Special Issue Article

QJEP

Quarterly Journal of Experimental Psychology
1–11


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17470218261417305
qjep.sagepub.com
Sage

Effect of Prior Haptic Object Exploration on Eye Movements

Matteo Toscani[1], Mark Gather[1], Ellen Seiss[1] and Anna Metzger[1]

Abstract

Interaction with objects typically involves both vision and touch. Understanding how visual and haptic information interact during object exploration is essential to uncovering the mechanisms of multisensory shape perception. We investigated whether haptic exploration influences subsequent eye movements, using a cross-modal shape comparison task. Participants ($N = 22$) explored 3D replicas of bell peppers either haptically or visually, and subsequently viewed the same or a different object. We tracked eye movements during visual explorations. Comparing uni-modal visual to cross-modal, haptic-to-visual conditions, we found that prior haptic exploration led to significantly shorter fixations, longer and faster saccades, as well as larger coverage of the image with fixations during subsequent visual exploration – indicative of a broader, more distributed scanning pattern. These effects suggest that visual saliency is modulated by prior tactile experience, challenging purely unimodal or bottom-up models of attentional guidance.

Keywords

multisensory, saliency, visuo-haptic, shape perception

Received: 30 June 2025; revised: 19 December 2025; accepted: 5 January 2026

Introduction

We constantly interact with objects in our daily life, which enables us to affect the world around us according to our goals. Perception of objects' shape is crucial for this purpose. During this inherently active process, we sample information about objects with our sensory organs, which usually involves more than one sense. For example, we often do not only look at objects, but we also touch them to explore their properties at the same time. Though we obtain very different information from both senses (2D projection of the object on the retina vs. pressure and vibratory patterns distributed across the hand and proprioceptive information of finger positions), we can recognise objects that we have touched beforehand when we see them and vice versa (Norman et al., 2004), indicating the existence of multimodal object shape representations. Additional evidence for such multimodal representation comes from neurophysiological research. Buelte et al. (2008) asked participants to recognise objects visually or via touch, which they had seen or touched beforehand, accordingly (uni-modal recognition). In a different condition, participants were asked to recognise objects visually after touching them and vice versa (cross-modal recognition). Repetitive transcranial magnetic stimulation (rTMS) was applied over the anterior intraparietal sulcus (IPS) during the encoding of the stimuli (i.e. presentation of the reference stimulus for later recognition). The anterior IPS is suggested to play a crucial role in the integration of tactile and visual information during object manipulation. Cross-modal recognition was selectively deteriorated via rTMS during visual encoding for haptic recognition, while task performance was not significantly altered in the uni-modal conditions. These results indicate that additional multisensory representations or mapping mechanisms are required for cross-modal object recognition, which

[1]School of Psychology, Bournemouth University, Poole, UK

Corresponding author:
Anna Metzger, Bournemouth University, Poole BH12 5BB, UK.
Email: ametzger@bournemouth.ac.uk

are different from uni-modal ones. Interestingly, rTMS stimulation did not affect performance in the cross-modal recognition in the other direction, that is, when the stimulus was encoded haptically and recognised visually, indicating that the mapping from one modality to another is specific for each direction.

The nature of such cross-modal representations is not yet well understood. Uni-modal object recognition is typically view-dependent for both vision and touch (Newell et al., 2001). It was demonstrated that objects are better recognised when seen from a familiar view (Jolicoeur, 1985) or for novel objects from the same view as they were presented during the encoding phase (Newell et al., 2001). Though for haptic perception exploration is not restricted (i.e. the object can be touched from all sides), a view dependence was demonstrated too (Newell et al., 2001). Interestingly, for cross-modal recognition in both directions (from vision to touch and from touch to vision), view-dependence was inverted, that is, recognition was better when the object was rotated 180° about the vertical axis relative to its orientation during encoding with the different sense (Newell et al., 2001). These results suggest that both uni-modal representations differ between the senses and depend on exploration, which, in turn, affects cross-modal comparisons. Note, however, later studies showed that when objects are rotated about all axes, which resembles the natural situation more closely, cross-modal object recognition is view-independent, while uni-modal comparisons remain view-dependent (Lacey et al., 2007; Ueda & Saiki, 2012). These results imply that there might be different object representations for the cross-modal mapping, which are more general or abstract than uni-modal representations. Interestingly, such view-independent object recognition does not require a cross-modal comparison, as it could be triggered just by not telling participants if they would need to do cross-modal or uni-modal object recognition for both within-and cross-modal comparisons (Ueda & Saiki, 2007). These results indicate that potentially a different exploration of the object takes place in case it needs to be compared across senses. Indeed, Ueda and Saiki (2012) showed that eye movements in visual encoding in preparation for cross-modal object recognition were different from eye movements in the uni-modal one. They found more diffuse and longer fixations during encoding when participants expected cross-modal retrieval. However, it is not clear if, in this case, this reflects a task-driven, more conservative exploration in preparation for the more difficult comparison, which is consistent with longer and more diffuse fixations, rather than the construction of a different object representation. Crucially, knowledge is missing about how such cross-modal representations might guide subsequent object exploration in another sensory modality.

