we increased the sample size to 40 participants (37 female; mean age: 20.7; SD: 1.6). Participants were recruited from the Universidad de Granada in exchange for course credit. All of them signed informed consents approved by the local ethics committee. Four of them were unable to complete the entire experimental session and therefore were eliminated from the final sample.

Procedure

The overall procedure was the same as that described for Experiment 1 except for the following: to eliminate any potential interference from the sparse set on memory, we only used stimuli from the cluttered set. In addition, the identification question was removed, so that participants only had to perform the detection task that required them to respond as soon as they noticed any change during the flickering period. Following their response, the alternation of images stopped, and it was replaced by a fixation point, which indicated the beginning of the next trial after 1,000 ms. Proper task performance was assessed from accuracy in responding to both change and no-change trials. Each session had an approximate duration of 30 min.

Results

All participants reached the required threshold of 80% detection accuracy. We did not conduct a specific comparison between experiments concerning this result, but the absence of any participant below the threshold suggests that the inclusion of trials in which change detection was easier (i.e., the sparse trials) may have biased participants in Experiment 1 against reporting more subtle changes in cluttered trials.

Detection

Analysis of detection responses replicated the previous *detection cost*: more efficient responses for incongruent than congruent trials, $t_{(35)} = 8.05$, $p < 0.001$, Cohen's $d = 1.34$.

Recognition

Overall d' and beta were 0.80 and 2.35, respectively. Contrary to Experiments 1A and 1B, but replicating previously published results, analysis of hit rates showed the expected *recognition benefit*: targets from congruent scenes were remembered better than those from incongruent ones, $t_{(35)} = 2.21$, $p = 0.034$, Cohen's d =0.34.

Discussion

Experiment 2 aimed at replicating the previously reported *recognition benefit*, which was surprisingly absent in Experiments 1A and 1B, and at testing whether *recognition benefit* would appear in conditions in which identification was not required. According to our predictions, participants in this experiment showed the standard *recognition benefit*, even though overall recognition scores were considerably lower in this case (38%) as compared to that found in previous studies when an identification task was included after detection (e.g., 62%, in Ortiz-Tudela et al., 2016). This lower recognition rate is most likely due to the shallower processing of the stimulus (Craik and Lockhart, 1972) and can be taken as indirect proof of the reduced level of identification achieved during the change detection task. Although it is not possible to claim that identification processes were completely absent when the task was removed (since these are most likely automatic and dynamically engaged when any stimulus is processed), it is fair to assume that they were at least minimized in this experiment. More importantly, the fact that we observed the *recognition benefit* under these conditions rules out the possibility that this benefit is due to post-detection processes engaged during the identification task itself.

It is worth noting that, since participants did not carry out the identification task, it is not possible to further correct the detection responses to discard incorrectly detected trials (i.e., trials in which the participant reported a change but were not able to locate it). However, given that these are rare and that participants had high accuracy in discarding the no-change trials, we argue that they are unlikely to have a meaningful impact on the results. Nevertheless, future studies with a location report task (e.g., *via* mouse click) will be needed to further clarify this issue.

The results of Experiment 2 show that the *recognition benefit* arises independently of the identification task and that it is more clearly observed when object competition is present in the scenes (i.e., when targets are presented in a cluttered context). However, the nature of this memory effect is still puzzling. Indeed, congruent trials are generally displayed for longer periods of time (Hollingworth and Henderson, 2000; LaPointe et al., 2013; Ortiz-Tudela et al., 2016), but incongruent objects tend to attract eye gaze and to be looked at longer than congruent ones (Henderson et al., 1999; LaPointe and Milliken, 2016). The literature on mere exposure effect shows that extended exposure strengthens memory simply by virtue of longer processing time. What happens then under the circumstances in which the *recognition benefit* is produced? Are incongruent targets looked at longer than the congruent ones, but still they get more poorly remembered? Or might it be the case that, under these particular conditions, congruent targets produced longer dwelling times, and *recognition benefit* arises as a by-product of this extended exposure? Some *post-hoc* analyses have been conducted before as tentative attempts at controlling these and related issues (Krebs et al., 2015; Rosner et al., 2015; Ortiz-Tudela et al., 2016), but no study to date aimed at ruling out this possibility by directly measuring gaze patterns.

On Experiment 3 we used a high temporal resolution eye tracking system to record eye movements while the scenes were being processed to gain more insights into the mechanisms underlying each one of these processes.

Experiment 3

Building upon previous studies recording eye movements in change detection tasks (Henderson et al., 1999; Hollingworth et al., 2001; LaPointe and Milliken, 2016), in Experiment 3 we used the full paradigm (i.e., the three sequential tasks) to be able to characterize the gaze patterns associated respectively with the *detection cost*, the *identification benefit*, and the *recognition benefit*. This characterization provides an indirect measure of potential underlying mechanisms such as attentional capture, processing time or cognitive effort (see Methods section for more on this).

Material and methods

Participants

As this experiment made use of the standard procedure for obtaining the three semantic indexes, the minimum sample size usually required to measure them (LaPointe et al., 2013; Ortiz-Tudela et al., 2016, 2018a; see Experiment 1 for analysis of power). Twenty students (15 female; mean age: 20.65; SD: 3.8) from the Universidad de Granada volunteered to take part in the study. All of them signed informed consent according to the local ethics committee.

Design

The overall structure of the experiment resembles that of the previous studies in the present paper: a change detection + identification task was followed by a distracter task, and then by a surprising recognition test. As in Experiment 2, we only used cluttered scenes as stimuli. The duration of the session was 1 h approximately.

