Cognition 254 (2025) 105968

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Visual mental imagery of nonpredictive central social cues triggers automatic attentional orienting

Shujia Zhang [a,b], Li Wang [a,b,*], Yi Jiang [a,b]

[a] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[b] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Social attentional orienting
Visual mental imagery
Eye gaze
Biological motion
Arrow

ABSTRACT

Previous research has demonstrated that social cues (e.g., eye gaze, walking direction of biological motion) can automatically guide people's focus of attention, a well-known phenomenon called social attention. The current research shows that voluntarily generated social cues via visual mental imagery, without being physically presented, can produce robust attentional orienting similar to the classic social attentional orienting effect. Combining a visual imagery task with a dot-probe task, we found that imagining a non-predictive gaze cue could orient attention towards the gazed-at hemifield. Such attentional effect persisted even when the imagery gaze cue was counter-predictive of the target hemifield, and could be generalized to biological motion cue. Besides, this effect could not be simply attributed to low-level motion signal embedded in gaze cues. More importantly, an eye-tracking experiment carefully monitoring potential eye movements demonstrated the imagery-induced attentional orienting effect induced by social cues, but not by non-social cues (i.e., arrows), suggesting that such effect is specialized to visual imagery of social cues. These findings accentuate the demarcation between social and non-social attentional orienting, and may take a preliminary step in conceptualizing voluntary visual imagery as a form of internally directed attention.

1. Introduction

Being social creatures, humans are always ready to detect interactive social partners' focus of attention via various social cues (e.g., eye gaze and walking direction). This exquisite ability, known as social attention, is fundamental to social interaction and adaptive functions because it enables us to learn about others' inner states and where the important events occur in the environment (Birmingham & Kingstone, 2009; Nummenmaa & Calder, 2009; Shi et al., 2010). In order to characterize the properties of social attention, Friesen and Kingstone (1998) introduced a central cueing task modified from classic Posner cueing paradigm (Posner, 1980). Typically, a gaze cue is centrally presented, followed by a target appeared on either side of screen. People generally make faster response when the target appears on the gazed-at hemifield, reflecting a gaze cueing effect. Interestingly, although the gaze cue appears in the central location and the gaze-triggered attentional effect persists over a long interval (Frischen et al., 2007; Liu et al., 2021), much like endogenous attention, it also shares many properties such as reflexivity with exogenous attention (Driver et al., 1999; Friesen & Kingstone, 2003; Friesen et al., 2004; Langton & Bruce, 1999; Tipples, 2008). It has been argued that social attention might be unique and beyond the classic dichotomy of covert attention (i.e., endogenous and exogenous), hewing out a new direction for visual attention research (Frischen, Bayliss, et al., 2007; Ji et al., 2020; Wang et al., 2020).

For most of the previous research, social cues were externally presented and perceived by observers. However, we humans do not always focus on events in the here and now, and we spend nearly half of our time in self-generated thoughts about what is not going on around us (Killingsworth & Gilbert, 2010). Therefore, the potential effect of social cue representation generated in our inner mind's world (i.e., via voluntary visual mental imagery) makes an interesting question. Voluntary visual imagery refers to the ability of constructing mental representations in the absence of these corresponding stimuli. It often produces a subjective feeling of 'seeing', but is much vaguer and involves certain different mechanisms compared with visual perception (Cole et al., 2022; Koenig-Robert & Pearson, 2021; Pylyshyn, 2003). Despite these differences, voluntary imagery demonstrated a functional similarity with visual perception in many aspects (Finke, 1980, 1989; Grush, 2004; Pearson, 2019; Pearson et al., 2008; Tartaglia et al., 2009) including guiding attention (e.g., Cochrane & Milliken, 2019; Cochrane,

---

* Corresponding author at: State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.
E-mail address: wangli@psych.ac.cn (L. Wang).

https://doi.org/10.1016/j.cognition.2024.105968
Received 18 December 2023; Received in revised form 18 July 2024; Accepted 24 September 2024

Nwabuike, et al., 2018; Cochrane et al., 2019; Cochrane, Wang, et al., 2021; Cochrane, Zhu, et al., 2018; Liao et al., 2023). For instance, merely imagining the search process for a certain target could improve subsequent attentional selection for this target (Reinhart et al., 2015). Other studies also suggest a modulation effect of imagery cues on visual search when the search item matched the imagined cue (Cochrane, Townsend, et al., 2021; Cochrane, Wang, et al., 2021). Even a task-irrelevant but imagery-matching content could attract attention, leading to facilitation effects in visual search (Liao et al., 2023; Moriya, 2018) and attentional blink in rapid serial visual presentation paradigm (Pashler & Shiu, 1999). These pioneers unanimously demonstrate that imagery is as powerful as visual perception in guiding attention, but whether social attention, which is beyond the traditional dichotomy of exogenous and endogenous attention, could occur within the realm of visual imagery remains unclear. As a form of sensory simulation, visual mental images are widely involved in many cognitive processes (e.g., visual working memory) and daily behaviors (e.g., spatial navigation). The internal representation of social cues and the potential attentional effect guided by such representation may help people prepare for upcoming social interaction and understand others' intentions better. Investigating social attention via imagery can be an important step for understanding the role of social cues in daily socio-cognitive behaviors.

In the current study, we directly assessed the imagery-induced social attention by combining a visual mental imagery task with a dot-probe task. Participants were required to imagine a leftward or rightward eye gaze, and then discriminate the hemifield of a target flashed at either the left or right side of the screen. We went further by adopting a counter-predictive design (75% incongruent) to see whether the attentional orienting effect, if observed, was to some degree reflexive and automatic. Besides, the walking direction of biological motion (BM) was also used as the imagery content to see if any observed effect could be generalized to another type of social cue. We also investigated whether low-level motion signal embedded in dynamic eye gaze, or any semantic directional information, would confound above effects. Finally, to investigate whether any observed attentional effect was specific to social cues, we employed a non-social arrow cue and compared its effect with that induced by the gaze cue in a within-subject manner. Previous research has suggested that non-social arrow cue and social gaze cue are quantitatively similar in guiding attention, while whether these two are qualitatively different remained an open-to-debate question (for a recent review, see Chacon-Candia et al., 2023). Comparing social and non-social cues in the realm of visual imagery might be an innovative way to elucidate the longstanding dispute on the specificity of social attention (Friesen et al., 2004; Frischen, Bayliss, et al., 2007; Ristic et al., 2007).

2. Method

2.1. Participants

A prior power analysis (paired samples *t*-test) using G*Power in Version 3.1.9.7 (Faul et al., 2007) suggested that at least 15 participants would afford 80% power with alpha at 0.05 (two-tailed) to detect a high imagery-induced attentional effect (cohen's $d = 0.8$) which was observed in previous research (Cochrane, Townsend, et al., 2021). This effect size was also comparable with previous research adopting similar dot-probe task to detect attentional effect (Hietanen et al., 2006; Ji et al., 2020; Ji et al., 2022; Shi et al., 2010). We further increased the sample size to 20 per experiment to adequately detect the potential effects in the current study.

A total of 120 naïve adults aged between 20 and 31 years old ($M \pm SD = 23.19 \pm 1.95$) took part in the current study. The current research consisted of 6 experiments (i.e., Experiments 1, 2a, 2b, 3, 4, and 5) and each experiment included 20 participants (10 females). All participants had self-reported normal or corrected-to-normal vision and they all gave written informed consent in accordance with procedure and protocols approved by the institutional review board of the Institute of Psychology, Chinese Academy of Sciences. Participants received monetary compensation to appreciate their participation.

2.2. Apparatus and stimuli

Participants were seated on a height-adjustable chair approximately 57 cm from a 24-in. LCD monitor (1920 × 1080 pixels, 60 Hz). Their heads were stabilized by a chin and head rest. Stimuli were generated and presented using Psychtoolbox extensions, version 3 for MATLAB (the MathWorks) (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). In Experiments 1 and 2, a female face exemplar with neutral expression was chosen from the NimStim Set of Facial Expressions (Tottenham et al., 2009). This face image was first manipulated in a face-image modeling software (i.e., FaceGen Modeller, Version 3.4) to extract facial features, and these features were fused on the built-in head template of the software. Then, using Adobe Photoshop software, we converted the face image to greyscale and cropped it with an elliptic tool (about $4.2^\circ \times 6.1^\circ$) to remove the features outside it. The gaze direction (leftward or rightward) was generated by moving the irises and pupils of both eyes to the canthi. In Experiment 3, the to-be-imagined BM exemplar was adopted from Vanrie and Verfaillie (2004). This BM sequence (subtended approximately $3.6^\circ \times 6.1^\circ$) was created by capturing the motion of a walking actor and consisted of 13 white point-light dots which depicted the motions of head and major joints including shoulders, elbows, wrists, hips, knees, and ankles. In Experiment 4, the to-be-imagined exemplar was a stationary white dot ($0.6^\circ \times 0.6^\circ$, RGB: 255, 255, 255). In Experiment 5, participants were required to imagine a face with averted eye gaze (gaze block) or a double-headed arrow (arrow block). Their eye movements were recorded at a sampling rate of 1000 Hz using Eyelink Portable Duo (SR Research, Canada) from the beginning of imagery until response was made in the dot-probe task. The face exemplar used in gaze block was the same with Experiments 1 and 2. For arrow block, the to-be-imagined exemplar was a double-headed arrow filled with texture (subtended approximately $3.4^\circ \times 2.0^\circ$, see Fig. 1). The Gabor patch subtended $1.2^\circ \times 1.2^\circ$ in visual angle (Gussian *SD* of the window = 4 pixels, contrast = 1, initial phase = 0°, spatial frequency = 8.33 cycles/degree) was used as a target in all experiments. All stimuli were presented on a black background (RGB: 0, 0, 0), and a fixation cross ($0.5^\circ \times 0.5^\circ$, RGB: 255, 255, 255) was centrally presented to help participants maintain central fixation.