Any perceptual representation is based on acquired sensory input, which is determined by how we actively sample objects. When exploring objects, we look at some parts of the object more than at others. We also touch some parts more than others. This means that some parts are more salient to vision and touch. Investigations of visual saliency suggest that it is, to some extent, driven by *bottom-up processes*, that is, it can be predicted from stimulus properties (Itti & Koch, 2001; Treue, 2003), implying a certain degree of automatisation. For example, high luminance contrast and moving stimuli automatically attract gaze. However, more recent findings suggest that there are also *top-down influences* on eye movements from higher-level factors such as task demand or value (Hayhoe & Ballard, 2005; Schütz et al., 2011). For instance, when participants are asked to estimate object colour, they fixate the parts of the image which are most informative for this task (Toscani et al., 2013a, 2013b). There also seem to be specific task-dependent eye-movement strategies, similar to exploratory procedures in haptic perception (Aizenman et al., 2024). Furthermore, the visual system seems to be finely and adaptively tuned to extract task-relevant information, as evident from cases when it is faced with conflicting perceptual tasks or dynamic scenes. For instance, when participants were judging the lightness of a moving stimulus, fixation landing positions were balanced to allow both – following the object with the eyes and looking at its most informative regions (Toscani et al., 2016). And when they were asked to alternate between gloss and speed judgements of a moving stimulus, with gloss-diagnostic or speed-diagnostic features dynamically changing position, they could constantly change their fixation allocation towards the most diagnostic regions for each task (Toscani et al., 2019). We have recently shown that later and long latency saccades are mostly task-driven and contribute to higher performance in the perceptual task (Metzger et al., 2024).

Less is known about haptic saliency. Contrary to visual saliency, touch perception relies mostly on *top-down*, task-dependent stereotypical exploration (Lederman & Klatzky, 1987). For instance, lateral movements across the surface are used to perceive roughness, whereas contour following is used to perceive shape. Such movements seem to be tuned to extract task-relevant information, as they correlate with higher performance for the specific task. Such an exploration strategy is sensible for a perceptual system with a small field of "view," as bottom-up saliency requires pre-attentive inspection of large portions of space. However, haptic explorations seem to be less systematic when the whole hand is used as compared to one finger explorations (Morash, 2016), suggesting that systematic movements might be complemented by *bottom-up saliency* in natural exploration when broader sensory input is available. In fact, we found evidence for foveation-like behaviour in whole-hand haptic search, consisting of a first quick and coarse exploration of the search space followed by detailed exploration of potential targets with the index finger (Metzger, Toscani, Valsecchi, et al., 2021;

Metzger et al., 2019). Crucially, we could show that such behaviour is functional to information gain, as it was more prominent in difficult search and the index finger revealed the highest sensitivity (Metzger et al., 2020). In line with this, we showed that touched locations on the stimulus can be, to some extent, predicted from stimulus properties (Metzger, Toscani, Akbarinia, et al., 2021).

However, although it is known that multimodal representations of objects exist, whether and how visual and haptic saliency interact with each other to drive sensory exploration has not yet investigated. In fact, saliency is often operationalised as a performance advantage rather than exploratory movements; for example, salient items are the ones to which we react faster. For instance, in visual search paradigms, participants are typically asked to detect a target item among distractors when a target differs from distractors in a salient feature such as colour, orientation, motion or luminance visual attention. In these tasks, participants are automatically drawn to the unique item, leading to a highly efficient search (Treisman & Gelade, 1980; Wolfe, 1994). Similarly, in haptic search paradigms, participants explore objects or textures with their hands to find a target differing from distractors. Studies have shown that salient features such as roughness, temperature, material properties or shape can lead to faster and more efficient detection (Lederman & Klatzky, 1997; Overvliet et al., 2008; Plaisier & Kappers, 2010; Plaisier et al., 2008) and sharp edges and vertices for three-dimensional objects (Plaisier et al., 2009). While performance-based studies suggest that attention can interact across touch and vision (Driver & Spence, 1998; Spence et al., 1998, 2000), there is currently no evidence that saliency-driven sensory exploration interacts across modalities.

Here, we investigated whether touch saliency affects subsequent visual exploration. We recorded gaze behaviour when participants looked at objects (replicas of natural bell peppers), which they had seen or touched before, to decide whether their shapes are same or different. Based on previous research, we predict that exploration of objects in one sensory modality affects subsequent exploration of the object in another sensory modality, that is, that visual and haptic saliency interact. More specifically, we hypothesise that visual explorations will be at least partially guided by prior haptic exploration.

Methods

Participants

Twenty-two Bournemouth University students (13 females, mean age 21.6 years, age range 18–33 years) volunteered to participate in the experiment by signing up for the study on the SONA platform. We recruited a total of 36 participants, but excluded 14 because they did not complete at least one trial per condition per shape (please see the "Procedure" section). An *a priori* power analysis was conducted using G*Power (Faul et al., 2009) to determine the required sample size for a repeated measures ANOVA with one group and three measurements. The analysis was based on a significance level of $\alpha = .05$, a desired power of 0.95, and a large effect size ($\eta^2 = .14$). The choice of large effect sizes is justified by our previous study on visual saliency, which showed a large effect of task on eye movement allocation (Metzger et al., 2024). The results of the power analysis indicated that a minimum of 18 participants would be required to detect a statistically significant effect. Participants provided informed consent and were compensated for their time with course credits at the rate of 1 credit per hour. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Bournemouth University (ID: 48806).

Design

We used a within-subject design in combination with a two-interval forced choice (2IFC) paradigm. Participants were presented with 1 of 12 bell peppers. In the first interval, the exploration was either visual or haptic; in the second interval, exploration was always visual. This resulted in two *experimental conditions* (cross-modal vs. uni-modal comparison). However, as we are interested in visual exploratory behaviour we differentiate three *viewing conditions*: Haptic 1st (H 1st), relating to visual exploration of the stimulus during the second interval, when the object was explored haptically in the first interval. Visual 1st (V 1st), relating to visual exploration of the stimulus in the first interval in visual comparisons, and Visual 2nd (V 2nd), relating to visual exploration of the stimulus in the second interval in the visual comparison. Dependent variables were eye movement parameters as described by Greene et al. (2012), that is, spatial distribution of fixation density, fixation number, fixation duration, saccadic duration, saccadic amplitude and speed.

Setup

The participants were sitting comfortably at a desk 80 cm away from the screen (Figure 1A). The head position was stabilised by a chin and forehead rest. The haptic stimuli and the exploring hands were covered by a box and a curtain during each trial. In between the trials and when not in use, the haptic stimuli were hidden behind the screen and additional shields. In trials with haptic exploration, they were handed to the participant by the experimenter below the curtain, ensuring that they could not be seen by the participant. The visual stimuli were displayed on a 24-inch BenQ XL monitor with a resolution of 1920×1080 pixels (100 Hz). Experimental software was implemented in Matlab R2019a and Psychtoolbox (Kleiner et al., 2007).