Eye movement recording

A high sampling frequency (250 Hz) SMI [SensoMotoric Instruments (SMI), 1991] system was used to record participants' eye movements during the change detection task. A maximum of 1 degree of tracking error was accepted for every participant during a calibration phase at the beginning of the experiment. Participants' heads were placed on a chinrest 60 cm away from the monitor to avoid unwanted movements and to allow for a comfortable posture. Regions of interest were defined for each scene as a rectangular area encapsulating the object. This area extended vertically from the highest to the lowest pixel in the object image and horizontally from the left-most to the right-most pixel. All of the eye-tracking measures reported here were pulled from the built-in SMI's software [BeGaze; SensoMotoric Instruments (SMI), 1991]. BeGaze's default method of event detection for high-speed eye tracking data uses saccades (computed with a velocity-based algorithm) as primary event; fixations and blinks are, in turn, derived from saccades. Blinks are defined as saccades with a pupil diameter of 0 mm [see SensoMotoric Instruments (SMI),

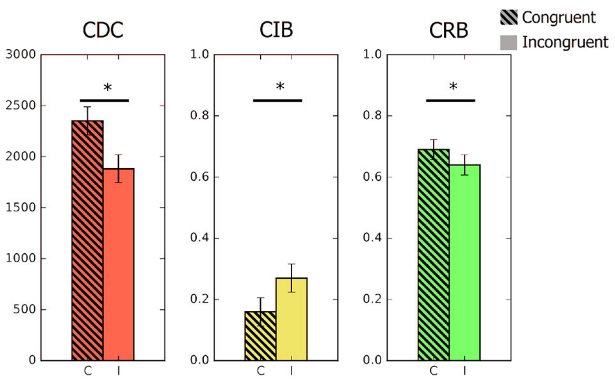

FIGURE 3
Graphic representation of the three semantic indexes in Experiment 3. From left to right: CDC, congruity *detection cost* (efficiency index); CIB, congruity *identification benefit* (errors); CRB, congruity *recognition benefit* (hits). Asterisks denote significant differences between congruent and incongruent trials ($p < 0.05$).

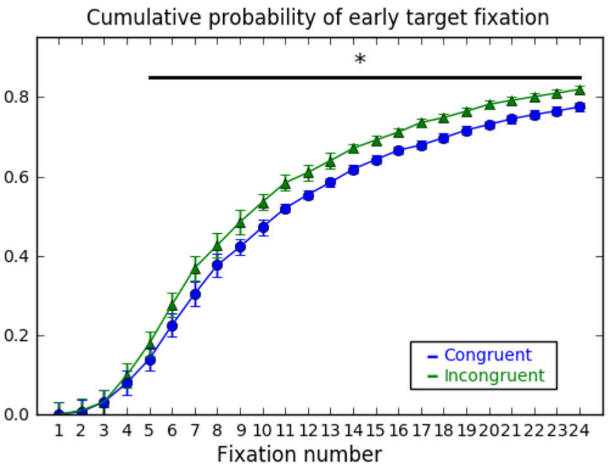

FIGURE 4
Cumulative probability of early target fixation for congruent and incongruent trials. Probability of having fixated the target as a function of ordinal fixation number. $^*p < 0.05$.

1991 for more details on how these measures are computed]. Post-processing of saccades, fixations and pupil diameter was performed with custom MATLAB scripts.

Results

Two participants were removed from the analyses due to an eye-tracking malfunction which caused the loss of all the session data.

The change detection analyses adequately reproduced the expected results. Namely, we obtained the usual *detection cost* of more efficient detections on incongruent trials than on congruent ones, $t_{(17)} = 9.34$, $p < 0.001$, Cohen's $d = 2.20$. We were also able to measure the *identification benefit*, $t_{(17)} = 4.82$, $p < 0.001$, Cohen's $d = 1.14$, and the *recognition benefit*, $t_{(17)} = 3.06$, $p = 0.007$, Cohen's $d = 0.72$ (see Table 1 and Figure 3 for a graphic representation of the three indexes).

Eye-movement measures

We selected three key variables of interest obtained from the raw pattern of gaze data. Namely, *target first fixation latency, on-target dwell time, and pupil dilation*. To analyse these three variables of interest during the study phase we took a step-by-step approach. First, we analyzed the three measures of interest (i.e., target first fixation latency, target dwell time, and pupil dilation) for all correctly detected trials, separately for congruent and incongruent targets. Then, we performed the same analyses for the subset of these correctly detected trials that corresponded to correctly identified targets. Finally, we examined the same variables of interest for the subset of these correctly identified trials that were also correctly remembered. Together with these three variables of interest, we also report here other secondary variables often used by researchers in similar areas for the sake of convergence of results. It is important to note that for all these three analyses, including those involving the memory results, the eye movements of interest were those recorded at the time of encoding. During the analysis stage, the eye-tracking measures were retroactively coded as a function of memory performance. This procedure is usually referred to as the "subsequent memory approach" and it is very common in the neuroimaging literature aiming at exploring encoding processes that have either a successful (i.e., eventually remembered) or unsuccessful (i.e., eventually forgotten) outcome (Brewer et al., 1998; Paller and Wagner, 2002).

Detection

Target first fixation latency

We analyzed target first fixation latency for successfully detected congruent and incongruent objects. Mimicking RT, fixation latency was shorter for incongruent objects (1,419 ms) than for congruent ones (1,755 ms), $t_{(17)} = 5.13$, $p < 0.001$, Cohen's $d = 1.21$.

To further explore this result and its relation with the *detection cost*, we run a Pearson correlation between the congruity effect on response times and on-target first fixation latency and found a high correlation of $r = 0.69$, $t_{(16)} = 3.76$, $p = 0.002$, Cohen's $d = 0.89$., suggesting that shorter first fixation latencies are at the basis of the *detection cost*. Besides, we also assessed the probability of fixating the target as a function of the number of fixations, which has also been used as a measure of early attention attraction (LaPointe and Milliken, 2016). The cumulative probability of fixating the target object within the first four fixations on the scene reached 10% and was equally probable for congruent and incongruent targets, $t < 1$. However, from the 5th fixation onwards the probability of fixating the target object started growing significantly faster for incongruent than for congruent targets (see Figure 4).