2.3. Procedure and design

2.3.1. Experiment 1

Firstly, all participants filled in The Vividness of Visual Mental Imagery Questionnaire (VVIQ; Marks, 1973) where they rated the vividness of their subjectively experienced visual images. VVIQ has been verified to have fairly acceptable reliability and validity (McKelvie, 1995) and has been widely used in imagery research. This 16-item scale is summarized in a vividness score between 16 and 80 for each participant, where a score of 16 indicates extremely high vividness and 80 indicates low vividness.

Before the formal experiment, there was a familiarization session which included four trials of self-paced familiarization of the to-be-imagined female face exemplar, four trials of self-paced imagery practice, and two practicing trials. Previous research on the topic of voluntary imagery has adopted similar procedures to ensure that participants formed perceptually detailed memories of the to-be-imagined stimuli and fully understood the experiment instructions (e.g., Ganis & Schendan, 2008; Lee et al., 2012; Xie et al., 2020). Then in the subsequent formal experiment, participants could selectively retrieve information from the stored memory to voluntarily generate the mental image for the cued stimulus. At the beginning of the familiarization trials, participants were given both written and oral instructions to carefully inspect and memorize the stimuli presented on the screen for as long as they needed

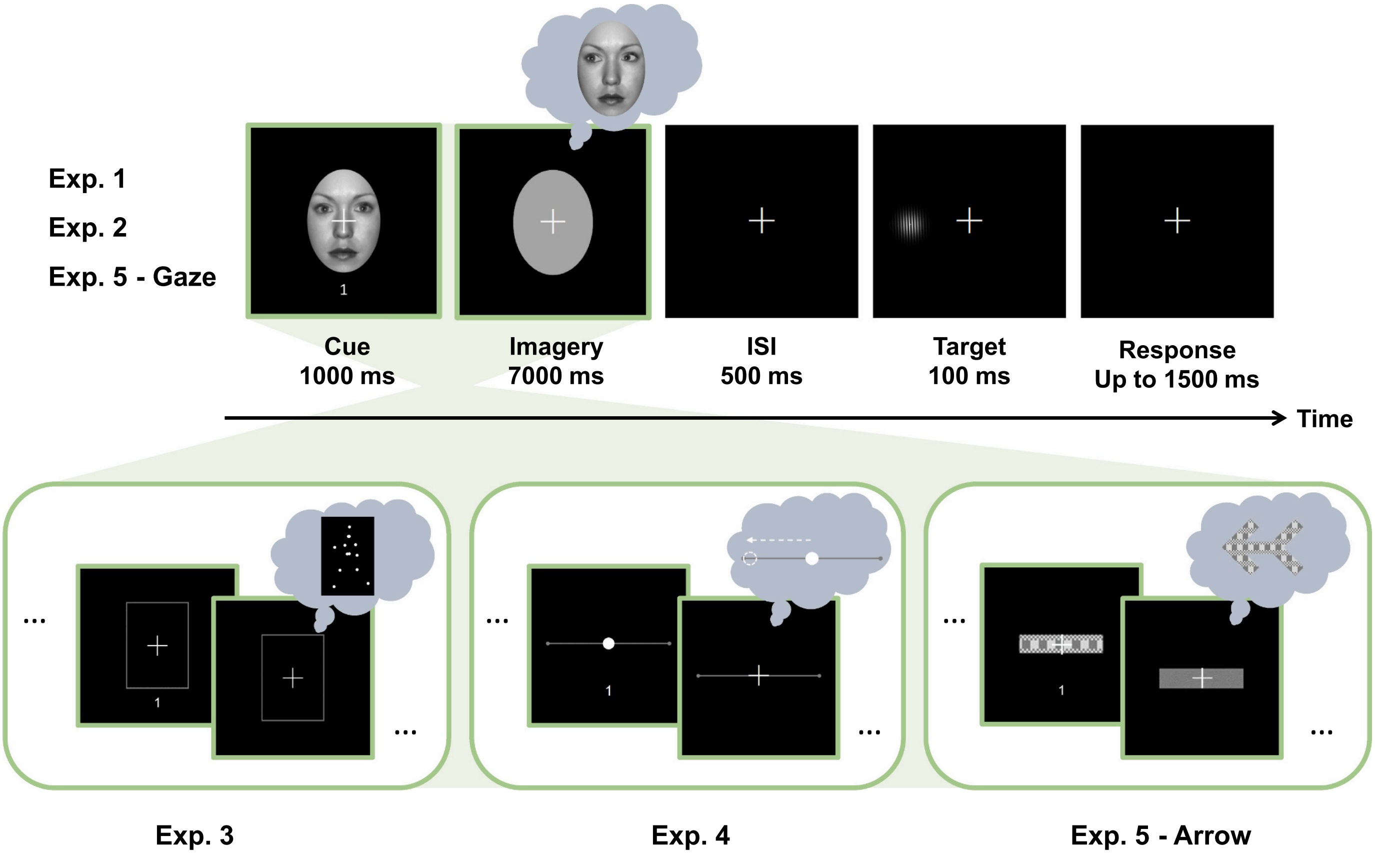

Fig. 1. The main procedure illustration of Experiments 1–5, taking numeric cue '1' representing 'left' as an example. In Experiments 1 and 2, participants were instructed to voluntarily imagine a female face with leftward or rightward eye gaze according to a numeric cue ('1' or '2'). Then after a blank inter-stimulus interval (ISI), they should respond to the location of a peripheral target by pressing assigned keys.

because they would be required to imagine them from memory in the subsequent experiment. For each time, participants were shown the numeric cue ('1' or '2') and the to-be-imagined female face with straight eye gaze (the same with cue display in formal experiment, see Fig. 1), followed by the corresponding image of female face with averted eye gaze (leftward or rightward). Participants observed for as long as they needed while maintaining central fixation, and pressed space key to move on to the next time of familiarization. After familiarization trials were the self-paced imagery practice trials, where participants learned to generate mental images while keeping central fixation. For each trial, participants saw the female face with straight eye gaze and the numeric cue below the face for 1000 ms, followed by a gray ellipse of the same size with the female face. Participants were instructed to maintain central fixation and imagine the female slowly turning her eyes avertedly as vividly as possible. They were encouraged to imagine for approximately seven seconds, and press space key to indicate they had formed vivid image in mind. Then, the actual image would be presented on screen so that participants could compare their visual image with the actual stimulus. Participants were encouraged to notice differences between the mental and actual images, and they were free to adjust their internal images for as long as they wanted. After these familiarization and imagery practice trials, participants received both written and oral instructions for the formal experiment, retold the instructions to ensure the correct understanding of their task, and conducted two practicing trials to adapt to the task in the formal experiment. It should be mentioned that averted eye gaze cues would not physically appear in the formal experiment, which was different from visual working memory paradigm, where the key stimulus would appear in, always at the beginning of each trial, and then disappear. Besides, although we showed participants the images of averted eye gaze cues during the familiarization, the formal experiment was delayed for at least 5 min since the last exposure to a physically presented averted eye gaze image. This temporal delay is much longer than the memory duration of visual sensory memory (about 1 s or less than that; Sperling, 1960) and visual working memory (< 1 min; Atkinson & Shiffrin, 1971; Baddeley, 1990; Sakai, 2017; Shiffrin & Atkinson, 1969). These aspects in the laboratory paradigm differentiate voluntary imagery from visual working memory from the perspective of operational definition.

In the formal experiment, participants were instructed to complete an imagery task followed by a dot-probe task. An example trial of Experiment 1 is illustrated in Fig. 1A. In the cue display, a to-be-imagined female face image (about 4.2° × 6.1°) was centrally presented for 1000 ms, together with a numeric cue ('1' or '2') below it. A gray ellipse of the same size was then shown for 7000 ms, during which participants were instructed to imagine the female turning her eyes avertedly towards left or right according to the numeric cue (e.g., '1' for 'left' and '2' for 'right'). The specific meaning of numeric cue was counterbalanced between participants. We set the imagery duration to be 7000 ms based on previous imagery literatures (e.g., Chang et al., 2013; Keogh & Pearson, 2017; Koenig-Robert & Pearson, 2019; Pearson et al., 2008). Participants were required to hold their imagery within the range of gray ellipse until the gray ellipse disappeared, which meant the end of imagery period. After an inter-stimulus interval (ISI) of 500 ms, a Gabor patch (1.2° × 1.2°) as a target flashed for 100 ms on either of the left or right visual field at a distance of 4.5° from the fixation. Participants were asked to localize the target by pressing the arrow keys on the keyboard, i.e., pressing the left arrow key if the target flashed on the left visual field and vice versa. They had to make responses as quickly as possible on the premise of avoiding making mistakes. The current study used this localization task rather than a detection task or a discrimination task mainly for two reasons. First, among these tasks, the detection task has the shortest response times (RTs), the discrimination task has the longest RTs, and the localization task was in the middle (Friesen & Kingstone, 1998), suggesting that the localization task has the moderate difficulty. Moreover, these three tasks trigger different magnitudes of the gaze cueing effect, with the localization task triggering a larger gaze cueing effect than the detection task (McKay et al., 2021) or the discrimination task (Chen et al., 2021). In short, compared with the other two tasks, the localization task has the moderate difficulty and can trigger the largest gaze cueing effect.