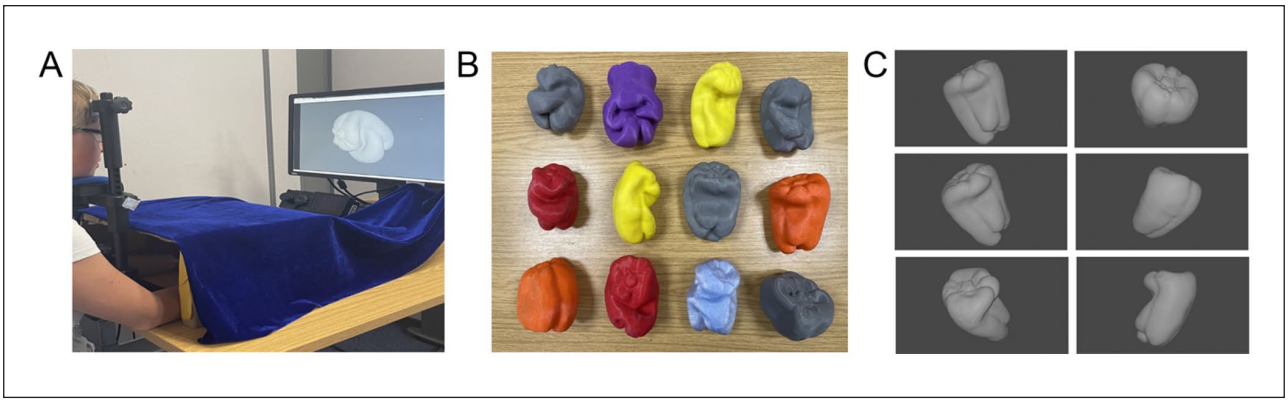

Figure 1. Set up and stimuli. (A.) Experimental setup. The head of the participant was stabilised by a chin and forehead rest. The haptic stimuli and the exploring hands were covered by a box and a curtain. In each trial with haptic exploration, one object was handed to the participant below the curtain from behind the screen. Visual stimuli were displayed on a computer screen. Eye movements were recorded by the Eyelink 1000 eye tracker. (B.) The haptic stimuli were 12 replicas of bell peppers with natural shape variations. (C.) Example of a visual stimulus with six views of the 3D model of the bell pepper showing the full rotation of the object in 60° steps, from top left to bottom right.

Eye-Tracking

Gaze position was recorded throughout the experiment using a desktop-mounted EyeLink 1000 eye tracker (SR Research Ltd., Osgoode, Ontario, Canada) at a sampling rate of 500 Hz. Participants viewed the display binocularly, but only the right eye was tracked. The eye tracker was calibrated at the start of each session and validated. Calibration was accepted only if the mean validation error was below 0.5° of visual angle. At the beginning of each trial, calibration accuracy was checked. Recalibration was conducted when the error exceeded 1.5°; otherwise, a drift correction was applied. Eye movements were classified using EyeLink's standard saccade detection algorithm, with velocity and acceleration thresholds of 30°/s and 8,000°/s^2, respectively. Consecutive samples without saccades were averaged into single fixations. Saccades and fixations were assigned a frame only if they (a) started after the frame was presented and (b) ended before the frame was removed.

Stimuli

We used a replica of bell peppers (Capsicum annum) in their original size as used in Norman et al. (2004) as they present naturally complex objects, rich in natural variations in shape and novel to the participants. The haptic stimuli (Figure 1B) were 3D printed using cornstarch-based polylactic acid using a MakerBot Replicator (5th Gen) Desktop 3D Printer. Their weight was between 60 and 100 g. As the haptic stimuli were never compared with each other, variations in the weight were non-informative for the perceptual task. The visual stimuli were rendered in Blender 4.4.3. Each pepper was first oriented in a way that its main axis was diagonal, and then it was fully rotated stepwise in 60° steps around the vertical axis, resulting in 6 views of the pepper (see example in Figure 1C). This way the rotation exposed both the top and bottom of the pepper to the participant's view, optimising the visibility of the pepper's shape. The renderings were presented in full screen with the peppers being seen at approximately 14° of visual angle.

Procedure

The procedure is outlined in Figure 2. Upon arrival, participants read a Participant Information Sheet, which included all the relevant information about the study, including the instructions for the experiment, before giving written informed consent on the Participant Agreement Form. Initial nine-point calibration was performed prior to the experiment and validated to have an error $<0.5°$ visual angle. The experiment followed a 2IFC design. In the first interval, participants explored 1 of the 12 peppers either haptically or visually for 6 s, to keep performance in the dynamic range (Norman et al., 2004). The pepper number was displayed to the experimenter on a portion of the screen invisible to the participant. In the cross-modal condition, participants were informed via text on the screen that they had to explore haptically. The experimenter handed the pepper below the curtain covering the stimulus and the hands to the participant and pressed the space bar on the keyboard to start the timer for the exploration. Haptic exploration was unrestricted and with both hands. A tone indicated the end of the exploration, and participants placed the pepper down onto the table, which was then taken back by the experimenter. In the uni-modal condition, the trial started with a screen presenting a fixation