Target dwell time (ms)

Target dwell time was assessed by adding the total amount of fixation and saccades time (i.e., the total time spent exploring in the target area) for congruent and incongruent trials. No differences

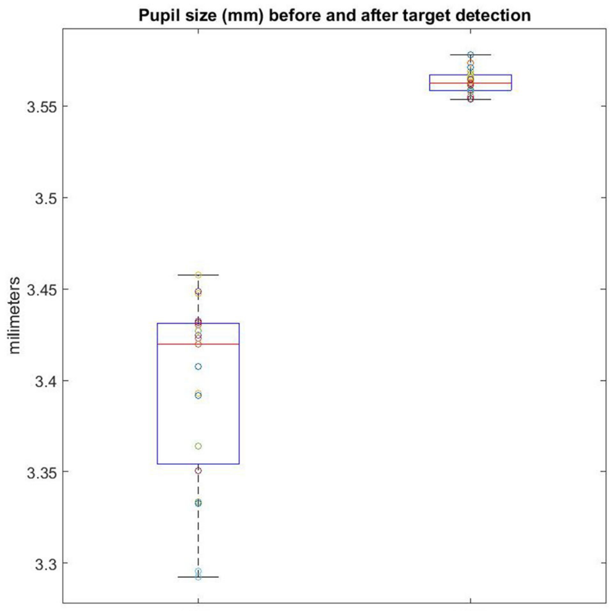

FIGURE 5
Pupil size (mm) before and after conscious reports of target detection. Each colored dot represents an individual participant.

were found in the time spent looking at the target for congruent (1,899 ms) or incongruent objects (1,867 ms), $t < 1$.

Pupil dilation

Pupil size has been used as a proxy for cognitive effort (Kahneman and Beatty, 1966; Peavler, 1974) with larger pupil diameter for effortful responses. Our results revealed larger average pupil dilation for those trials in which the target was detected (3.47 mm) compared to trials in which participants were unable to detect any change (2.48 mm), $t_{(17)} = 6.13$, $p < 0.001$, Cohen's d =1.44. However, in order to ascertain whether increased pupil dilation is either cause or consequence of detecting the change, we compared mean pupil dilation before and after the target was found. The comparison revealed that pupil dilation following a properly detected target was smaller before (3.39 mm) than after (3.56) target detection, $t_{(19)} = 10.83$, $p < 0.001$, Cohen's $d = 3.04$ (Figure 5). This result suggests that increased pupil dilation can be a direct consequence of consciously detecting the target object (Critchley et al., 2005; Braem et al., 2015; Wessel and Aron, 2017). In any case, among the detected trials, no differences were found between congruent and incongruent objects, $t < 1$.

Identification

Target first fixation latency

The latencies of the first fixations for those targets that were correctly identified and those that were not did not differ from one another, $t < 1$. However, among the identified ones we still measured shorter latencies for incongruent objects (1,406 ms) than for congruent objects (1,748 ms), $t_{(17)} = 4.66$, $p < 0.001$, Cohen's $d = 1.10$.

Target dwell time (ms)

Total target dwell time for identified (1,927 ms) and unidentified (1,869 ms), did not differ significantly, $t_{(17)} = 1.01$, $p = 0.324$, Cohen's $d = 0.24$. When we considered only correctly identified objects, target dwell time did not differ between congruent (1,929 ms) and incongruent objects (1,802 ms), $t < 1$.

Pupil dilation

No differences in average pupil size were found either between identified and unidentified, $t < 1$, or between congruent and incongruent objects when these were identified, $t_{(17)} = 1.20$, $p = 0.24$, Cohen's $d = 0.28$.

Recognition

Target first fixation latency

No differences in target first fixation latencies were found between remembered and forgotten items, $t < 1$, among those that were correctly detected and identified. As in the previous analyses, when we took into consideration exclusively correctly remembered items, we also found the same pattern of shorter latencies for incongruent targets (1,441 ms) than for congruent ones (1,750 ms), $t_{(17)} = 4.78$, $p < 0.001$, Cohen's $d = 1.13$.

Target dwell time (ms)

No differences in dwell time were found either between remembered (1,886 ms) and forgotten (1,874 ms) items, $t_{(17)} = 1.17$, $p = 0.258$, Cohen's $d = 0.28$, or between congruent (1,880 ms) and incongruent (1,822 ms) remembered items, $t_{(17)} = 1.26$, $p = 0.22$, Cohen's $d = 0.30$.

Pupil dilation

Average pupil size in the change detection phase did not vary between later remembered (3.36 mm) and later forgotten trials (3.36 mm), $t < 1$. Finally, no differences were found between congruent and incongruent later remembered objects, $t_{(17)} = 1.19$, $p = 0.25$, Cohen's $d = 0.28$.

Discussion

Experiment 3 aimed at characterizing the gaze patterns associated with the three effects that semantic congruity causes on detection, identification, and recognition, namely the *detection cost*, *identification benefit*, and *recognition benefit*. We showed that the first fixations on the target region were shorter for incongruent than for congruent trials and that this difference was highly correlated with the strength of the *detection cost*. Moreover, we also showed that the cumulative probability of fixating the target over the first n-fixations grew faster for incongruent than for congruent targets, which is also an indicator of early gaze (and arguably attention)-attraction toward the target region (LaPointe and Milliken, 2016). This result suggests that the shorter time required to fixate an incongruent target, due to the attraction provoked by semantic mismatch, is responsible for the *detection cost*. Thus, considering that it has been previously shown that scene viewing is guided by meaning maps even more strongly than by saliency maps (Henderson and Hayes, 2017), we argue that in our paradigm, attention is, by default, endogenously guided by meaning in search for a change; however, whenever a mismatch between the meaning map and the bottom-up semantic information of the incongruent object occurs attention is quickly attracted toward this object.

