Consciousness and Cognition 143 (2026) 104101

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Investigating the reversed congruency effect of gaze through the lens of joint attention

Anna Lorenzoni *, Mario Dalmaso

Department of Developmental and Social Psychology, University of Padova, Italy

ARTICLE INFO

Keywords:
Social attention
Eye-gaze
Arrows
Joint attention
Spatial Stroop task

ABSTRACT

Recent evidence has highlighted differences in attentional shifting elicited by gaze and arrow stimuli. Specifically, in a spatial Stroop task, arrows consistently produce a standard congruency effect (SCE; faster reaction times on congruent vs. incongruent trials). In contrast, averted gaze stimuli produce a reversed congruency effect (RCE; faster reaction times on incongruent vs. congruent trials). The origin of this discrepancy is still under debate. One hypothesis suggests that the RCE for gaze is driven by joint attention (the shared focus of two individuals on the same object). According to this view, during incongruent trials, gaze stimuli appear to look toward the fixation stimulus, thus creating an episode of joint attention with the participant. To investigate this possibility, we conducted four experiments where participants identified the direction of a peripheral target (gaze or arrow). We also varied the fixation stimulus (symbolic cross or real-world object) and added an oculomotor component, to enhance joint attention contexts. Results consistently showed an SCE for arrows and an RCE for gaze, regardless of fixation type and the request to perform an eye movement, suggesting that joint attention, as operationalised in the present study, may not fully explain the RCE elicited by gaze.

1. Introduction

Humans are extremely sensitive to eye-gaze stimuli (e.g., Adams Jr & Kleck, 2003; Baron-Cohen, 1996; Emery, 2000; Hietanen et al., 2016). From birth, our visual system shows an attentional preference for faces, particularly those with a direct gaze (see, e.g., Johnson et al., 1991; Shultz et al., 2018). While direct gaze typically serves as a signal to initiate communication (see e.g., Hamilton, 2016), averted gaze can trigger attentional shifts toward the same spatial location, a phenomenon known as social attention (e.g., Dalmaso et al., 2020a; Marotta et al., 2012; Oyama & Okubo, 2022).

To determine whether gaze cues can elicit unique forms of attentional responses, several works have been conducted comparing eye-gaze stimuli to non-social directional cues (e.g., arrows; Marotta et al., 2018; Tipples, 2002; for a review, see Frischen et al., 2007). These signals have been widely compared using spatial cueing paradigms based on Posner's task (e.g., Friesen & Kingstone, 1998; Driver et al., 1999; see also Posner, 1980). These paradigms usually involve the presentation of a centrally located directional stimulus indicating left or right. After a time interval (i.e., stimulus onset asynchrony; SOA), a target appears at the cued (i.e., a congruent trial) or uncued (i.e., an incongruent trial) location, and participants have to respond to the target. Results typically show faster and more accurate responses for cued than uncued targets, thus revealing a so-called cueing effect. At the behavioural level, several studies reported that both eyes and arrows elicit quantitatively similar cueing effects (e.g., Kuhn & Kingstone, 2009; Blair et al., 2017; Dalmaso et al., 2020b; for a review and meta-analysis, see Chacón-Candia et al., 2023). According to some authors, these similarities could be ascribed to a domain-general attentional mechanism that would treat the two cues to the same extent (e.g., Callejas et al., 2014; Cole & Millett, 2019; Heyes, 2014; Marotta et al., 2019). However, it is important to note that differences between social and non-social cues may be qualitative rather than quantitative in nature (Chacón-Candia et al., 2023). For example, social cues can selectively enhance working memory performance (e.g., Gregory & Jackson, 2017), and social attention can be further modulated by social characteristics such as ethnicity, social status, and dominance (e.g., Zhang et al., 2021; Lorenzoni et al., 2023; for a review, see Dalmaso et al., 2020a).