The whole experiment consisted of 120 trials, taking about 25 min to complete. Participants were allowed to take a short break approximately per 7 min. Half of the trials were congruent, which meant the target hemifield was congruent with the imagery-cued hemifield. And the other half of trials were incongruent, namely the target hemifield and the imagery-cued hemifield were opposite to each other. Imagery-cued hemifields (left vs. right) were balanced in congruent and incongruent conditions. Another five catch trials, in which participants were required to indicate the direction of the imaged gaze (left or right), were interspersed in the experiment to assess whether participants did make efforts to generate mental images according to the cue. It's also worth noting that before starting the formal experiment, participants were told to fixate on the central cross through the whole experiment. Previous studies utilizing eye-movement recording techniques demonstrated that participants did have the ability to keep central fixation during imagery task (Gregori Grgic et al., 2016), even if they were not required to do so (Laeng & Teodorescu, 2002). Besides, all participants were explicitly informed that the imagined gaze direction couldn't predict the subsequent target hemifield.

2.3.2. *Experiment 2*

The procedure and design of Experiments 2a and 2b were basically the same as those of Experiment 1 except that 75% trials were incongruent and 25% trials were congruent, resulting in counter-predictive design. Participants in Experiment 2a were not informed of the counter-predictivity of imagery cue whereas participants in Experiment 2b were explicitly informed before the experiment.

2.3.3. *Experiment 3*

The procedure and design of Experiment 3 were the same as those of Experiment 1 except for the imagery stimuli. The cue display of this experiment consisted of a central fixation cross, a light gray frame (4.4° × 6.4°, RGB: 20, 20, 20), and a numeric cue. Participants were asked to imagine a point-light BM walker walking towards left or right according to numeric cue within the light gray frame. Before the formal experiment, a familiarization session was also conducted, which ensured that participants had perceptually detailed memories of the to-be-imagined BM exemplar and fully understand the experiment instructions.

2.3.4. *Experiment 4*

In the cue display of Experiment 4, participants were presented with the to-be-imagined white dot placed at the center of a gray line (6° in width, RGB: 128, 128, 128) with a numeric cue below it. Participants were required to imagine the dot slowly moving along the gray line presented on the screen and hold this scene in imagery when the dot reached the end of line. In the familiarization session, the moving speed of the white dot exemplar was 0.6°/s. Other aspects of Experiment 4 were the same as Experiment 1.

2.3.5. *Experiment 5*

Experiment 5 adopted a within-participant design to compare the attentional effect induced by social cue (i.e., eye gaze) and non-social cue (i.e., arrow). This experiment was divided into two blocks. In the gaze block, participants were instructed to imagine the female face with averted eye gaze, without the requirement of imagining the dynamic process of turning eyes avertedly. In the arrow block, participants were asked to imagine a double-headed arrow (subtended approximately 3.4° × 2.0°) filled with grid texture. Compared with Experiments 1–4, Experiment 5 had the following modifications. First, participants' eye movements were recorded from the beginning of imagery until the response was made in the dot-probe task. Recordings were used to exclude trials where participants failed to maintain central fixation. Second, catch trials were replaced with trial-by-trial imagery vividness rating and effortness rating. After the completion of dot-probe task, participants first reported the vividness of their imagery on that trial on a 4-point scale by pressing corresponding keys (1 = *almost no imagery*, 2 = some *weak imagery*, 3 = *moderate imagery*, 4 = *strong imagery almost like perception*), and then reported their subjective impression of the efforts they exerted to form their mental images again on a 4-point scale (1 = *almost no effort*, 2 = *some effort*, 3 = *moderate effort*, 4 = *tried very hard to generate a mental image*). Previous research has shown that people have a good metacognition on their own mental imagery, and can reliably evaluate the vividness of single episode of imagination (Pearson et al., 2011). Each block contained 120 trials with the same design as Experiment 1, and took approximately 45 min to complete. The sequence of gaze block and arrow block was counterbalanced between participants. Participants were allowed to take a short break approximately per 7 min, and when one block was finished, participants would take a break for at least 30 min before starting the next.

In order to further probe whether and what participants imagined during experiments, we asked participants to complete a post-experiment task, which consisted of a short-answer question plus a comparison task (referred from Ishai et al., 2000; Moriya, 2018; Xie et al., 2020). The short-answer question asked participants to describe the mental images they generated during the gaze and arrow blocks. And the comparison task presented four face/arrow images, which included the imagery exemplars and three new face/arrow images differed from each other in identity. Participants were instructed to select the one that is more closely matched the imagined exemplar.

2.4. *Transparency and openness*

We report how we determined the sample size, the data exclusion, all manipulations, and all measures in the study. Data were analyzed using SPSS Version 26.0.0 and G-power Version 3.1. The original data and analytic codes for the current study are publicly accessible at Science Data Bank (ScienceDB) and can be accessed at https://doi.org/10.57760/sciencedb.o00115.00100. Materials used in the current research are widely available. There is not a preregistration for the current study.

3. Results

3.1. *Data exclusion*

For Experiments 1–4, all participants accomplished all trials. For Experiment 5, the last one gaze-cue trial for one participant and the last four arrow-cue trials for another participant were omitted due to technical problem. For further analysis, trials with incorrect responses and RTs less than 100 ms or more than 1000 ms were excluded. The percentage of trials excluded was 0.58% in Experiment 1, 0.50% in Experiment 2a, 0.62% in Experiment 2b, 0.79% in Experiment 3, 0.62% in Experiment 4, and 2.27% in Experiment 5 (2.42% for gaze block and 2.13% for arrow block).

In Experiment 5, we further excluded the trials where participants failed to maintain central fixation. We parsed eye fixations from the eye-movement recordings based on the default criteria of Eyelink. For each sample, velocity and acceleration are computed and compared against the default threshold (30°/sec and 8000°/sec^2, respectively). Samples were labeled as fixational sample if neither the velocity nor the acceleration were above the threshold. Fixational samples with horizontal position less than 1° deviation from the central fixation was further labeled as central fixational sample. We first excluded the trials whose ratio of central fixational samples to all fixational samples was less than 90%, basically ensuring participants maintained central fixation most of the time. Then, as central fixation is particularly emphasized for the dot-probe task, we inspected the fixation position trace from the end of imagery until the response was made in dot-probe task, and excluded trials identified with non-central fixational samples during this period. Individual traces are publicly accessible at https://doi.org/10.57760/sciencedb.o00115.00100. The overall percentage of trials excluded was 5.74%, with 4.84% in gaze block and 6.64% in arrow block, respectively.

3.2. VVIQ

The range of participants' VVIQ scores was 19 to 60 in Experiment 1, 21 to 43 in Experiment 2a, 19 to 43 in Experiment 2b, 22 to 41 in Experiment 3, 18 to 52 in Experiment 4, and 18 to 46 in Experiment 5. And the average VVIQ score was 34.75 ± 9.18 ($M \pm SD$) in Experiment 1, 31.60 ± 5.64 in Experiment 2a, 33.25 ± 6.00 in Experiment 2b, 33.05 ± 4.98 in Experiment 3, 33.10 ± 9.45 in Experiment 4, and 31.30 ± 9.06 in Experiment 5. One-way analysis of variance (ANOVA) didn't find statistically significant difference of participants' imagery capability across six experiments, $F(5, 114) = 0.54$, $p = 0.746$, $\eta_p^2 = 0.02$. Therefore, the discrepancy of attentional effects induced by different kinds of imagined cues was not due to group difference in the ability of generating mental images.

3.3. Catch trials

For Experiments 1–4, five catch trials were interspersed in the experiment. The overall performance in catch trials was generally high ($M \pm SE = 92.2\% \pm 1.1\%$), suggesting that participants did generate imagery according to the numeric cue. Specifically, average accuracy for catch trials was 92% in Experiment 1, 89% in Experiment 2a, 93% in Experiment 2b, 95% in Experiment 3, and 92% in Experiment 4. One-way ANOVA found no significant difference across five experiments, $F(4, 95) = 0.83$, $p = 0.508$, $\eta_p^2 = 0.03$.

3.4. Attentional effect analysis

All participants achieved high detection accuracy in the dot-probe task, as the accuracy was 99.6% ± 0.2% ($M \pm SE$) in Experiment 1, 99.6% ± 0.2% in Experiment 2a, 99.7% ± 0.1% in Experiment 2b, 99.5% ± 0.2% in Experiment 3, 99.6% ± 0.2% in Experiment 4, and 99.0% ± 0.2% in Experiment 5 (99.1% ± 0.2% in the gaze block and 98.8% ± 0.2% in the arrow block).