In addition, we showed that no differences in dwell time were found for congruent and incongruent targets, which reflects that longer processing time is not responsible for the *identification benefit*. Rather, eased access to the objects' semantic content is most likely due to pre-activation by semantically related information either from context or from neighboring objects (Davenport and Potter, 2004; Eger et al., 2007; Henderson and Hayes, 2017). In a similar fashion, none of the other explored gaze measures during the processing of the scene were related to the *recognition benefit* and therefore we argue that this asymmetry between the remembering of congruent and incongruent trials is most likely due to the integration of the object identity into a coherent schema established by the scene and not to longer processing time. Accordingly, it is often found in the schema integration literature that information that matches pre-existing structures is more easily retrieved than one that is hard to reconcile with prior information (Van Kesteren et al., 2012; Brod et al., 2013; Gronau and Shachar, 2015).

Combined analysis

Finally, to enable a robust examination of the presence of the three indices reported above, we collapsed all four experiments (for cluttered scenes only) and used generalized or linear mixed-effect models (with lme4 in R; Bates et al., 2015) to account for individual variability in either the participants or the stimuli. For all three indices we modeled participants and stimuli (and their interaction with congruity) as random effects and our manipulated variable (i.e., congruity) as a fixed effect. Model comparison was performed to test for the significance of the random effects included. The winning model was determined following a backwards model selection procedure where a maximal model (i.e., the model including all possible random effects and interactions) is contrasted for explanatory power against a reduced model (i.e., obtained by removing one random effect from the maximal model). The two models are compared using a likelihood ratio test with an alpha level of 0.2 (Matuschek et al., 2017). If a significant decrease in model fit is observed, the removed effect is accepted to significantly contribute to the overall fit of the model and it is kept in for further analysis; if no significant decrease is observed, the removed effect is discarded, and another reduction step is performed. In this new reduction step, the previously reduced model now becomes the complex one and a new reduced model is created by removing one effect from the complex model. This process is repeated until no further random effects can be removed without significantly decreasing model fit. Each reduction step is labeled with a sequential number in the reports (e.g., *reduced 2* is a model obtained in the second reduction step). The order of removal went from interactions to main effects and from slopes to intercepts. In case a given reduction step involved two models with the same number of parameters, two alternative models were created (labeled as "a" and "b" in the reports) and both were separately tested against the complex one. Once the winning model was obtained, its parameters were estimated with Maximum Likelihood and the significance of the fixed effects was tested through analysis of deviance using Wald chi-square. Only the winning models (and their corresponding statistical comparisons) are reported in this section but see online materials at https://github.com/ortiztud/three_indices for the full set of scripts to reproduce the results.

Results

Detection

Trial-level detection accuracy scores were submitted to the modeling procedure with the winning model including random slopes and intercepts for participants and stimuli (AIC complex: 2,408.3; BIC complex: 2,459.5; against reduced 1a, $\chi^2(2) = 127.992$, $p < 0.001$, AIC reduced 1a: 2,532.3, BIC reduced 1a: 2,570.7; against reduced 1b, $\chi^2(2) = 10.050$, $p < 0.001$, AIC reduced 1b: 2,414.3, BIC reduced 2b: 2,452.7). We observed a main effect of congruity, $\chi^2(1) = 15.787$, $p < 0.001$, with higher accuracy for incongruent than congruent trials, $z = 3.973$, $p < 0.001$, $\beta = 3.61$.

In a similar way, trial-level RTs for correct detections were submitted to the modeling procedure with the winning model including random intercepts for participants and random slopes only for stimuli, $\chi^2(2) = 519.420$, $p < 0.001$ (AIC complex: 14,112, reduced 1: 14,625; BIC complex: 14,156, reduced 1: 14,650). As expected, we observed a main effect of congruity, $\chi^2(1) = 45.910$, $p < 0.001$, with faster RTs for incongruent than congruent trials, $t_{(217)} = -6.776$, $p < 0.001$, $\beta =$ -0.561.

Identification

As Experiment 2 did not include the identification task, trial-level identification accuracy scores for Experiments 1a, 1b, and 3 were submitted to the modeling procedure with the winning model including random slopes and intercepts for participants and stimuli (AIC complex: 2,194.7; BIC complex: 2,241.4; against reduced 1a, $\chi^2(2) = 4.515$, $p < 0.001$, AIC reduced 1a: 2,195.2, BIC reduced 1a: 2,230.3; against reduced 1b, $\chi^2(2) = 5.172$, $p < 0.001$, AIC reduced 1b: 2,195.9, BIC reduced 1b: 2,230.9). We observed a main effect of congruity, $\chi^2(1) = 14.762$, $p < 0.001$, with higher accuracy for incongruent than congruent trials, z $= 3.842$, $p < 0.001$, $\beta = 0.737$.

Recognition

Mimicking previous steps, trial-level recognition accuracy scores were submitted to the modeling procedure with the winning model including random intercepts for participants and stimuli (AIC complex: 4,901.1; BIC complex: 4,926.2; against reduced 2a, $\chi^2(2) = 136.391$, $p < 0.001$, AIC reduced 2a: 5,035.5, BIC reduced 2a: 5,054.3; against reduced 2b, $\chi^2(2) = 125.261$, $p < 0.001$, AIC reduced 2b: 5,024.3, BIC reduced 2b: 5,043.2). Exploring the winning model revealed a main effect of congruity, $\chi^2(1) = 10.499$, $p = 0.002$, with higher accuracy for congruent than incongruent trials, $z = -3.137$, $p = 0.002$, $\beta = -0.246$.