Interestingly, some recent studies using a spatial Stroop task have further revealed qualitative differences in attentional processes between eye-gaze and non-social stimuli (e.g., Cañadas & Lupiáñez, 2012; Marotta et al., 2018; Tanaka et al., 2023, 2024). The spatial Stroop task

* Corresponding author at: Department of Developmental and Social Psychology (DPSS), Via Venezia 8, 35131, Padova, Italy.
E-mail address: anna.lorenzoni@unipd.it (A. Lorenzoni).

https://doi.org/10.1016/j.concog.2026.104101
Received 11 May 2026; Received in revised form 16 July 2026; Accepted 21 July 2026
Available online 25 July 2026
1053-8100/

(e.g., Lu & Proctor, 1995) examines the conflict between the processing of task-relevant (i.e., the direction indicated by a stimulus) and task-irrelevant (i.e., the location of the stimulus) spatial information. In the case of eye-gaze stimuli, participants are generally asked to discriminate the direction (e.g., left or right) indicated by the gaze stimulus, which can be located to the left or right of a central fixation point. The same is also required for arrow stimuli acting as a control condition. Typically, the results show a standard congruency effect (SCE) for arrows, yielding faster and more accurate responses when the direction indicated by the stimulus is congruent with its spatial location. Interestingly, this also holds for biological stimuli such as pointing gestures (Bonventre & Marotta, 2023; Dalmaso et al., 2023). However, for eye-gaze stimuli, the reverse pattern is observed, with faster and more accurate responses when the direction indicated by the gaze is incongruent with its spatial location. Although this reversed congruency effect (RCE) has been replicated several times, in different laboratories, and with different experimental conditions (see, e.g., Chacón-Candia et al., 2020; Hemmerich et al., 2022; Ishikawa et al., 2021, 2025a, 2025b; Tanaka et al., 2024), its nature is still debated, and different accounts have been proposed to explain its nature.

According to some works, the RCE for gaze would be caused by low-level mechanisms. In this regard, Román-Caballero et al. (2021a) showed that the RCE occurred even with non-social stimuli when these were presented against perceptually complex backgrounds (e.g., arrows with mosaics). Román-Caballero et al. (2021a) argued that target background segregation and delays in processing spatial information would play a crucial role in the emergence of the RCE. Specifically, complex backgrounds would slow target extraction, weakening the spatial Stroop effect. Moreover, gaze targets embedded in whole faces elicit a stronger RCE than isolated eyes (Román-Caballero et al., 2021b), suggesting that facial context interferes with target selection and accelerates location code decay. Extending this perspective, Tanaka et al. (2025, 2026) proposed a dual-stage model, which involves an initial target-background segregation (Román-Caballero et al., 2021a, 2021b) followed by a response inhibition (Ridderinkhof, 2002a, 2002b; Ridderinkhof et al., 2004; Tipper, 1985; Tipper & Driver, 1988), thus corroborating a perceptual explanation for the RCE for gaze.

Other works provided evidence supporting a more social interpretation of this effect. Indeed, it has been shown that the RCE for gaze can be influenced by social variables such as emotional facial expressions (Jones, 2015; Marotta et al., 2022; Torres-Marín et al., 2017), developmental trajectories across different age groups (Aranda-Martín et al., 2022), individual differences in social anxiety (Ishikawa et al., 2021), and the nature of the target faces (e.g., human, animal, or robotic; Ishikawa et al., 2025b). In this regard, three accounts based on a social interpretation of eye-gaze have been pushed forward. One is the 'eye-contact' hypothesis (Marotta et al., 2018), which suggests that in the incongruent condition the eyes, directed inwards, may be perceived as making direct eye contact with the participant (see also Cañadas & Lupiáñez, 2012), aligning with a documented bias in interpreting gaze stimuli as looking at us, especially during episodes of perceptual ambiguity (Mareschal et al., 2013). This perception could facilitate responses to incongruent stimuli, given that direct gaze is a well-documented social cue known to capture and enhance attention more effectively than averted gaze (e.g., Böckler et al., 2014). Converging evidence for the eye-contact hypothesis comes from a recent study by Marotta and Chacón-Candia (2026), who showed that the RCE was significantly stronger for partially averted gaze stimuli than for fully averted gaze stimuli, suggesting that subtle gaze deviations are more likely to be implicitly misperceived as direct eye contact, thereby facilitating responses on incongruent trials. Another is the 'joint distraction' hypothesis (Hemmerich et al., 2022), which posits that on congruent trials, gaze stimuli, because of their particularly relevant social nature, would cause stronger attentional shifts than arrows away from the task, requiring reorientation and thereby increasing response latencies. This suggests that the RCE would emerge due to the disruptive effects of congruent gaze stimuli rather than the facilitation on incongruent gaze stimuli. Finally, Edwards et al. (2020) proposed the 'joint attention' hypothesis, which suggests that, on incongruent trials, the peripheral stimulus would direct participants' attention toward the object of fixation, triggering an episode of joint attention that, in turn, would facilitate response execution. The joint attention hypothesis constitutes the backbone of the present work and will be discussed in detail in the following section.