RTs of Experiment 1 were entered into paired samples t-test with the congruency (congruent vs. incongruent) as the independent variable. Results revealed that participants made faster responses under congruent condition ($M \pm SE = 408.46 \pm 12.53$ ms) compared with incongruent condition ($M \pm SE = 418.86 \pm 13.77$ ms), $t(19) = 3.59$, $p = 0.002$, cohen's $d = 0.80$, 95% confidence interval for cohen's d (95% CI_d) = [0.29, 1.30] (Fig. 2A), indicating a significant attentional effect. In other words, though not physically presented, the mental representation of eye gaze cue generated via visual imagery induced attentional orienting effect.

Further, in Experiment 2, we would like to examine whether this imagery-induced gaze cueing effect was automatic and reflexive, which is the property of classic gaze cueing effect (Driver et al., 1999; Friesen et al., 2004; Friesen & Kingstone, 1998; Tipples, 2008). To achieve this, we adopted counter-predictive design which has been generally used to measure the automaticity of attentional orienting (e.g., Driver et al., 1999; Friesen et al., 2004; Tipples, 2008). Specifically, counter-predictive design means a target is much more likely to appear on the opposite side of the gaze-cued visual field, making it strategically more beneficial to avoid orienting attention towards the direction of imagined gaze cue. Therefore, faster responses at the direction of the imagined movement of the eyes (i.e., congruent trials), even if the target appearing there was below chance levels, could be thought to reflect the automaticity of attentional orienting. In Experiment 2a, a significant attentional orienting effect was still observed even when the imagined cues were counter-predictive of the target hemifield (congruent: $M \pm SE = 401.19 \pm 11.64$ ms, incongruent: $M \pm SE = 408.11 \pm 11.58$ ms; $t(19) = 4.39$, $p < 0.001$, cohen's $d = 0.98$, 95% $\mathrm{CI}_d = [0.44, 1.51]$; see Fig. 2B). In Experiment 2b, participants were explicitly informed that the target would appear in the hemifield opposite to that indicated by imagery cue on 75% of trials. Strikingly, imagery-induced gaze cueing effect persisted (congruent: $M \pm SE = 429.22 \pm 17.09$ ms, incongruent: $M \pm SE = 435.14 \pm 18.41$ ms; $t(19) = 2.11$, $p = 0.048$, cohen's $d = 0.47$, 95% $\mathrm{CI}_d = [0, 0.93]$; see Fig. 2C). These two experiments (Experiments 2a and 2b) together disclosed the reflexive nature of imagery-induced gaze cueing effect.

To directly compare the magnitude of attentional effects observed across the above three experiments, we calculated the normalized

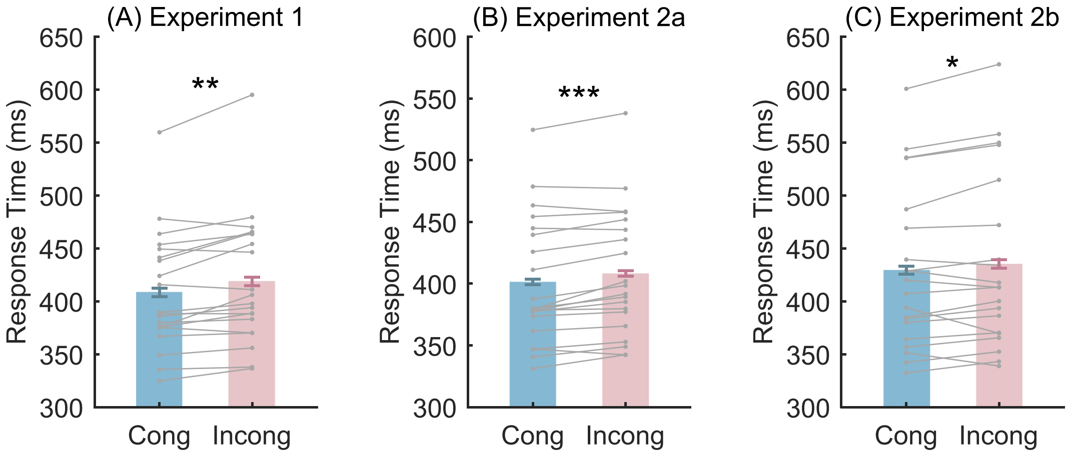

Fig. 2. Results of Experiments 1, 2a, and 2b. (A) Experiment 1: non-predictive gaze cue. Voluntarily imagining eye gaze cues that were non-predictive of the target hemifield triggered attentional orienting effect. (B) Experiment 2a: non-informed counter-predictive gaze cue. Attentional orienting was also observed when the imagined eye gaze cues were counter-predictive (75% incongruent) of the target hemifield, indicating that imagery-induced gaze cueing is reflexive. (C) Experiment 2b: informed counter-predictive gaze cue. Attentional orienting still occurred when participants were explicitly informed of the counter-predictive design, further confirming the reflexivity of imagery-induced gaze cueing effect. 'Cong' refers to congruent condition, and 'Incong' refers to incongruent condition. Error bars indicate 95% within-subject confidence intervals following Cousineau-Morey corrections. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

attention effect index produced by dividing the difference in the mean RT obtained under the incongruent condition versus that in the congruent condition by their sum, expressed as ($\mathrm{RT}_{\mathrm{incongruent}}$ - $\mathrm{RT}_{\mathrm{congruent}}$)/($\mathrm{RT}_{\mathrm{incongruent}}$ + $\mathrm{RT}_{\mathrm{congruent}}$). A subsequent one-way ANOVA with experiment as the independent variable suggested that the attentional effects were comparable across the three experiments, $F(2, 57) = 1.10$, $p = 0.340$, $\eta_p^2 = 0.04$. We also calculated Bayes Factor to provide complementary evidence for null hypothesis H0. Results showed that H0 is about 3.3 times more likely than the alternative H1, $BF_{01} = 3.305$. Further, independent samples *t*-test showed no statistical difference between Experiments 2a and 2b, $t(38) = 0.71$, $p = 0.485$, cohen's $d = 0.32$, 95% $\mathrm{CI}_d = [-0.40, 0.84]$, $BF_{01} = 2.658$. These lines of results convergently elucidate that imagined gaze cue triggers attentional orienting effect that is automatic and to some degree immune to top-down cognitive control.

However, whether this effect was restricted to gaze cues, or could be extended to other kinds of social cues, remained unknown. BM walker which conveys critical social information is an ideal candidate for tackling this issue. It has been suggested that walking direction of BM can trigger reflexive attentional orienting effect like eye gaze cue does (Shi et al., 2010; Yu et al., 2020). Therefore, in Experiment 3, we instructed participants to imagine dynamic BM. Similar to Experiment 1, imagined BM cues also elicited an attentional orienting effect (congruent: $M \pm SE = 412.36 \pm 13.48$ ms, incongruent: $M \pm SE = 430.95 \pm 15.36$ ms; $t(19) = 3.65$, $p = 0.002$, cohen's $d = 0.82$, 95% $\mathrm{CI}_d = [0.30, 1.32]$; see Fig. 3A). We also compared the normalized attentional effect observed in the current experiment with that induced by imagined eye gaze cue (Experiment 1). Independent-sample *t*-test showed statistically insignificant difference, $t(38) = 1.39$, $p = 0.172$, cohen's $d = 0.44$, 95% $\mathrm{CI}_d = [-0.19, 1.07]$, $BF_{01} = 1.508$. Together, results of Experiments 1–3 indicated that different types of social cues (i.e., eye gaze and BM walking direction) generated via visual mental imagery consistently gave rise to attentional orienting effect.

Since participants were instructed to imagine the dynamic process that a female turned her eyes avertedly, one may argue that it was the low-level motion signal embedded in the imagery cue that triggered attentional orienting effect. Besides, the semantic directional information conveyed by numeric cues was also likely to cause above attentional effects. These alternations could be basically excluded by Experiment 4 where participants were instructed to imagine a moving dot. Participants' responses showed no statistically difference under congruent condition ($M \pm SE = 413.95 \pm 14.73$ ms) and incongruent condition ($M \pm SE = 410.20 \pm 13.78$ ms), $t(19) = 1.10$, $p = 0.287$, cohen's $d = 0.25$, 95% $\mathrm{CI}_d = [-0.20, 0.69]$, $BF_{01} = 2.541$ (Fig. 3B). These results were in line with previous research, which showed low-level motion signal to be much less powerful as compared to gaze cue when higher-order cognitive function was involved (Gregory & Jackson, 2017; Nie et al., 2018). As a control, the current Experiment 4 was also compared with previous Experiments 1–3. We calculated normalized attentional effects of these experiments and entered them into one-way ANOVA. A significant effect was found, $F(4, 95) = 5.48$, $p < 0.001$, $\eta_p^2 = 0.19$, and post-hoc pairwise comparison (LSD Method) suggested the imagery-based attentional effect in Experiment 4 was smaller than Experiment 1 ($p = 0.005$), Experiment 2a ($p = 0.024$), Experiment 2b ($p = 0.074$), as well as Experiment 3 ($p < 0.001$).