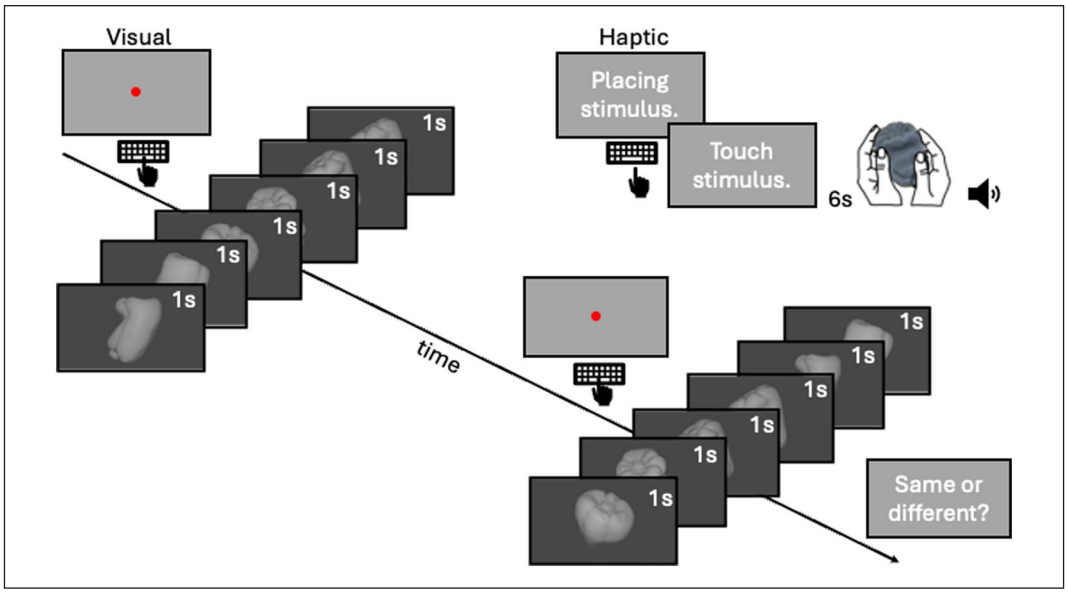

Figure 2. Procedure. The uni-modal condition started with a red fixation dot. Once participants fixated on the dot, they pressed the space bar. This started the presentation of the visual stimulus, which consisted of 6 views of a full stepwise rotation of 1 of 12 bell peppers. Each view was presented for 1 s. In the cross-modal condition, participants were informed that the experimenter would place the object into their hands. Once the stimulus was placed, the experimenter pressed the space bar, which prompted participants to start exploring the stimulus haptically. A beeping tone indicated the end of the exploration, and participants placed the stimulus on the table for the experimenter to collect. In both conditions, the 6 s long exploration of the first stimulus was followed by a screen with the red fixation dot. Once participants fixated on the dot, they pressed the space bar, which started the presentation of the second visual stimulus. Here, the same or a different bell pepper was shown in six views of its full stepwise rotation, starting at a random initial view. At the end, participants responded verbally if the peppers presented in the two intervals were the same or different.

dot. Once participants fixated, they pressed the space bar to start the presentation of the first stimulus. Each view of the pepper was shown for 1 s in the order of rotation. In the second interval, the stimulus was always explored visually. Visual presentations started at a random initial view (among the six views) to promote shape over sequence comparisons. Once both intervals were completed, participants decided whether the two peppers were the "same" or "different." They gave their response verbally, and the experimenter recorded the answers by pressing "s" or "d" on the keyboard. No feedback about the correctness of the response was provided. In most of the trials, the objects were the same in both intervals to ease the qualitative comparison between gaze patterns. However, one-third of the trials were catch trials with different objects to keep participants engaged in the task. The presentation of the stimuli and experimental conditions was blocked. In each block, each pepper was compared with one of the other peppers (randomly assigned) under the 2 experimental conditions (uni-modal vs. cross-modal comparison) for 3 times (repetitions), resulting in overall 72 trials per block: 2 experimental conditions $\times$ 12 peppers $\times$ 3 repetitions. The stimuli and experimental conditions were presented in random order within each block. The experiment was terminated after 2 hr allowing completion of approximately three blocks; however, the exact trial number varied between participants for this reason. We excluded participants who did not complete at least one trial per condition, per shape. Participants were informed they were able to take breaks when they required, and a small break was offered after 1 hr.

Analysis

We first wanted to assess whether there were systematic differences in the fixated regions depending on prior tactile or visual, or no prior exploration – which corresponds to the initial visual exploration (H 1st, V 2nd and V 1st, respectively). Fixation density maps were computed for each trial and for each of the six images corresponding to the six rotations by assigning a value of 1 to the fixated pixel and 0 to all other pixels. To account for eye-tracker measurement error and the size of the fovea, the images were then filtered using a Gaussian filter with a sigma of 0.5 degrees of visual angle (dva), and subsequently normalised so that the total sum of pixel values equalled 1.

We assessed systematic differences of classic eye movement statistics (Greene et al., 2012) between viewing conditions. For each participant and viewing condition, we computed the average fixation duration, saccadic amplitude,