1.1. *The possible role of joint attention in the RCE*

Joint attention can be defined as the shared focus of two individuals on an object. A key feature of joint attention lies in its triadic nature: it involves two interacting agents and a shared external referent, such as an object (Carpenter et al., 1998; Mundy & Newell, 2007; Tomasello, 1995). In other words, joint attention is typically achieved when one person uses another's gaze (or other signals such as pointing gestures) to locate and attend to the same object (Emery, 2000; Moore & Dunham, 1995). The presence of a concrete object is therefore not incidental, but constitutive of joint attention, as it provides the common focus through which attentional and intentional states are aligned (Emery, 2000; Mundy, 2018). A growing body of research has demonstrated that the presence of a meaningful object facilitates the establishment and maintenance of joint attention episodes, enhancing social engagement processes (e.g., Bayliss et al., 2013; Farroni et al., 2000; Teufel et al., 2010). These findings suggest that the presence of a shared, meaningful, object acts as a fundamental anchor for social orienting processes based on joint attention.

Based on this, Edwards et al. (2020) investigated whether the RCE for gaze could result from a joint attention episode rather than from other possible mechanisms, in particular eye contact (Marotta et al., 2018). In two experiments based on a spatial Stroop-like task, participants were asked to discriminate the direction (left or right) of gaze in faces presented peripherally to the left or right of fixation. Crucially, the authors manipulated how these faces gazed toward or away from the fixation point, thus creating joint attention (i.e., both the faces and the participant looked at the same location) versus non-joint attention conditions. In Experiment 1, participants fixated either a central or a peripheral symbolic cross, while the target face always appeared in a lateral position. The results showed that, when eye-gaze stimuli were directed centrally (i.e., a face looking inwards), responses were faster for faces establishing (vs. non-establishing) joint attention (i.e., gaze oriented toward fixation), replicating earlier findings (e.g., Cañadas & Lupiáñez, 2012; Marotta et al., 2018). However, this facilitation disappeared when participants fixated on a peripheral location (i.e., a face looking outwards). In Experiment 2, the fixation cross was replaced with images of real-world objects (e.g., a mug) to reinforce the joint attention context. In this case, responses were faster for joint (vs. non-joint) attention trials regardless of whether the object appeared centrally or peripherally (i.e., regardless of whether the face was looking inwards or outwards). Overall, these findings support the view that joint attention episodes, especially when involving real-world objects, enhance the processing of the target face and lead to the RCE. According to Edwards et al. (2020), these results also provide evidence that appears to disconfirm the eye contact hypothesis (Marotta et al., 2018), as the facilitation in gaze direction discrimination emerged not only when the gaze was directed inwards (a condition typically associated with potential eye contact), but also when it was directed outwards (a condition not associated with eye contact). Considering these theoretical and empirical insights, the manipulation of fixation type (i.e., a symbolic cross vs. a real-world object) represents a direct test of whether the triadic structure of joint attention (specifically, the presence of a meaningful shared referent) modulates the RCE. If joint attention is implied in the emergence of the RCE, its magnitude should be enhanced when an object of shared focus is available to anchor the interaction, compared to when attention is directed toward an abstract and symbolic fixation cross.