Discussion

The cross-experiments analysis provides a richer picture of the described effects. The LMM approach taken here allows the consideration of variability in overall performance across individuals in all three tasks and overall detectability, identifiability and memorability across all items included. Moreover, we were able to also test for the presence of significant variability in our participants and stimuli reactions to the congruity manipulation in all three tasks. Indeed, the model comparison approach revealed that these are important factors to consider as they explained different portions of the variance in all three tasks (see Supplementary material for the full report of the model comparison procedure).

More interestingly, after controlling for the between participants and stimuli variability, our three targeted indices remained significant. Namely, semantically congruent objects were detected worse but identified and recognized better than semantically incongruent ones. These results are, therefore, robust across different participants (i.e., four different samples), experimental setups (i.e., standard behavioral vs. eye tracking) and paradigm idiosyncrasies (i.e., inclusion of different stimulus types—Experiment 1a and 1b and alterations to the main paradigm—Experiment 2).

General discussion

Meaning maps have been shown to guide attention in scene viewing beyond the influence of perceptual saliency maps (Henderson and Hayes, 2017), thus acting as a tool for selecting which information to prioritize. Nevertheless, we still do not have a clear understanding of which processes are affected by semantic information and when and how that influence is exerted. Is semantic information only capable of influencing behavior after conscious access has taken place? Can semantic properties of objects bias conscious access itself? What is the relationship between these different processes? The lack of integrative theoretical models is most likely responsible for the mixed pattern of results found in previous studies about object detection, identification and remembering (Henson and Gagnepain, 2010; Van Kesteren et al., 2010; LaPointe et al., 2013; Stein and Peelen, 2015; Stein et al., 2016). Although extensive research has been conducted on these three processes, very few studies have attempted at studying them together by pursuing the flow of processing from early perception to remembering. In this study we aimed at jointly studying three key steps across the stream of processing; namely, from the moment in which one becomes aware of the presence of an object (i.e., detection), through that in which we gain access to its identity (i.e., identification) to that in which one is confronted with the need to retrieve it from memory (i.e., delayed recognition).

To approach these three processes, we have made use of three sequential behavioral indices, which result from the semantic relationship between selected objects and the natural scenes in which they are embedded. The first of these three indices, the *detection cost*, is defined as the impaired detection of certain objects when their semantic properties overlap with their surrounding scenes. The second one, the *identification benefit*, refers to an improved identification of certain objects precisely when they are embedded in semantically matching scenes. Finally, the *recognition benefit* reflects a relative advantage in retrieving objects when they had been presented within a congruent background. In order to deepen our knowledge of the processes underlying those effects, we conducted four experiments in which we explored their dependency on the structure of the scenes and the participant's task set (Experiments 1A and 1B), the relationship between the *recognition benefit* and the identification task (Experiment 2), and gaze patterns associated with each index (Experiment 3).

Experiments 1A and 1B showed that using scenes involving multiple objects (i.e., cluttered scenes) was a condition *sine qua non* to measure the *detection cost*. In other words, the *detection cost* only appears when competing objects are present together with the to-be-detected one and, therefore, when object selection is not trivial. The change detection task requires participants to explore different objects in search for a changing one; the *detection cost* effect indicates that such an exploration does not take place randomly, but it tends to prioritize processing of incongruent objects. Even if this exploration is performed on the basis of semantic maps (Henderson and Hayes, 2017) and congruity guides standard search (Peelen and Kastner, 2014), incongruent objects act as a circuit-breaker of the search process and capture attention. In other words, context incongruent objects tend to win the race for attention. Thus, in the same way that an abrupt onset captures spatial attention, incongruent objects seem to automatically segregate themselves from the semantic map and capture attention.

In addition to the implications for the detection process, Experiments 1A and 1B also showed that object competition is not needed to measure the *identification benefit*. Our identification results did not differ when object competition was removed by using sparse scenes. The propagation of activation from the gist of the scene to the identity of the target is most likely underlying the observed benefit in identification obtained in congruent conditions. However, without a neutral condition (i.e., not congruent and not incongruent), it is hard to ascertain whether the *identification benefit* actually reflects an absolute advantage for the processing of congruent targets, or rather an impairment produced in the identification of incongruent targets (see Ortiz-Tudela et al., 2016 for a discussion in favor of the latter account).

Experiment 2 provided a strong dissociation between the *recognition benefit* index and the *identification benefit* one. In this experiment, we eliminated the requirement of performing an identification task after the change detection task. Previous studies that have shown the *recognition benefit* effect could not ascertain whether the effect was due to processes occurring while participants were trying to detect the change, or it rather arose later, as a by-product of the subsequent identification task. By removing the identification task altogether and still measuring the *recognition benefit*, one can safely rule out that such offline re-elaboration was responsible for the better remembering of the congruent objects. Even though we cannot claim that identification processes were completely absent from the task, identifying those targets was not part of the explicit requirements, and thus their importance was reduced with respect to previous experiments (Ortiz-Tudela et al., 2016). Accordingly, compared to previous studies, Experiment 2 produced a much lower overall recognition rate [38% compared to an average of 62% in the experiments reported in Ortiz-Tudela et al. (2016)]. This result attest to the shallower processing (Craik and Lockhart, 1972) promoted by the conditions arranged in Experiment 2, and therefore indicate that the *recognition benefit* does not require the deep semantic elaboration brought about by the explicit requirement to explicitly identify the target.