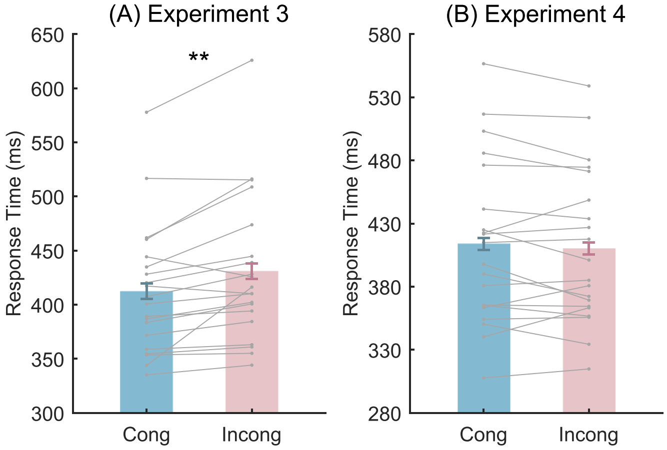

Fig. 3. Results of Experiments 3 and 4. (A) Experiment 3: imagery-induced social attention effect was generalized to the walking direction of biological motion. (B) Experiment 4: the observed imagery-mediated social attention in Experiments 1–3 can't be attributed to low-level motion signal, as the imagery of simple motion failed to induce attentional orienting effect. 'Cong' refers to congruent condition, and 'Incong' refers to incongruent condition. Error bars indicate 95% within-subject confidence intervals following Cousineau-Morey corrections. $^{**}p < 0.01$.

However, Experiment 4 might be argued to differ from other experiments in terms of movement complexity, which then led to difference in imagery vividness and effortness. What's more, whether this effect is specific to cues with social relevance, or can be generalized to non-social cues (e.g., arrow), remains an interesting but unresolved question. Despite lacking social relevance, directional arrow cues are also capable of orienting spatial attention in the modified Posner's central cueing paradigm with non-predictive design (e.g., Tipples, 2002), similar with the social eye gaze cues. This discovery sparked a long-standing debate about whether the attentional effects induced by gaze cues are indeed special compared to those induced by arrow cues (Friesen et al., 2004; Frischen, Bayliss, et al., 2007; Ristic et al., 2007). Hence, researchers have frequently employed eye gaze and arrow cues to ascertain whether social attention is unique and distinct from nonsocial attention, despite the perceptual mismatches of two types of cues.

Therefore, the aim of Experiment 5 was to compare the attentional effects induced by imagined social and non-social cues under more delicate control. We adopted static eye gaze cue and arrow cue in separate experimental blocks. In order to make the comparison more stringent, we monitored participants' eye movements and excluded the trials where participants failed to maintain central fixation. Trial-by-trial imagery vividness and effortness were recorded to see if there were any difference in gaze and arrow blocks. First, we tested the imagery-induced attentional effect in the dot-probe task (Fig. 4). In gaze block, participants made faster responses under congruent condition ($M \pm SE = 463.45 \pm 25.75$ ms) compared with incongruent condition ($M \pm SE = 484.97 \pm 25.11$ ms), $t(19) = 5.59$, $p < 0.001$, cohen's $d = 1.25$, 95% $\mathrm{CI}_d = [0.65, 1.83]$, replicating the imagery-induced attentional orienting. By contrast, imagining arrow cue failed to elicit such effect, as RTs under congruent condition ($M \pm SE = 484.16 \pm 23.23$ ms) and RTs under incongruent condition ($M \pm SE = 484.67 \pm 23.07$ ms) showed no statistically difference, $t(19) = 0.16$, $p = 0.875$, cohen's $d = 0.04$, 95% $\mathrm{CI}_d = [-0.40, 0.47]$, $BF_{01} = 4.255$. Crucially, the interaction between cue type (gaze vs. arrow) and congruency (congruent vs. incongruent) was significant, $F(1, 19) = 22.41$, $p < 0.001$, $\eta_p^2 = 0.54$, highlighting the distinction between the gaze and arrow blocks. Moreover, the magnitude of normalized attentional effect in the gaze block was also larger than that in the arrow block, $t(19) = 4.97$, $p < 0.001$, cohen's $d = 1.11$, 95% $\mathrm{CI}_d = [0.54, 1.66]$, in line with previous analyses. These results together demonstrated the dissociation between social and non-social cues, suggesting such imagery-induced attentional effect is tuned to cues with social relevance.

Besides, to see if there existed any differences in the imagined eye gaze and arrow cues which may confound the results, we compared the trial-by-trial ratings between two blocks. The vividness rating between gaze block ($M \pm SE = 3.11 \pm 0.10$) and arrow block ($M \pm SE = 3.01 \pm 0.14$) was comparable, $t(19) = 0.86$, $p = 0.403$, cohen's $d = 0.19$, 95% $\mathrm{CI}_d = [-0.25, 0.61]$, $BF_{01} = 3.109$. So did effortness rating, with 2.82 ± 0.18 in gaze block and 2.78 ± 0.16 in arrow block, $t(19) = 0.36$, $p =$

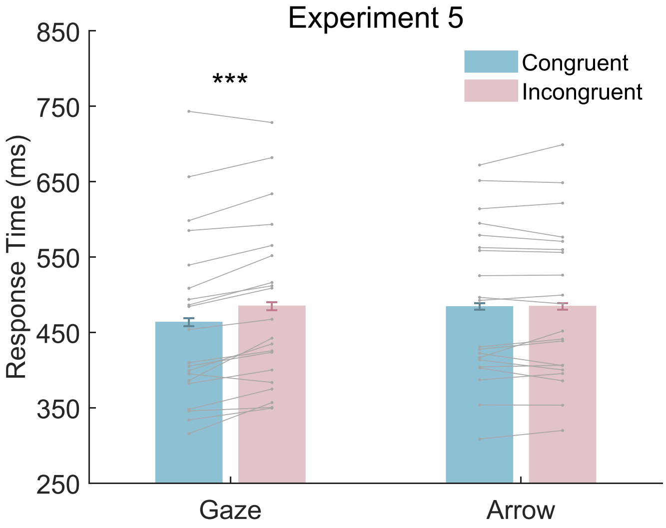

Fig. 4. Results of Experiment 5. Imagery-induced social attention was replicated using gaze cue (the left two bars), but couldn't be generalized to arrow cue (the right two bars). Error bars indicate 95% within-subject confidence intervals following Cousineau-Morey corrections. $^{***}p < 0.001$.

0.725, cohen's $d = 0.08$, 95% $\mathrm{CI}_d = [-0.36, 0.52]$, $BF_{01} = 4.064$. Interestingly, the trial-by-trial vividness rating in the gaze block was significantly correlated with the imagery-based gaze cueing effects, $r = 0.47$, $p = 0.039$ (please see the next section for more details). No correlations were found between the trial-by-trial effortness rating and attentional effects in both the gaze and arrow blocks ($ps > 0.4$), suggesting that participants spared equal efforts in the gaze and arrow blocks, and the imagery effortness had nothing to do with any observed effect. In the post-experiment task, all participants' descriptions were consistent with the experimental requirements and they all correctly chose the imagery exemplar in comparison task, further confirming that all participants formed mental images according to instructions.

3.5. *The correlations between imagery vividness and attentional effects*

If the significant attentional effects observed in the current study were indeed attributed to voluntary visual imagery, a significant correlation between the imagery vividness and the attentional effect was in expectation. The current study adopted two types of imagery vividness measurements. One was the VVIQ, and the VVIQ scores were obtained in all Experiments 1–5. Another was the trial-by-trial vividness rating, which was additionally administrated in only Experiment 5.

The correlations between the VVIQ and the normalized attentional effects for all experiments were illustrated in Table 1. It can be seen that marginally significant correlations were obtained in Experiment 2b ($p = 0.062$) and in Experiment 5-Gaze block ($p = 0.094$). Then we moved on to the trial-by-trial vividness rating in the Experiment 5. For the gaze block where significant attentional effect was observed, trial-by-trial vividness rating was significantly correlated with the normalized attentional effect, $r = 0.47$, $p = 0.039$, suggesting that more vivid mental representation of gaze cue orients attention more efficiently. For the arrow block where no arrow-mediated attentional effect was found, no such a significant correlation was identified, $r = 0.19$, $p = 0.416$. Therefore, the role of visual mental imagery gained strong support from the significant correlation based on the trial-by-trial imagery vividness rating in the Experiment 5-Gaze block, and nuanced support from the marginally significant correlations based on VVIQ in Experiment 2b and Experiment 5-Gaze block.

Table 1
The correlations between VVIQ and the magnitude of normalized attentional effects.