1.2. The current study

The study by Edwards et al. (2020) provides valuable insights into the possible role of joint attention in modulating the RCE. However, some aspects remained unexplored and were therefore systematically addressed in the present work. First, in Edwards et al. (2020), non-social control stimuli (such as arrows) were not included, making it impossible to determine whether the observed facilitation was specific to gaze-based social attention or reflected a more general attentional process. Given that arrows, by definition, do not elicit joint attention, their inclusion is essential to test the social specificity of the effect. Second, although Edwards et al. (2020) tested both symbolic fixation crosses and real-world objects, these manipulations were implemented across separate experiments rather than within a single, directly-comparable, paradigm. Consequently, the extent to which the type of fixation (symbolic vs. object-based) modulates the RCE under identical task demands remains unknown. By integrating both fixation types within the same paradigm, our study provides a direct test of whether the defining triadic feature of joint attention (i.e., the shared focus on a meaningful external object) selectively enhances the RCE for gaze stimuli. Third, and crucially, in Experiments 3 and 4 we introduced a novel oculomotor component designed to actively engage the mechanisms underlying joint attention to a larger extent. At the beginning of each trial, participants were required to fixate on the lower part of the screen and then to make an eye-movement upward toward the central fixation item. This dynamic manipulation mimics the natural initiation of joint attention episodes, which involve orienting one's attention toward the same object or event being looked at by another agent (Emery, 2000; Pfeiffer-Leßmann et al., 2012; Tomasello & Carpenter, 2007). Prior evidence shows that such active eye movements are critical for the emergence of social attention phenomena, such as the 'gaze-leading' effect (Edwards et al., 2015), which typically occur only when participants initiate joint attention toward a real-world object. Thus, the introduction of an oculomotor phase represents an important methodological innovation, increasing both the ecological validity and the theoretical precision of our paradigm. In summary, here we systematically replicated and extended Edwards et al. (2020) by i) introducing non-social control cues; ii) directly comparing symbolic and object-based fixation types within the same paradigm; iii) incorporating a dynamic oculomotor component to explicitly foster the emergence of joint attention. Together, these extensions allow us to test whether the RCE for gaze reflects a genuinely social mechanism tied to shared attention.

Four experiments were conducted. Although the four experiments shared a common framework, each was designed to address a distinct theoretical or methodological question. Experiments 1 and 2 focused on whether potential carry-over effects due to the intermixed versus blocked presentation of fixation categories could mask differences between real-world and symbolic fixation items. This manipulation allowed us to assess the influence of contextual organization on social orienting responses. Experiments 3 and 4 extended this investigation by adding an oculomotor component, thereby testing whether the active initiation of eye movements toward the fixation item (i.e., an important feature of joint attention; Edwards et al., 2015) modulates the relationship between fixation type and congruency effects. Taken together, this incremental, yet theory-driven, progression ensured that each experiment provided a unique contribution to understanding whether perceptual, contextual, or oculomotor factors could shape the emergence of the RCE for gaze.

In the first experiment, we devised a modified version of the spatial Stroop task developed by Marotta et al. (2018), in which we manipulated the nature of the fixation point, which could be either a symbolic cross or a real-world object. Consistent with previous studies (e.g., Cañadas & Lupiáñez, 2012; Marotta et al., 2018), we expected to observe an SCE for arrow stimuli and an RCE for eye-gaze stimuli. We did not anticipate any modulatory role of fixation type for arrow stimuli, as they lack social intentionality. In contrast, for eye-gaze stimuli, we expected a larger RCE when the fixation point was a real-world object than when it was a symbolic cross, in line with the defining features of joint attention (Gibson & Sztybel, 2015; Makris et al., 2013). Critically, this manipulation provides a direct test of competing accounts of the RCE. Indeed, whereas the joint-attention account predicts a larger RCE in the presence of a shared attentional focus on the same object, both the eye-contact account (e.g., Marotta et al., 2018), which is independent of the fixation object, and the perceptual account (e.g., Román-Caballero et al., 2021a), which is based on low-level perceptual processes, predict no modulation by fixation type.