Experiment 3 characterized the gaze patterns associated with each of the three main effects studied in this article. We found that the *detection cost* is mostly driven by shorter first fixation latencies and fewer saccades required to detect an incongruent target than to do so for a congruent one. Indeed, previous reports of the effect focusing exclusively on response times were unable to clarify whether faster detection of a change in incongruent trials could be due either to earlier attention attraction or to quicker decision process produced after the target was fixated. In other words, faster detection times in response to an incongruent trial may be produced because, once detected, participants responded right away, whereas responding to a congruent target might require them to wait until the next display is shown, in order to ascertain that this is precisely the object that is undergoing the change. However, our data point otherwise. The shorter latencies to fixate incongruent targets, together with the higher cumulative probability of early target fixation for incongruent targets, and the lack of significant differences in dwell time between congruent and incongruent targets, all suggest that incongruent objects act as efficient attractors of attention (Hollingworth and Henderson, 2000; LaPointe and Milliken, 2016).

Interestingly, our findings are in agreement with LaPointe and Milliken (2016) which also found no differences in dwell time between congruent and incongruent items in a change detection paradigm. They aimed at testing two major accounts proposed for explaining differences in object detection times. One of them, the attention attraction hypothesis, states that context-incongruent objects effectively draw attention and therefore speed detection responses; the other one, the attention disengagement hypothesis, proposes that while the scanning of an image would take place randomly, incongruent objects, when found, force attention to linger on them as a mean of extracting more information and when the change occurs attention is already at the appropriate location. Both LaPointe and Milliken (2016)'s results and the present ones seem to be more consistent with the former claim. Future studies restricting eye movements could further extend this finding and reveal whether shorter first fixation latencies are a necessary requirement for the *detection cost* or are just one of many other components.

In contrast to what has been observed for the *detection cost*, neither the *identification benefit* nor the *recognition benefit*, seem to be strongly related to any of the eye gaze measures considered in the current study. Exposure and processing time are potential key variables for both processes and therefore key factors to consider when drawing conclusions about identification or memory effects. The fact that the amount of time spent looking at the objects was equivalent for congruent and incongruent objects in spite of the asymmetry found in the variables measured suggests that exposure or processing time has little to no influence in generating said asymmetries.

Finally, it is worth noting that in our stimulus set, somehow reflecting the real world, there is a marked asymmetry between the amount of congruent and incongruent objects. Indeed, in complex ecological scenes, incongruent objects are, by definition, less frequent than congruent ones. One may argue that this unbalanced proportion could be biasing participants' strategies and affecting some of the processes that we are measuring here. However, the early arising of these effects and the fact that none of them is affected by the manipulation of predictability of the nature of the trials in Experiment 1 weakens the argument that conscious volitional strategies might be responsible for the observed effects. In addition, removing this asymmetry to equate the number of congruent and incongruent objects without disrupting the scene information is virtually impossible. Congruent objects are a constituent part of the scene itself and therefore largely increasing the number of incongruent objects would alter the gist of the scene; in other words, it would change the meaning map that guides the search (Henderson and Hayes, 2017). Some intermediate attempts can nonetheless be taken, for instance adding a few incongruent non-target objects, to reduce the efficiency of a *search-for-incongruity* strategy. Future studies including such manipulation would most likely produce very interesting results concerning the weight of the strategic component on these processes.

In the next subsections, we summarize the major implications of our findings for object detection, object identification and delayed object recognition processes.

Object detection

Conscious detection of a stimulus is arguably the first gate into awareness. Although still controversial, evidence in favor of pre-conscious semantic influences on detection has been continuously increasing. Here we have provided new evidence by showing that, in the presence of many non-target objects, some form of semantic analysis is done prior to conscious detection. When there is competition among a set of objects that are potential targets, the selection process is not randomly performed: rather, it is by default guided by the scene's semantic structure (Peelen and Kastner, 2014; Henderson and Hayes, 2017). This standard search is immediately stopped upon encountering scene-mismatching elements that are prioritized in the analysis since they convey more informational value (e.g., incongruent objects; Santangelo et al., 2015). Eye gaze and selective attention get directed toward those more informative items or regions of the space. Whether object selection occurs first and causes attentional allocation, or whether attention gets allocated and then a given object can be selected, is still unsolved. What can be safely assumed so far is that semantic processing (at least at the category level) leads to conscious detection by accordingly biasing attentional allocation and object selection (Mudrik et al., 2011; Stein and Peelen, 2015).

It is worth noting that, although Stein and Peelen's (2015) results also reflected influences of semantic information on the access to awareness, their manipulation seemed to exert its influence in the opposite direction. In their paradigm, an anticipatory cue was given to participants in the context of a bCFS setup; when the category of the object to be found matched the cue, shorter suppression times were measured. In other words, prior semantic knowledge improved object detection. This result seems in principle to be at odds with the *detection cost* reported here. One might argue that the presence vs. absence of object competition could account for the different results obtained in these two paradigms. However, in the sparse condition of Experiment 1, when most competing objects were eliminated, the *detection cost* was obliterated but no benefit from congruity arose either. Hence, another line of argument is needed to fully explain the overall pattern.

Finding a congruity-driven cost on detection in cluttered scenes, no effect of congruity on sparse scenes, and a congruity benefit when the target appears surrounded by random noise, can all be understood as different points in a continuum of contextual complexity, across which different processes probably operate. For instance, in the bCFS paradigm participants need to report the appearance of any object under a gradually fading mask. Under these circumstances, the ability to differentiate a portion of the display as a meaningful object is crucial, and thus semantic cues that help to disambiguate any *objectness* will speed responses. By way of analogy, consider the scenario of looking for shapes in the clouds. If one is told to look for a dog, it is not difficult to understand that one would find the shape of a dog more easily than something else.

On the other side of the continuum, looking for changes in a cluttered scene makes grasping the *objectness* of an item entirely irrelevant for the task at hand, since a lot of different objects are unambiguously presented. Under these conditions, competition among semantically related objects, or prioritization of highly informative targets, can bias the process to produce the *detection cost*. Halfway through this continuum (i.e., from segregating one object from random noise to selecting one among different clear candidates), is our sparse condition. The sparse scene scenario, in which neither visual noise nor distractor objects compete for the detection of the changing object, would constitute an intermediate case in which neither the cost of competition nor the benefit of semantic cueing would be expected (either due to the absence of the two or to both mechanisms canceling each other across different trials).