Experiment	Pearson's r	p value
Exp.1	−0.27	0.255
Exp.2a	0.13	0.584
Exp.2b	−0.43	0.062
Exp.3	0.20	0.392
Exp.4	0.21	0.374
Exp.5, Gaze block	−0.38	0.094
Exp.5, Arrow block	−0.16	0.503

Still, it can be seen that though both are measurements of imagery vividness, the trial-by-trial vividness rating is more sensitive than VVIQ in the current study. Indeed, many studies focused on the effect of visual imagery on visual search as well as other perceptual processes also reported significant analytical results using trial-by-trial vividness rating, instead of VVIQ score, as a factor (Cochrane, Townsend, et al., 2021; Cochrane, Wang, et al., 2021; Keogh & Pearson, 2017; Pearson et al., 2011). The reason may be that VVIQ defines an individual's propensity to form visual mental imagery globally, which may lead to decreased sensitivity for the specific cognitive and neural processes underlying the phenomenological experience of vividness occurring within each trial (D'Angiulli et al., 2013). In contrast, the trial-by-trial vividness rating measures the subjective experience at a particular moment (i.e., the current trial), and is structurally constrained by the requirements of the imagery task. Therefore, the trial-by-trial vividness rating is perhaps the most efficient mean by which imagery vividness can be studied (Hertzog & Dunlosky, 2006; Runge et al., 2015; Runge et al., 2017).

4. Discussion

The human ability of automatically shifting attention according to social cues (i.e., social attention) has been a central focus of research in the past decades. Here, we examined whether internally generated social cues, without any external input, could also orient attention towards the cued hemifield. By combining a visual imagery task with a dot-probe task modified from Posner's central cueing paradigm, we found that voluntarily imagining a non-predictive central eye gaze cue triggered attentional orienting to the cued hemifield. This effect still existed when the imagined gaze cue was counter-predictive of the target hemifield, suggesting that the observed attentional effect was reflexive and automatic. Further experiment excluded the alternation of semantic information and low-level motion signals. Importantly, this imagery-mediated social attention could be generalized to another type of social cue, i.e., the walking direction of BM, but vanished when the imagery content was non-social arrow cue. Taken together, these findings demonstrated an automatic attentional orienting effect induced by imagery-generated social cues, mimicking the well-known social attention phenomenon mediated by perceptual social cues.

In the past decades, a series of literatures have demonstrated how the internal representation of visual mental imagery impacted visual attention (e.g., Cochrane et al., 2019; Cochrane, Nwabuike, et al., 2018; Cochrane, Townsend, et al., 2021; Cochrane, Wang, et al., 2021; Cochrane, Zhu, et al., 2018; Liao et al., 2023; Moriya, 2018). The current study is in line with these pioneers on this view with novel evidence. In previous research, the locus of attention was generally tied to the specific location of the imagery-matching stimuli. In contrast, the attentional effects in the current research were specific to the hemifield indicated by the directional information conveyed by imagined social cue. This phenomenon is observed when the imagined contents were irrelevant and even harmful (i.e., counter-predictive of target location) to the external attentional task, highlighting its automaticity. Another remarkable difference of the current study is that whereas previous findings were all illustrations of classic types of attention, our study revealed a new and special form of attention that is beyond classic endogenous and exogenous dichotomy of attention. This exquisite attentional ability helps people to learn about other's inner mental state and key events occur around them (Birmingham & Kingstone, 2009;

Nummenmaa & Calder, 2009; Shi et al., 2010). Such attention orienting observed in the realm of visual mental imagery may help people keep track of others' intentions better and prepare for high-level socio-cognitive behaviors, marking great evolutionary significance.

Besides, our research also evokes further discussion on the conceptualization of visual mental imagery as a prominent form of internally directed attention, which refers to the attention that operates based on the representation of internally activated stimuli (Chun et al., 2011). As an internal mental process, voluntary imagery shares this conception and has been identified with the prominent features of visual attention (e.g., selectivity and limited capacity, see Ceja & Franconeri, 2023; Kalkstein et al., 2011; Keogh & Pearson, 2017; Munro & Strohminger, 2021). Moreover, the current research revealed that a special form of attention, i.e., social attention, has also been robustly observed with voluntary imagery. Various social cues (e.g., eye gaze, BM walking direction) were adopted to testify the imagery-mediated attentional orienting, and such effects demonstrated reflexivity as the external social attentional effects do. By revealing that imagery-induced social attention exhibits a similar property of external social attention, our results provided further support for the conceptualization of imagery as internal attention. More broadly speaking, voluntary visual imagery acts similarly with visual working memory, which has been conceptualized as a form of internally directed attention (Johnson et al., 2013; Kiyonaga & Egner, 2013, 2014). For instance, the imagery-generated color cue affects visual search, no matter whether the imagined feature is directly relevant with the search task (e.g., Cochrane, Nwabuike, et al., 2018; Liao et al., 2023; Moriya, 2018). And so does the cue representation maintained in visual working memory (e.g., Olivers et al., 2006; Soto et al., 2005; Soto et al., 2008, 2008). Researchers have also investigated if social cues maintained in visual working memory could orient spatial attention (Ji et al., 2022; McDonnell & Dodd, 2013). For example, Ji et al. (2022) combined a working memory task with a dot-probe task and found that holding eye gaze cues in working memory guided attention to the cued location while the arrow cues maintained in working memory yielded null effects, similar with the current study. Reason for their similar effects might be that visual working memory and visual mental imagery have shared internal representations (Albers et al., 2013; Tong, 2013), as the imagined contents can be accurately cross-decoded from early visual cortex using algorithms trained on afferent perception or visual working memory (Albers et al., 2013; Dijkstra et al., 2017; Koenig-Robert & Pearson, 2019; Naselaris et al., 2015). Still, it should be emphasized that visual imagery and working memory are different constructs measured by different paradigms. In these working memory studies, the representation maintained in visual working memory is still derived from external visual input. Here, we demonstrated an internal social attention triggered by visual mental imagery, which generates internal representation without any external input.

Notably, the imagery-induced attentional orienting effect reported in the current research is specific to social cues but not non-social arrow cues, while both gaze cueing and arrow cueing effects have been widely reported in the perceptual central cueing paradigm (e.g., Galfano et al., 2012; Ristic et al., 2002; Tipples, 2002). In some other paradigms, the perception-based attentional effects have been observed with imagined stimuli without social relevance (e.g., color, shape; Liao et al., 2023; Ongchoco & Scholl, 2019; Pashler & Shiu, 1999). The key reason why imagined arrows didn't orient spatial attention as imagined eye gaze cue within the current paradigm might reflect the inherent difference between social and non-social attention, not that 'imagining it' is a special manipulation which diminishes the arrow cueing effect. More specifically, in the perceptually central cueing task, although observers were explicitly told the non-informativeness of central cues, they could still associate the cue direction with the target location in a simple single task. Such an association introduces some voluntary components, leading to the general attentional effects which conceal the intrinsic differences between eye gaze and arrow cues. While in the current context of dual-task paradigm and long period of imagery task, we reasoned that the cue could be well disconnected from the target and thus the qualitative difference between social and non-social cues could be probed. Therefore, null effects induced by imagined non-social cues implied that the non-social attentional effects may involve some voluntary processes (see also Liu et al., 2021), resonating well with literatures that also cast doubt on the reflexivity of non-social attention (Friesen et al., 2004; Hietanen et al., 2006; Hietanen et al., 2008; Ji et al., 2022; Liu et al., 2021). Looking from another angle, our findings help to answer the question regarding 'the specificity of social attention', which is still equivocal and under discussion nowadays. Albeit a group of literatures have revealed the distinctions of social attention compared to non-social attention from many aspects (Friesen et al., 2004; Hietanen et al., 2006; Ji et al., 2020; Kingstone et al., 2000; Ristic et al., 2007; Wang et al., 2020), some research obtained opposite conclusion (e.g., Galfano et al., 2012; Sato et al., 2009; Tipples, 2008). Regarding this controversial question, a very recent meta-analytic study (Chacon-Candia et al., 2023) concluded that while the social and non-social attentional effects were quantitatively similar as reflected by standard cueing paradigm, their qualitative discrepancy still existed and could only be probed relying on variations of classic cueing paradigm, which has been suggested to be more sensitive than classic cueing paradigm (Birmingham & Kingstone, 2009). However, as called in Chacon-Candia et al. (2023), more research and paradigms exploring potential qualitative discrepancy are still in need to reach a conclusion. The current research, responding to this call, modified the classic cueing paradigm with a voluntary visual imagery task and showed clear difference between social and non-social cues, espousing the specificity of social attention and corroborating the potential existence of 'social attention detector' in human brain (Ji et al., 2020; Wang et al., 2020).

The candidate neural basis of the imagery-based internal social attention remains to be explored. Visual mental imagery and visual perception shared partial neural mechanisms (for a recent review, see Pearson, 2019). This is indeed the case in face imagery studies which suggested a large portion of occipital and temporal cortices (e.g., the fusiform face area, FFA), which are usually active during perception of faces, are also strongly activated during face imagery (O'Craven & Kanwisher, 2000). Similarly, the superior temporal sulcus (STS), a perceptual region that responds strongly while viewing BM, is suggested to be activated during biological-motion imagery (Grossman & Blake, 2001). These empirical results lead to the possibility that imagery-based social attention relies on similar neural substrates involved in the perception-induced social attention such as the STS. In order to disentangle this issue, future study may adopt neuroimaging techniques to identify the neural network dedicated to the internal imagery-based social attention, and could compare it with that of classic social attention induced by externally presented cues.