2. Experiment 1

Participants completed a spatial Stroop task, with eye-gaze and arrow stimuli presented in two distinct blocks, where the fixation item could be a symbolic cross or a real-world object presented intermixed. According to our hypothesis based on joint attention, we expected a greater RCE for eye-gaze stimuli presented alongside a real-world object than a symbolic cross.

2.1. Methods, results, discussion

2.1.1. Participants

We followed the guidelines proposed by Brysbaert and Stevens (2018) for linear mixed-effect models to determine the minimum number of participants needed to be tested. According to these guidelines, 1600 observations (at least) per experimental cell should be recorded. Based on our experimental design (see next section), we established that the minimum sample size was equal to 50. Sixty participants (26 females; *mean age* = 25.2 years, *SD* = 4.37; four left-handed) were recruited online through the Prolific crowdsourcing platform. Data collection for all experiments was conducted between May and July 2024. The participants were compensated with 2.5 €. The research protocol was approved by the Ethics Committee for Psychological Research of the University of Padova and was conducted in accordance with the Declaration of Helsinki.

2.1.2. Apparatus, stimuli, and procedure

We developed the experiment using PsychoPy (Peirce et al., 2019), which was delivered online using the Pavlovia platform (Bridges et al., 2020). Two target stimuli were used: the face of a young adult Caucasian female, with a neutral expression and the gaze averted leftwards or rightwards (about 300 px width × 300 px height), and two black arrows both pointing leftwards or rightwards (each arrow was about 176 px width × 48 px height). The face stimulus was extracted from the standardised MR2 face database (Strohminger et al., 2016; see also Dalmaso et al., 2021). For the two targets, the parts conveying spatial information (e.g., eye-gaze for the facial stimulus and the two 'heads' for the arrows) were of the same size. In addition, two categories of fixation items were used: real-world objects versus a symbolic stimulus. For the real-world objects, 8 kitchen objects (such as pans, bowls, ladles, etc.) were used (for a similar approach, see also, e.g., Bayliss et al., 2013), while the symbolic stimulus was a black cross. All stimuli were presented on a white background (see Fig. 1).

At the beginning of each trial (see also Fig. 1), the black fixation cross or a real-world object appeared at the centre of the screen for 1000 ms. Then, the face or the arrow (depending on the block) appeared either leftwards or rightwards (±0.3 normalised units; timeout: 1500 ms) with respect to the centre of the screen. Participants were asked to classify, as quickly and accurately as possible, the direction in which the target was pointing by pressing the 'F' key (with the left index finger) for 'left' and the 'K' key (with the right index finger) for 'right'. They were also asked to keep their eyes at fixation and ignore the location (left or right) where the target could appear, as it was task-irrelevant. Missing and incorrect