Object identification

If one needs to produce a fast and unspecific response whenever something changes, prioritizing novelty seems in order; thus, any information that helps to segregate the new element from its old surrounding environment will aid detection. However, if the task requires instead to identify a target, then it seems much more appropriate to integrate all available information, including the background image, and the semantic properties of all neighboring objects, to disambiguate the to-be-identified targets (Biederman, 1972; Eger et al., 2007; Esterman and Yantis, 2010; Gorlin et al., 2012). Early explorations of this effect showed that when an object is surrounded by a congruent context, its identification is better than when less contextual information is available (Biederman, 1972). Thus, the identification process rather than being negatively affected by object competition is instead positively affected by the propagated activation from the background scene. This pattern of identification responses stands in sharp contrast with that observed for detection responses where semantic incongruity improves performance.

Delayed object recognition

Many different factors determine which information is eventually remembered or forgotten. Here we used recognition just as one of the many ways in which it is possible to assess memory performance. Our results indicate (1) that the congruency-driven *recognition benefit* is observed even under shallow encoding circumstances, not requiring the explicit identification of each object (Experiment 2) and (2) that gaze behavior is unrelated to this memory benefit (Experiment 3). Although more research is needed to further explore these memory results, we hypothesize that the *recognition benefit* arises as a consequence of the integration of the detected object into previous internal schemas (Van Kesteren et al., 2010).

Conclusions

Across four experiments using a change detection paradigm, we explored the influence of semantic congruity on three key stages in the stream of object visual processing. This stream ranges from detection to identification and eventual remembering. We used three behavioral effects: a congruency-driven *detection cost*, an *identification benefit*, and a *recognition benefit*. These behavioral effects were taken as empirical proxies for detection, identification and recognition and used to analyse the dependency of these processes with respect to the semantic factor. We combined online and offline behavioral measures together with eye movement recordings to characterize the gaze patterns associated with each of these three processes. Our results provide new evidence on the nature of these components of visual cognition and illustrate the ways in which semantic support and competition can modulate these successive stages in the flow of information processing. The data presented in this study not only provides insights into the mechanisms underlying key cognitive processes but also presents a challenging yet very promising combined approach by studying the same information across different stages rather than by isolating processes that in the real world are very likely to interact with each other.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://github.com/ortiztud/three_indices.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the University of Granada. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JO-T, LJ, and JL collaborated in the design of the experiments. JO-T carried out the data collection, analyses of results, and writing of the present manuscript. LJ and JL conceptually supervised the project and reviewed several drafts of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The present research was funded by the Spanish Ministerio de Economía y Competitividad with research grants to JL (PSI2017-84926-P and PID2020-114790GBI00) and LJ (PID-2020-116942GB-100), by the Goethe Research Academy for Early Career Researchers as part of their Fokus A/B support line awarded to JO-T, and by the Junta de Andalucía - Consejería de Universidad, Investigación e Innovación - with a research grant to JO-T (P21_00148).

Acknowledgments

We acknowledge Alba Sanz for her help in data collection in Experiments 1A and 1B.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcogn.2023.1125145/full#supplementary-material

References

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Biederman, I. (1972). Perceiving real-world scenes. *Science* 177, 77–80. doi: 10.1126/science.177.4043.77

Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., and Notebaert, W. (2015). Open your eyes for prediction errors. *Cogn. Affect. Behav. Neurosci.* 15, 374–380. doi: 10.3758/s13415-014-0333-4

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science* 281, 1185–1187. doi: 10.1126/science.281.5380.1185

Brod, G., and Shing, Y. L. (2019). A boon and a bane: comparing the effects of prior knowledge on memory across the lifespan. *Dev. Psychol.* 55, 1326–1337. doi: 10.1037/dev0000712

Brod, G., Werkle-Bergner, M., and Shing, Y. L. (2013). The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. *Front. Behav. Neurosci.* 7, 139. doi: 10.3389/fnbeh.2013.00139

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., and Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic response. *Neuroimage* 27, 885–895. doi: 10.1016/j.neuroimage.2005.05.047

Davenport, J. L., and Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol. Sci.* 15, 559–564. doi: 10.1111/j.0956-7976.2004.00719.x

Eger, E., Henson, R., Driver, J., and Dolan, R. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cereb. Cortex* 17, 2123–2133. doi: 10.1093/cercor/bhl119

Esterman, M., and Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cereb. Cortex* 20, 1245–1253. doi: 10.1093/cercor/bhp188

Gorlin, S., Meng, M., Sharma, J., Sugihara, H., Sur, M., and Sinha, P. (2012). Imaging prior information in the brain. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7935–7940. doi: 10.1073/pnas.1111224109

Greve, A., Cooper, E., Tibon, R., and Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *J. Exper. Psychol. Gen.* 148, 325–341. doi: 10.1037/xge0000498

Gronau, N., and Shachar, M. (2015). Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1095–1111. doi: 10.1037/xhp0000071

Henderson, J. M., and Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* 1, 743–747. doi: 10.1038/s41562-017-0208-0

Henderson, J. M., Weeks, P. A., and Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 210–228. doi: 10.1037/0096-1523.25.1.210

Henson, R. N., and Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus* 20, 1315–1326. doi: 10.1002/hipo.20857

Hollingworth, A., and Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cogn.* 2014, 37–41. doi: 10.1080/135062800394775

Hollingworth, A., Schrock, G., and Henderson, J. M. (2001). Change detection in the flicker paradigm: the role of fixation position within the scene. *Mem. Cogn.* 29, 296–304. doi: 10.3758/BF03194923