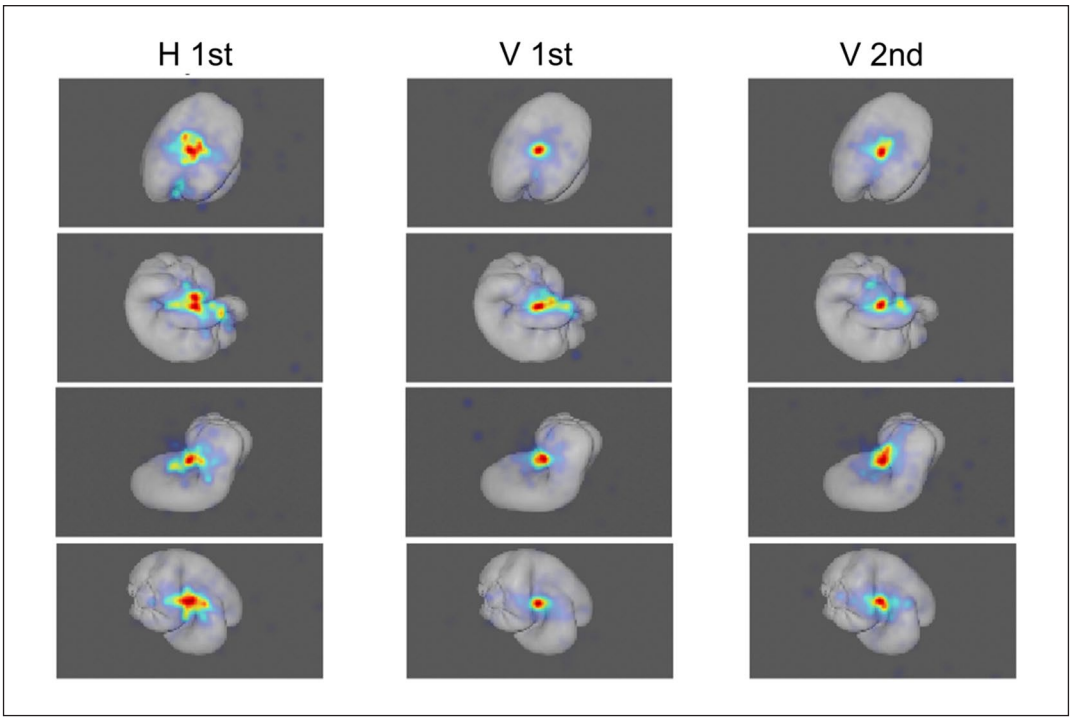

Figure 3. Example heatmaps in the three different viewing conditions averaged across participants. H 1st, eye-movements during the second interval, when the object was explored haptically in the first interval. V 1st, eye movement in the first interval in visual comparisons. V 2nd, eye movements in the second interval in the visual comparisons.

saccadic speed and number of fixations per trial across shapes and rotations. Additionally, as a measure of coverage, we computed the area of a PCA ellipse (Metzger & Toscani, 2022) fitted to fixation locations on each object and each condition across frames and repetitions. We then ran a one-way repeated-measures ANOVA for each of these measures. In addition to the ANOVAs, we ran post hoc Bonferroni-Holm-corrected comparisons to test for differences between each viewing condition.

Results

Performance of participants was in the dynamic range (78% correct on average) and significantly higher than chance, $t(21) = 15.36$, $p < .001$. There was a slightly better performance in the uni-modal condition, 79% as compared to the cross-modal condition, 0.76%, but the difference is not significant, $t(21) = 0.74$, $p = .467$, $BF_{01=}3.5$, suggesting moderate evidence for the H_0.

Figure 3 shows example heat maps for the three viewing conditions. The heat maps appear more similar between the initial and subsequent visual explorations than in the case where participants first touched the stimulus before visually inspecting it.

Figure 4 shows the average fixation duration, saccadic amplitude, saccadic speed and the number of fixations for each viewing condition.

Fixation duration is shorter in the H 1st viewing condition, as confirmed by the ANOVA, $F(2,42) = 4.26$, $p = .021$, $\eta^2 = .168$. Post hoc comparisons showed a significant difference between the H 1st and V 1st viewing conditions, $t(42) = 2.91$, $p = .017$, Cohen's $d = 0.51$.

Saccadic amplitude is larger in the H 1st viewing condition than in both the other two viewing conditions as revealed by the ANOVA, $F(2,42) = 51.94$, $p < .001$, $\eta^2 = .712$. Post hoc comparisons showed a significant difference between the H 1st viewing condition and the V 1st, $t(42) = 8.74$, $p < .001$, Cohen's $d = 1.66$, and V 2nd, $t(42) = 8.91$, $p < .001$, Cohen's $d = 1.65$, viewing conditions.

Saccadic speed is higher in the H 1st viewing condition than in the other two viewing conditions as revealed by the ANOVA, $F(2,42) = 39.39$, $p = < .001$, $\eta^2 = .652$. Post hoc comparisons showed a significant difference between the H 1st viewing condition and the V 1st $t(42) = 6.62$, $p < .001$, Cohen's $d = 1.11$ and V 2nd, $t(42) = 8.43$, $p < .001$, Cohen's $d = 1.93$, viewing conditions.

Coverage ellipse area is higher in the H 1st condition than in the other two viewing conditions as revealed by the ANOVA, $F(2,42) = 17.64$, $p = < .001$, $\eta^2 = .457$. Post hoc comparisons showed a significant difference between the H 1st viewing condition and the V 1st $t(42) = 3.89$, $p < .001$, Cohen's $d = 0.79$, and V 2nd, $t(42) = 5.83$, $p < .001$, Cohen's $d = 1.19$, viewing conditions.