Still, the present study possesses some limitations. First, the control stimuli used in the current study may not completely control for some details, such as the speed of motion or the richness of social cues. Future study should consider using non-social cues involving a quick shift or using dynamic arrow cues to function as a better control. Besides, our sample only included the volunteers who self-identify as neurotypical. And thus, it should be cautious when extending the current findings to the neurodivergent minors, especially those with autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), or aphantasia. Individuals with ASD and ADHD showed difficulty in social attention (Braithwaite et al., 2020; Yang et al., 2024), and the aphantasics exhibit deficits in voluntary imagery (Zeman, 2024; Zeman et al., 2015). Therefore, these populations may not demonstrate the imagery-based social attention. Withal, this remains an open question which requires systematic and comprehensive investigations to answer in the future.

To summarize, the present study revealed that spatial attention could be reflexively guided by social cues generated via voluntary visual mental imagery. Such effect was largely modulated by social relevance and couldn't be generalized to non-social cues, corroborating the specificity of social attention. Importantly, our finding is not simply a type of classic attention triggered by social information, but represents a unique behavior which is distinct from traditionally identified exogenous/endogenous attention and tightly associated with human adaptive functioning. Moreover, the observed effects echoed with previous literatures on imagery-attention interactions, together indicating that the voluntary visual imagery could be a candidate form of internal attention. In conclusion, our inquiry takes a further step in showing how our internal mind's world impacts and interacts with the outside physical world.

Author contributions

S. Zhang performed software, investigation, data curation, formal analysis, and writing - original draft, and contributed equally to conceptualization, methodology, and visualization, and served in a supporting role in writing - review and editing under the supervision of L. Wang and Y. Jiang. L. Wang contributed equally to conceptualization, methodology, project administration, funding acquisition, visualization, and writing - review and editing. Y. Jiang contributed equally to funding acquisition, methodology, project administration, and writing - review and editing.

Open practices statement

We report how we determined the sample size, the data exclusion, all manipulations, and all measures in the study. Data were analyzed using SPSS Version 26.0.0 and G-power Version 3.1. The data and analytic codes for the current study are publicly accessible at https://doi.org/10.57760/sciencedb.o00115.00100. Materials used in the current research are widely available. There is not a preregistration for the current study.

CRediT authorship contribution statement

Shujia Zhang: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Li Wang:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Yi Jiang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition.

Declaration of competing interest

None.

Data availability

We have made the original data and analytic codes for the current research publicly accessible at https://doi.org/10.57760/sciencedb.o00115.00100.

Acknowledgements

This research was supported by grants from the STI2030-Major Projects (No. 2021ZD0203800, 2022ZD0205100), the National Natural Science Foundation of China (No. 32430043, 32371106), the Interdisciplinary Innovation Team of the Chinese Academy of Sciences (JCTD-2021-06), the Key Research and Development Program of Guangdong, China (2023B0303010004, and the Fundamental Research Funds for the Central Universities.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*(2), 82–90. https://doi.org/10.1038/scientificamerican0871-82

Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar, & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 54–73). Cambridge University Press. https://doi.org/10.1017/CBO9780511665547.004.

Birmingham, E., & Kingstone, A. (2009). Human social attention: A new look at past, present, and future investigations. *Annals of the New York Academy of Sciences, 1156*, 118–140. https://doi.org/10.1111/j.1749-6632.2009.04468.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Braithwaite, E. K., Gui, A., & Jones, E. J. H. (2020). Social attention: What is it, how can we measure it, and what can it tell us about autism and ADHD? *Progress in Brain Research, 254*, 271–303. https://doi.org/10.1016/bs.pbr.2020.05.007

Ceja, C. R., & Franconeri, S. L. (2023). Difficulty limits of visual mental imagery. *Cognition, 236*, Article 105436. https://doi.org/10.1016/j.cognition.2023.105436

Chacon-Candia, J. A., Roman-Caballero, R., Aranda-Martin, B., Casagrande, M., Lupianez, J., & Marotta, A. (2023). Are there quantitative differences between eye-gaze and arrow cues? A meta-analytic answer to the debate and a call for qualitative differences. *Neuroscience and Biobehavioral Reviews, 144*, Article 104993. https://doi.org/10.1016/j.neubiorev.2022.104993

Chang, S., Lewis, D. E., & Pearson, J. (2013). The functional effects of color perception and color imagery. *Journal of Vision, 13*(10). https://doi.org/10.1167/13.10.4

Chen, Z., McCrackin, S. D., Morgan, A., & Itier, R. J. (2021). The gaze cueing effect and its enhancement by facial expressions are impacted by task demands: Direct comparison of target localization and discrimination tasks. *Frontiers in Psychology, 12*, Article 618606. https://doi.org/10.3389/fpsyg.2021.618606

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

Cochrane, B. A., & Milliken, B. (2019). Imagined event files: An interplay between imagined and perceived objects. *Psychonomic Bulletin & Review, 26*(2), 538–544. https://doi.org/10.3758/s13423-019-01572-2

Cochrane, B. A., Nwabuike, A. A., Thomson, D. R., & Milliken, B. (2018). An imagery-induced reversal of intertrial priming in visual search. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 44*(4), 572–587. https://doi.org/10.1037/xlm0000470

Cochrane, B. A., Siddhpuria, S., & Milliken, B. (2019). Cueing color imagery: A critical analysis of imagery-perception congruency effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 45*(8), 1410–1421. https://doi.org/10.1037/xlm0000653

Cochrane, B. A., Townsend, B., Lee, R. C., Legere, J. K., Milliken, B., & Shedden, J. M. (2021). Visual imagery influences attentional guidance during visual search: Behavioral and electrophysiological evidence. *Attention, Perception, & Psychophysics, 83*(1), 58–66. https://doi.org/10.3758/s13414-020-02208-7

Cochrane, B. A., Wang, C., Pratt, J., Milliken, B., & Sun, H. J. (2021). Comparing imagery and perception: Using eye movements to dissociate mechanisms in search. *Attention, Perception, & Psychophysics, 83*(7), 2879–2890. https://doi.org/10.3758/s13414-021-02336-8

Cochrane, B. A., Zhu, H., & Milliken, B. (2018). Strategic visual imagery and automatic priming effects in pop-out visual search. *Consciousness and Cognition, 65*, 59–70. https://doi.org/10.1016/j.concog.2018.07.008

Cole, G. G., Samuel, S., & Eacott, M. J. (2022). A return of mental imagery: The pictorial theory of visual perspective-taking. *Consciousness and Cognition, 102*, Article 103352. https://doi.org/10.1016/j.concog.2022.103352

D'Angiulli, A., Runge, M., Faulkner, A., Zakizadeh, J., Chan, A., & Morcos, S. (2013). Vividness of visual imagery and incidental recall of verbal cues, when phenomenological availability reflects long-term memory accessibility. *Frontiers in Psychology, 4*, 1. https://doi.org/10.3389/fpsyg.2013.00001

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *The Journal of Neuroscience, 37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509–540. https://doi.org/10.1080/135062899394920

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Finke, R. A. (1980). Levels of equivalence in imagery and perception. *Psychological Review, 87*(2), 113–132. https://doi.org/10.1037/0033-295x.87.2.113

Finke, R. A. (1989). *Principles of mental imagery*. The MIT Press.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490–495. https://doi.org/10.3758/bf03208827

Friesen, C. K., & Kingstone, A. (2003). Abrupt onsets and gaze direction cues trigger independent reflexive attentional effects. *Cognition, 87*(1), B1–B10. https://doi.org/10.1016/s0010-0277(02)00181-6

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology. Human Perception and Performance, 30*(2), 319–329. https://doi.org/10.1037/0096-1523.30.2.319

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Frischen, A., Smilek, D., Eastwood, J. D., & Tipper, S. P. (2007). Inhibition of return in response to gaze cues: The roles of time course and fixation cue. *Visual Cognition, 15* (8), 881–895. https://doi.org/10.1080/13506280601112493

Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *The Quarterly Journal of Experimental Psychology, 65*(10), 1895–1910. https://doi.org/10.1080/17470218.2012.663765

Ganis, G., & Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *NeuroImage, 42*(4), 1714–1727. https://doi.org/10.1016/j.neuroimage.2008.07.004

Gregori Grgic, R., Calore, E., & de'Sperati, C.. (2016). Covert enaction at work: Recording the continuous movements of visuospatial attention to visible or imagined targets by means of steady-state visual evoked potentials (SSVEPs). *Cortex, 74*, 31–52. https://doi.org/10.1016/j.cortex.2015.10.008

Gregory, S. E., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 43*(2), 237–249. https://doi.org/10.1037/xlm0000294

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research, 41*(10–11), 1475–1482. https://doi.org/10.1016/s0042-6989(00)00317-5

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *The Behavioral and Brain Sciences, 27*(3), 377–396. https://doi.org/10.1017/s0140525x04000093

Hertzog, C., & Dunlosky, J. (2006). Using visual imagery as a mnemonic for verbal associative learning. In *Imagery and Spatial Cognition: Methods, Models and Cognitive Assessment (pp. 259-280)*. https://doi.org/10.1075/aicr.66.20her

Hietanen, J. K., Leppanen, J. M., Nummenmaa, L., & Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain Research, 1215*, 123–136. https://doi.org/10.1016/j.brainres.2008.03.091

Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hamalainen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *NeuroImage, 33*(1), 406–413. https://doi.org/10.1016/j.neuroimage.2006.06.048