Kafkas, A., and Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia* 50, 3080–3093. doi: 10.1016/j.neuropsychologia.2012.08.001

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Kaiser, D., Stein, T., and Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonom. Bull. Rev.* 22, 1784–1790. doi: 10.3758/s13423-015-0833-4

Krebs, R. M., Boehler, C. N., De Belder, M., and Egner, T. (2015). Neural conflict-control mechanisms improve memory for target stimuli. *Cereb. Cortex* 25, 833–843. doi: 10.1093/cercor/bht283

Kroll, J. F., and Potter, M. C. (1984). Recognizing words, pictures, and concepts: a comparison of lexical, object, and reality decisions. *J. Verbal Learn. Verbal Behav.* 23, 39–66. doi: 10.1016/S0022-5371(84)90499-7

LaPointe, M. R. P., Lupianez, J., and Milliken, B. (2013). Context congruency effects in change detection: opposing effects on detection and identification. *Vis. Cogn.* 21, 99–122. doi: 10.1080/13506285.2013.787133

LaPointe, M. R. P., and Milliken, B. (2016). Semantically incongruent objects attract eye gaze when viewing scenes for change. *Vis. Cogn.* 24, 63–77. doi: 10.1080/13506285.2016.1185070

Lupiáñez, J., Martín-Arévalo, E., and Chica, A. B. (2013). Is inhibition of return due to attentional disengagement or to a detection cost? The detection cost theory of IOR. *Psicológica* 34, 221–252.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

Mooney, C. M. (1957). Age in the development of closure ability in children. *Can. J. Psychol.* 11, 219–226. doi: 10.1037/h0083717

Mudrik, L., Breska, A., Lamy, D., and Deouell, L. Y. (2011). Integration without awareness. *Psychol. Sci.* 22, 764–770. doi: 10.1177/0956797611408736

Ortiz-Tudela, J., Martín-Arévalo, E., Chica, A. B., and Lupiáñez, J. (2018a). Semantic incongruity attracts attention at a pre-conscious level: evidence from a TMS study. *Cortex* 102, 96–106. doi: 10.1016/j.cortex.2017.08.035

Ortiz-Tudela, J., Milliken, B., Botta, F., LaPointe, M. R. P., and Lupiañez, J. (2016). A cow on the prairie vs. *a cow on the street: long-term consequences of semantic conflict on episodic encoding. Psychol. Res.* 81, 1264–1275. doi: 10.1007/s00426-016-0805-y

Ortiz-Tudela, J., Milliken, B., Jimenez, L., and Lupianez, J. (2018b). Attentional influences on memory formation: a tale of a not-so-simple story. *Mem. Cognit.* 46, 544–557. doi: 10.3758/s13421-017-0784-2

Otero, S. C., Weekes, B. S., and Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology.* 48, 1346–1353. doi: 10.1111/j.1469-8986.2011.01217.x

Paller, K. A., and Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends Cogn. Sci.* 6, 93–102. doi: 10.1016/S1364-6613(00) 01845-3

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Mem. Cogn.* 3, 519–526. doi: 10.3758/BF03197524

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology* 11, 559–566. doi: 10.1111/j.1469-8986.1974.tb01114.x

Peelen, M. V., and Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250. doi: 10.1016/j.tics.2014.02.004

Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return: neural basis and function. *Cogn. Neuropsychol.* 2, 211–228. doi: 10.1080/02643298508252866

Quent, J. A., Greve, A., and Henson, R. (2021). Shape of U: the relationship between object-location memory and expectedness. *Preprint* 53, 1689–1699. doi: 10.31234/osf.io/xq37j

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8, 368–373. doi: 10.1111/j.1467-9280.1997.tb00427.x

Rosner, T., D'Angelo, M., MacLellan, E., and Milliken, B. (2015). Selective attention and recognition: effects of congruency on episodic learning. *Psychol. Res.* 79, 411–424. doi: 10.1007/s00426-014-0572-6

Santangelo, V., di Francesco, S. A., Mastroberardino, S., and Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Hum. Brain Mapp.* 36, 5003–5017. doi: 10.1002/hbm.22984

SensoMotoric Instruments (SMI) (1991). *BeGaze*. SensoMotoric Instruments Teltow, Germany.

Simons, D. J., and Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends Cogn. Sci.* 9, 16–20. doi: 10.1016/j.tics.2004.11.006

Spaak, E., Peelen, M. V., and De Lange, F. P. (2020). Scene context impairs perception of semantically congruent objects. *bioRxiv. [preprint]*. doi: 10.1101/2020.09.30.320168

Stein, T., and Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *J. Exp. Psychol. Gen.* 144, 1089–1104. doi: 10.1037/xge0000109

Stein, T., Siebold, A., and van Zoest, W. (2016). Testing the idea of privileged awareness of self-relevant information. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 303–307. doi: 10.1037/xhp0000197

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science* 314, 1311–1314. doi: 10.1126/science.1132028

Tsuchiya, N., and Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.* 8, 1906. doi: 10.1038/nn1500

Van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., and Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *J. Neurosci.* 30, 15888–15894. doi: 10.1523/JNEUROSCI.2674-10.2010

Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., and Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends Neurosci.* 35, 211–219. doi: 10.1016/j.tins,0.2012.02.001

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., et al. (2008). The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology.* 45, 130–140. doi: 10.1111/j.1469-8986.2007.00606.x

Wessel, J. R., and Aron, A. R. (2017). On the globality of motor suppression : unexpected events and their influence on behavior and cognition. *Neuron* 93, 259–280. doi: 10.1016/j.neuron.2016.12.013

Wynn, J. S., Ryan, J. D., and Moscovitch, M. (2019). Effects of prior knowledge on active vision and memory in younger and older adults. *J. Exp. Psychol. Gen.* 149, 518–529. doi: 10.1037/xge0000657