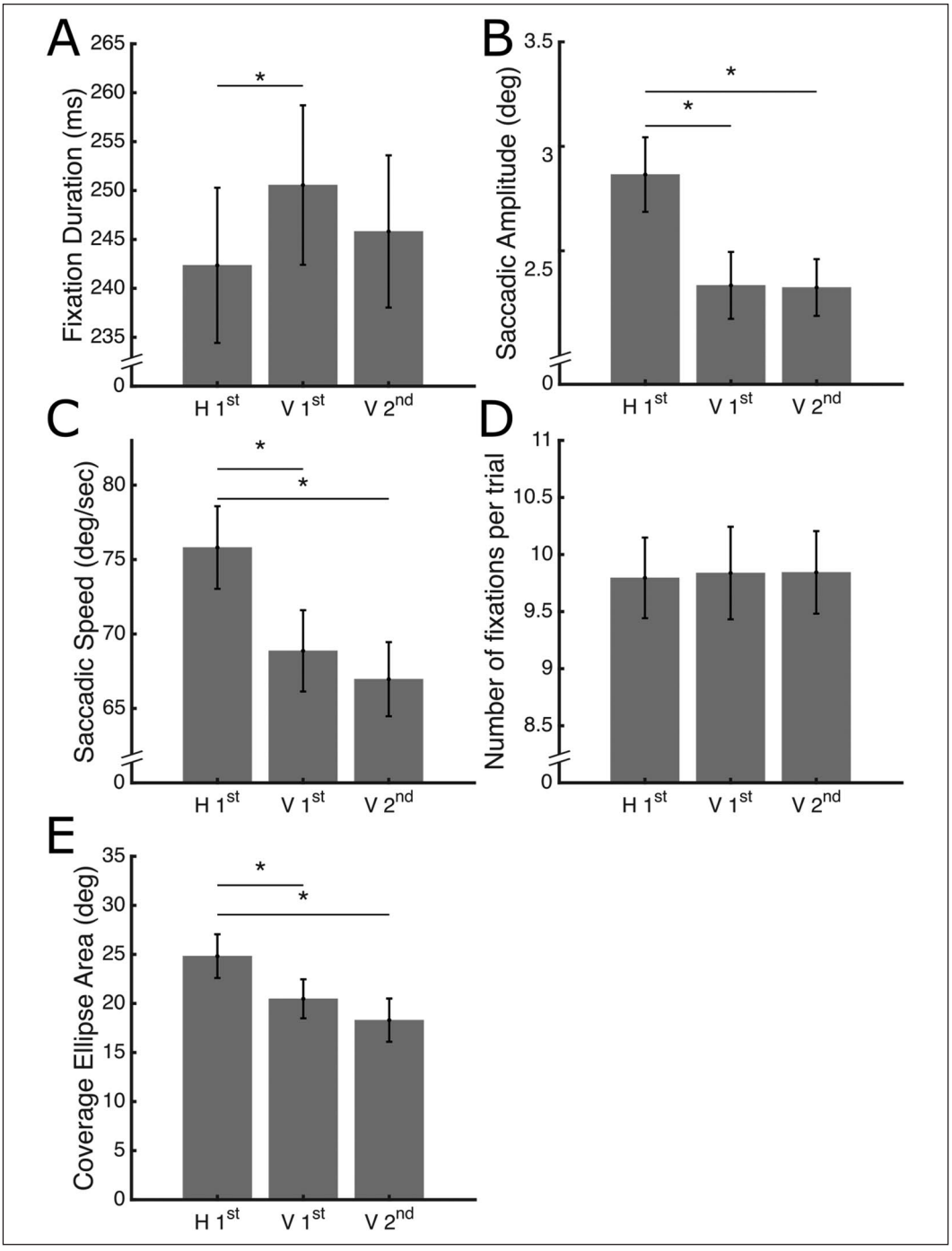

Figure 4. Average eye-movement parameters as a function of viewing condition. Data are averaged across objects and views for each participant and then across participants. The error bars represent the standard error of the mean. Significant post hoc comparisons are indicated with *. (A.) Fixation duration. (B.) Saccadic amplitude. (C.) Saccadic Speed. (D.) Number of fixations per trial. (E.) Coverage, computed as the area of a PCA ellipse.

No significant differences were found for the number of fixations per trial, $F(2,42) = 0.07$, $p = .934$, $\eta^2 = .003$. These results suggest that, while participants produce a similar number of fixations per trial across viewing conditions (around 1 per frame), touching the object first causes them to produce shorter fixations and longer and faster saccades, implying a broader exploration of the visual stimuli.

Discussion

We investigated whether visual exploration of an object would be affected by prior haptic exploration of the same object. In agreement with our hypothesis, we found that when participants first touched an object, the way they looked at it afterwards systematically differed from how

they looked at it the first time or after having looked at it once before. This means that visual exploration was not entirely driven by visual saliency, and haptic exploration influenced it in a way that cannot be merely explained by previous exposure. Specifically, we found shorter fixations, longer and faster saccades and larger coverage after haptic exploration, indicating a more distributed fixation pattern, consistent with our visual inspection of heatmaps.

The results are consistent with our hypothesis that visual saliency was affected by haptic exploration. While perceptual spaces are rather similar between vision and touch, different features are important for each sense (Gaissert et al., 2010). For instance, symmetry seems to be an important feature for vision, while convolutions (bulks) seem to be more important for touch, potentially being differently salient in each sense. Hence, while exploring the objects haptically, participants might have focused on different parts of the peppers, which they then needed to find in the visual counterpart. This could explain the more distributed fixation pattern.

Another reason why we observed such a pattern could be that haptic exploration is more serialised, resulting in a more distributed exploratory pattern (see examples in Metzger, Toscani, Akbarinia, et al., 2021). In fact, despite being able to entirely enclose the object in the hands, when attempting to extract object shape via touch, participants still need to follow its contours with the fingertips (Lederman & Klatzky, 1987). This is consistent with detailed exploration with more sensitive parts of the hand, following a coarse exploration (Metzger, Toscani, Valsecchi, et al., 2021; Metzger et al., 2019). Indeed, cross-modal object recognition is generally better in the other direction – when visually encoded objects are haptically retrieved (Lacey & Campbell, 2006), in line with the idea of a broader and more efficient capture of shape by vision.

However, it is possible that the difference arises not because of the different salient shape features between the two modalities, but because of the different amount of acquired information. For instance, a poorer prior visual exploration could have yielded a more diffuse subsequent exploration, as we found after prior haptic exploration. This could reflect a compensatory strategy, as we did not find a significant difference in performance between the conditions. While it is difficult to quantify the information acquired, measuring tactile exploration and relating it to performance would provide valuable insight and should be the focus of future research.