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural Systems for the Generation of visual images. *Neuron, 28*(3), 979–990. https://doi.org/10.1016/s0896-6273(00)00168-9

Ji, H., Wang, L., & Jiang, Y. (2020). Cross-category adaptation of reflexive social attention. *Journal of Experimental Psychology. General, 149*(11), 2145–2153. https://doi.org/10.1037/xge0000766

Ji, H., Yuan, T., Yu, Y., Wang, L., & Jiang, Y. (2022). Internal social attention: Gaze cues stored in working memory trigger involuntary attentional orienting. *Psychological Science, 33*(9), 1532–1540. https://doi.org/10.1177/09567976221094628

Johnson, M. R., Higgins, J. A., Norman, K. A., Sederberg, P. B., Smith, T. A., & Johnson, M. K. (2013). Foraging for thought: An inhibition-of-return-like effect resulting from directing attention within working memory. *Psychological Science, 24* (7), 1104–1112. https://doi.org/10.1177/0956797612466414

Kalkstein, J., Checksfield, K., Bollinger, J., & Gazzaley, A. (2011). Diminished top-down control underlies a visual imagery deficit in normal aging. *The Journal of Neuroscience, 31*(44), 15768–15774. https://doi.org/10.1523/JNEUROSCI.3209-11.2011

Keogh, R., & Pearson, J. (2017). The perceptual and phenomenal capacity of mental imagery. *Cognition, 162*, 124–132. https://doi.org/10.1016/j.cognition.2017.02.004

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science, 330*(6006), 932. https://doi.org/10.1126/science.1192439

Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science, 11*(2), 159–166. https://doi.org/10.1111/1467-9280.00232

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review, 20*(2), 228–242. https://doi.org/10.3758/s13423-012-0359-y

Kiyonaga, A., & Egner, T. (2014). The working memory stroop effect: When internal representations clash with external stimuli. *Psychological Science, 25*(8), 1619–1629. https://doi.org/10.1177/0956797614536739

Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 1–16. https://doi.org/10.1068/v070821

Koenig-Robert, R., & Pearson, J. (2019). Decoding the contents and strength of imagery before volitional engagement. *Scientific Reports, 9*(1), 3504. https://doi.org/10.1038/s41598-019-39813-y

Koenig-Robert, R., & Pearson, J. (2021). Why do imagery and perception look and feel so different? *Philosophical Transactions of the Royal Society of London B Biological Sciences, 376*(1817), 20190703. https://doi.org/10.1098/rstb.2019.0703

Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science, 26*(2), 207–231. https://doi.org/10.1016/s0364-0213(01)00065-9

Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*(5), 541–567. https://doi.org/10.1080/135062899394939

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage, 59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Liao, M.-R., Grindell, J. D., & Anderson, B. A. (2023). A comparison of mental imagery and perceptual cueing across domains of attention. *Attention, Perception, & Psychophysics, 85*(6), 1834–1845. https://doi.org/10.3758/s13414-023-02747-9

Liu, W., Yuan, X., Liu, D., Wang, L., & Jiang, Y. (2021). Social attention triggered by eye gaze and walking direction is resistant to temporal decay. *Journal of Experimental Psychology. Human Perception and Performance, 47*(9), 1237–1246. https://doi.org/10.1037/xhp0000939

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

McDonnell, G. P., & Dodd, M. D. (2013). Examining the influence of a spatially irrelevant working memory load on attentional allocation. *Journal of Experimental Psychology. Human Perception and Performance, 39*(4), 933–940. https://doi.org/10.1037/a0032111

McKay, K. T., Grainger, S. A., Coundouris, S. P., Skorich, D. P., Phillips, L. H., & Henry, J. D. (2021). Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin, 147*(12), 1269–1289. https://doi.org/10.1037/bul0000353

McKelvie, S. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery, 19*, 1–106.

Moriya, J. (2018). Visual mental imagery influences attentional guidance in a visual-search task. *Attention, Perception, & Psychophysics, 80*(5), 1127–1142. https://doi.org/10.3758/s13414-018-1520-0

Munro, D., & Strohminger, M. (2021). Are we free to imagine what we choose? *Synthese, 199*(5–6), 11847–11864. https://doi.org/10.1007/s11229-021-03314-1

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage, 105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Nie, Q. Y., Ding, X., Chen, J., & Conci, M. (2018). Social attention directs working memory maintenance. *Cognition, 171*, 85–94. https://doi.org/10.1016/j.cognition.2017.10.025

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences, 13*(3), 135–143. https://doi.org/10.1016/j.tics.2008.12.006

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience, 12* (6), 1013–1023. https://doi.org/10.1162/08989290051137549

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology. Human Perception and Performance, 32*(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Ongchoco, J. D. K., & Scholl, B. J. (2019). How to create objects with your mind: From object-based attention to attention-based objects. *Psychological Science, 30*(11), 1648–1655. https://doi.org/10.1177/0956797619863072

Pashler, H., & Shiu, L. P. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin & Review, 6*(3), 445–448. https://doi.org/10.3758/bf03210833

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews. Neuroscience, 20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology, 18*(13), 982–986. https://doi.org/10.1016/j.cub.2008.05.048

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the mind's eye: The metacognition of visual imagery. *Psychological Science, 22*(12), 1535–1542. https://doi.org/10.1177/0956797611417134

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897x00366

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences, 7*(3), 113–118. https://doi.org/10.1016/s1364-6613(03)00003-2

Reinhart, R. M., McClenahan, L. J., & Woodman, G. F. (2015). Visualizing trumps vision in training attention. *Psychological Science, 26*(7), 1114–1122. https://doi.org/10.1177/0956797615577619

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review, 9*(3), 507–513. https://doi.org/10.3758/bf03196306

Ristic, J., Wright, A., & Kingstone, A. (2007). Attentional control and reflexive orienting to gaze and arrow cues. *Psychonomic Bulletin & Review, 14*(5), 964–969. https://doi.org/10.3758/bf03194129

Runge, M., Bakhilau, V., Omer, F., & D'Angiulli, A. (2015). Trial-by-trial vividness self-reports versus VVIQ. *Imagination, Cognition and Personality, 35*(2), 137–165. https://doi.org/10.1177/0276236615587490

Runge, M. S., Cheung, M. W., & D'Angiulli, A. (2017). Meta-analytic comparison of trial-versus questionnaire-based vividness reportability across behavioral, cognitive and neural measurements of imagery. *Neuroscience of Consciousness, 2017(1), nix006.* https://doi.org/10.1093/nc/nix006

Sakai, K. (2017). A review of capacity limitation from visual perception to short-term visual memory of a single curved contour. *Journal of Psychological Researches, 7*(7), 3761–3379.

Sato, W., Kochiyama, T., Uono, S., & Yoshikawa, S. (2009). Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. *NeuroImage, 45*(3), 984–992. https://doi.org/10.1016/j.neuroimage.2008.12.052

Shi, J., Weng, X., He, S., & Jiang, Y. (2010). Biological motion cues trigger reflexive attentional orienting. *Cognition, 117*(3), 348–354. https://doi.org/10.1016/j.cognition.2010.09.001

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*(2), 179–193. https://doi.org/10.1037/h0027277

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology. Human Perception and Performance, 31*(2), 248–261. https://doi.org/10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*(9), 342–348. https://doi.org/10.1016/j.tics.2008.05.007

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1–29. https://doi.org/10.1037/h0093759

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human perceptual learning by mental imagery. *Current Biology, 19*(24), 2081–2085. https://doi.org/10.1016/j.cub.2009.10.060

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review, 9*(2), 314–318. https://doi.org/10.3758/bf03196287

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics, 70*(1), 77–87. https://doi.org/10.3758/pp.70.1.77

Tong, F. (2013). Imagery and visual working memory: One and the same? *Trends in Cognitive Sciences, 17*(10), 489–490. https://doi.org/10.1016/j.tics.2013.08.005

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers, 36*(4), 625–629. https://doi.org/10.3758/bf03206542

Wang, L., Wang, Y., Xu, Q., Liu, D., Ji, H., Yu, Y., & Jiang, Y. (2020). Heritability of reflexive social attention triggered by eye gaze and walking direction: Common and unique genetic underpinnings. *Psychological Medicine, 50*(3), 475–483. https://doi.org/10.1017/S003329171900031X

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual imagery and perception share neural representations in the alpha frequency band. *Current Biology, 30*(13), 2621–2627 e2625. https://doi.org/10.1016/j.cub.2020.04.074

Yang, F., Tian, J., Yuan, P., Liu, C., Zhang, X., Yang, L., & Jiang, Y. (2024). Unconscious and Conscious Gaze-Triggered Attentional Orienting: Distinguishing Innate and Acquired Components of Social Attention in Children and Adults with Autistic Traits and Autism Spectrum Disorders. *Research, 7*, Article 0417. https://doi.org/10.34133/research.0417

Yu, Y., Ji, H., Wang, L., & Jiang, Y. (2020). Cross-modal social attention triggered by biological motion cues. *Journal of Vision, 20*(10), 21. https://doi.org/10.1167/jov.20.10.21

Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. *Trends in Cognitive Sciences, 28*(5), 467–480. https://doi.org/10.1016/j.tics.2024.02.007

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery - congenital aphantasia. *Cortex, 73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019