Another possibility is that the different exploration of the visual stimulus after haptic exploration, as compared to uni-modal comparisons, as observed here, can be explained by the presence of a different cross-modal representation of the object. Indeed, previous results indicate that potentially cross-modal object representations are different from uni-modal ones, being more general or abstract (Lacey et al., 2007; Ueda & Saiki, 2012). Saliency and, therefore, attention allocation are tightly bound to the way our perceptual systems represent objects and shapes. Low-level, local visual features such as contrast and orientation explain only a modest amount of fixation behaviour, whereas object-based models explain significantly more (Einhäuser et al., 2008; Kümmerer et al., 2014, 2016). This challenges early models that assumed saliency is computed primarily in early visual areas based on simple local features such as colour, orientation and contrast (Itti & Koch, 2000; Li, 2002) and supports the view that attentional guidance depends on object-level information. Neurophysiological evidence shows that saliency is represented in frontal areas such as the frontal eye fields (FEF, Thompson & Bichot, 2005), and that FEF activity causally modulates saliency signals in area V4 (Armstrong et al., 2006), a region responsive to both simpler features and objects, and is an important candidate for saliency computation (Bichot et al., 2005). These findings suggest that saliency is computed in mid- and high-level visual areas, where object representations play a central role in guiding attention. However, based on our experiment, we cannot differentiate such an effect from an influence of just a different object representation in a different sense, as discussed before.

We focused our analyses on the trials in which the same object was presented in both intervals to be able to compare the gaze patterns on the same image in different viewing conditions and to be able to interpret potential differences. However, we expect that the observed pattern of results also be observed for comparisons in trials with different objects. Indeed, all results replicate when including all trials in the analyses (all $p < .05$). Such an observation is compatible with all potential explanations for the effect. (a) Matching the obtained haptic representation with the visual stimulus would render different parts of the visual stimulus salient, independent of whether the objects presented visually are the same or different. For instance, if feature a in the haptic stimulus is salient, and feature b is salient in the visual comparison stimulus, both features a and b would attract gaze, leading to different gaze behaviour, though the visual object might overall be different. (b) If the effect arises because participants replicated the generally more distributed haptic exploration visually or (c) because they acquired less information by touch, this also holds for a different object. Finally, (d) if a separate cross-modal representation of the object was constructed after haptic exploration, it still can affect the exploration of the comparison object because, haptic perception of shape is expected to be rather coarse and compatible with multiple visual stimuli, given that performance for visual retrieval of haptically encoded objects is worse than in the other direction (Lacey & Campbell, 2006).

We have chosen to do all analyses frame-wise and exclude eye-movements between frames because they are likely smooth pursuit movements elicited by a sense of

motion caused by the pseudo-rotation of the pepper across the frames. However, we have rerun all analyses, including all eye movements during each trial. As anticipated, some fixations likely reflect smooth pursuit, which led to an overall inflation of fixation duration in this analysis. Apart from this, all our results replicate.

While we find shorter fixation duration after haptic exploration, as compared to viewing the first stimulus in the uni-modal condition, we do not find a significant difference in the number of fixations. However, fixation duration and number of fixations per trial are not negatively correlated as would be expected. This is likely due to the fact that the ~8 ms shorter fixation duration after haptic exploration, while statistically significant, is too small to allow for an additional fixation. In line with this, we have reported both saccadic amplitude and saccadic speed for completeness; however, they are expected to correlate, that is, longer saccades are also faster as there is the main sequence for saccades (Bahill et al., 1975).

We have chosen bell peppers as stimuli, which were previously extensively used in research on visual and haptic, unimodal and multisensory object perception (Dowell et al., 2018; Norman et al., 2004, 2021). These stimuli are optimal for exploring how object shape representations are matched across senses, as they are novel to participants and exhibit natural, complex variations in shape, while remaining similar in simpler features such as volume. However, in everyday situations, we mostly interact with known objects, and saliency is determined by factors such as context and task (Rothkopf et al., 2016). Similarly, object affordances (i.e. actions which can be performed on an object, e.g. pulling a cord) affect gaze behaviour (Gomez & Snow, 2017). For instance, fixations focused on the tool's manipulation area, such as the handle of a hammer as people have prior knowledge of how to use it (Federico & Brandimonte, 2019). While such situations and objects present a more natural case, they make it more difficult to focus on understanding gaze behaviour employed in multisensory shape perception. For example, participants could use strategies to compare known objects, for example, by focusing on object features which are specific to an object type (e.g. the handles or heights of mugs vs. the shape and length of spoons as distinctive features).

In conclusion, we have shown that prior haptic exploration of a 3D shape influences visual exploration of the same shape. Participants produced longer saccades and fewer fixations, consistent with a more distributed exploration pattern. This implies that visual saliency is not only based on local *bottom-up* features or unimodal saliency maps but also on multimodal object representations.

ORCID iD

Anna Metzger https://orcid.org/0000-0002-5704-2821

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics committee (ID: 48806).

Consent to Participate

Written informed consent was provided.

Authors' Contributions

A.M. and M.T. conceived of and designed the experiment, M.G. carried out the experiment, A.M., M.T., and M.G. analysed the data, M.T. and A.M. wrote the manuscript, M.T., M.G., and E.S. provided critical feedback.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The research was supported by the Royal Society RG\R1\241159 – Research Grant.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Data Availability Statement

Data will be made publicly available on GitHub upon acceptance.

References

Aizenman, A. M., Gegenfurtner, K. R., & Goettker, A. (2024). Oculomotor routines for perceptual judgments. *Journal of Vision*, *24*(5), 3. https://doi.org/10.1167/jov.24.5.3

Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, *50*(5), 791–798.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3–4), 191–204. https://doi.org/10.1016/0025-5564(75)90075-9

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

Buelte, D., Meister, I. G., Staedtgen, M., Dambeck, N., Sparing, R., Grefkes, C., & Boroojerdi, B. (2008). The role of the anterior intraparietal sulcus in crossmodal processing of object features in humans: An rTMS study. *Brain Research*, *1217*, 110–118. https://doi.org/10.1016/j.brainres.2008.03.075

Dowell, C. J., Norman, J. F., Moment, J. R., Shain, L. M., Norman, H. F., Phillips, F., & Kappers, A. M. L. (2018). Haptic shape discrimination and interhemispheric communication. *Scientific Reports*, *8*(1), 377. https://doi.org/10.1038/s41598-017-18691-2

Driver, J., & Spence, C. (1998). Cross–modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1319–1331.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18–18.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Federico, G., & Brandimonte, M. A. (2019). Tool and object affordances: An ecological eye-tracking study. *Brain and Cognition*, *135*, 103582. https://doi.org/10.1016/j.bandc.2019.103582

Gaissert, N., Wallraven, C., & Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision*, *10*(11), 2–2.

Gomez, M. A., & Snow, J. C. (2017). Action properties of object images facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1115–1124. https://doi.org/10.1037/xhp0000390

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*, 1–8.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*(4), 289–303. https://doi.org/10.3758/BF03202498

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in psychtoolbox-3?*, Perception, *36*, 1–16.

Kümmerer, M., Theis, L., & Bethge, M. (2014). *Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet*. arXiv Preprint arXiv:1411.1045.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). DeepGaze II: Reading fixations from deep features trained on object recognition. *arXiv Preprint arXiv:1610.01563*.

Lacey, S., & Campbell, C. (2006). Mental representation in visual/haptic crossmodal memory: Evidence from interference effects. *Quarterly Journal of Experimental Psychology*, *59*(2), 361–376.

Lacey, S., Peters, A., & Sathian, K. (2007). Cross-modal object recognition is viewpoint-independent. *PLoS One*, *2*(9), Article e890. https://doi.org/10.1371/journal.pone.0000890

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*(3), 342–368. https://doi.org/10.1016/0010-0285(87)90008-9

Lederman, S. J., & Klatzky, R. L. (1997). Relative availability of surface and object properties during early haptic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(6), 1680.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16.

Metzger, A., Ennis, R. J., Doerschner, K., & Toscani, M. (2024). Perceptual task drives later fixations and long latency saccades, while early fixations and short latency saccades are more automatic. *Perception*, *53*(8), 501–511. https://doi.org/10.1177/03010066241253816

Metzger, A., & Toscani, M. (2022). Unsupervised learning of haptic material properties. *eLife*, *11*, Article e64876. https://doi.org/10.7554/eLife.64876

Metzger, A., Toscani, M., Akbarinia, A., Valsecchi, M., & Drewing, K. (2021). Deep neural network model of haptic saliency. *Scientific Reports*, *11*(1), 1395. https://doi.org/10.1038/s41598-020-80675-6

Metzger, A., Toscani, M., Valsecchi, M., & Drewing, K. (2019). Dynamics of exploration in haptic search. *2019 IEEE World Haptics Conference (WHC)*, 277–282. https://doi.org/10.1109/WHC.2019.8816174

Metzger, A., Toscani, M., Valsecchi, M., & Drewing, K. (2020). Foveation-like behavior in human haptic search. *Journal of Vision*, *20*, 1105.

Metzger, A., Toscani, M., Valsecchi, M., & Drewing, K. (2021). Target Search and Inspection Strategies in Haptic Search. *IEEE Transactions on Haptics*, *14*(4), 804–815. https://doi.org/10.1109/TOH.2021.3076847

Morash, V. S. (2016). Systematic movements in haptic search: Spirals, zigzags, and parallel sweeps. *IEEE Transactions on Haptics*, *9*(1), 100–110. https://doi.org/10.1109/TOH.2015.2508021

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*(1), 37–42. https://doi.org/10.1111/1467-9280.00307

Norman, J. F., Dukes, J. M., Shapiro, H. K., Sanders, K. N., & Elder, S. N. (2021). Temporal integration in the perception and discrimination of solid shape. *Attention, Perception, & Psychophysics*, *83*(2), 577–585. https://doi.org/10.3758/s13414-020-02031-0

Norman, J. F., Norman, H. F., Clayton, A. M., Lianekhammy, J., & Zielke, G. (2004). The visual and haptic perception of natural object shape. *Perception & Psychophysics*, *66*(2), 342–351. https://doi.org/10.3758/BF03194883

Overvliet, K., Smeets, J. B., & Brenner, E. (2008). The use of proprioception and tactile information in haptic search. *Acta Psychologica*, *129*(1), 83–90.

Plaisier, M. A., Bergmann Tiest, W. M., & Kappers, A. M. (2009). Salient features in 3-D haptic shape perception. *Attention, Perception, & Psychophysics*, *71*(2), 421–430.

Plaisier, M. A., & Kappers, A. M. (2010). Cold objects pop out! In *International conference on human haptic sensing and touch enabled computer applications* (pp. 219–224). Springer Berlin Heidelberg.

Plaisier, M. A., Tiest, W. M. B., & Kappers, A. M. (2008). Haptic pop-out in a hand sweep. *Acta Psychologica*, *128*(2), 368–377.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, *7*(14), 16. https://doi.org/10.1167/7.14.16

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 9.

Spence, C., Nicholls, M. E., Gillespie, N., & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between

touch, audition, and vision. *Perception & Psychophysics*, *60*(4), 544–557.

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(4), 1298.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, *147*, 249–262.

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013a). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences*, *110*(27), 11163–11168. https://doi.org/10.1073/pnas.1216954110

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013b). Selection of visual information for lightness judgements by eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130056.

Toscani, M., Yücel, E. I., & Doerschner, K. (2019). Gloss and speed judgments yield different fine tuning of saccadic sampling in dynamic scenes. *I-Perception*, *10*(6), Article 2041669519889070.

Toscani, M., Zdravković, S., & Gegenfurtner, K. R. (2016). Lightness perception for surfaces moving through different illumination levels. *Journal of Vision*, *16*(15), 21–21.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, *13*(4), 428–432. https://doi.org/10.1016/S0959-4388(03)00105-3

Ueda, Y., & Saiki, J. (2007). Viewpoint independence in visual and haptic object recognition. *The Japanese Journal of Psychonomic Science*, *26*(1), 11–19.

Ueda, Y., & Saiki, J. (2012). Characteristics of eye movements in 3-D object learning: Comparison between within-modal and cross-modal object recognition. *Perception*, *41*(11), 1289–1298. https://doi.org/10.1068/p7257

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.