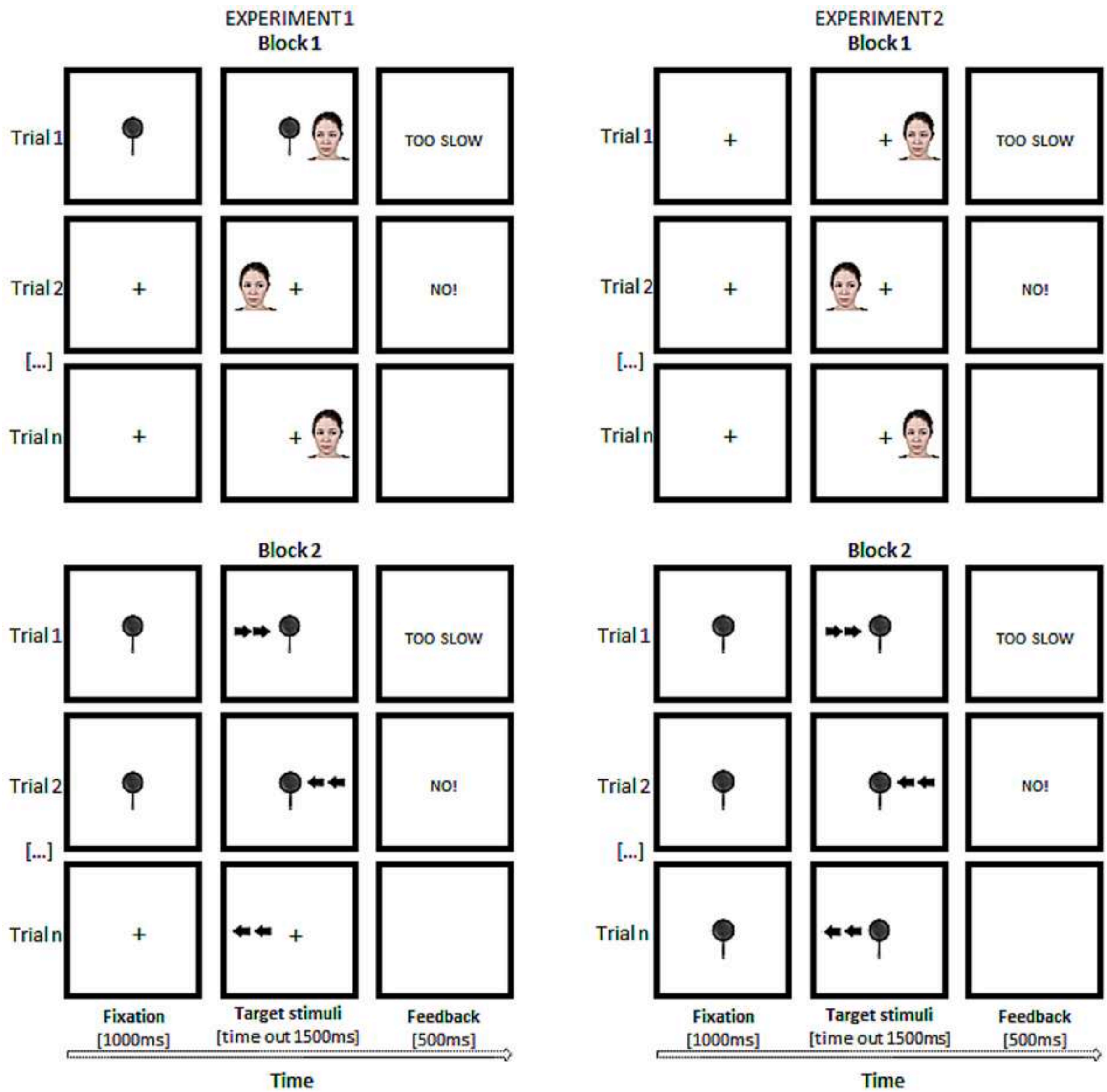

Fig. 1. Examples of trials and stimuli (not drawn to scale) used in the study. The left part of the picture depicts the procedure of Experiment 1, while the right part refers to Experiment 2. The face with the averted gaze (Block 1), or arrows (Block 2), appeared either to the left or right of the central item. While in both experiments, the target stimuli (face or arrows) were presented in different blocks (counterbalanced across participants), the fixation item (symbolic cross or real-world object) was intermixed within the block for Experiment 1 but presented in different blocks for Experiment 2. Visual feedback was given in all blocks for slow or incorrect responses, while a blank screen followed correct responses.

responses were signalled by a 500-ms visual feedback (the words 'TOO SLOW' and 'NO', respectively), and correct responses were followed by a 500-ms blank screen. The two targets were presented in two distinct blocks, whereas the two types of fixation items (i.e., cross or real-world objects) were presented randomly within the blocks and with the same frequency. Two versions of the task were created to counterbalance the order of target presentation. In one version, participants completed two consecutive blocks containing only face stimuli, followed by two consecutive blocks containing only arrow stimuli. In another version, the order was reversed (i.e., arrow first, face second). Thirty participants were randomly assigned to each of the two versions. The whole experiment included 256 trials (64 for each experimental block) and lasted about 15 min. A short break was granted at the end of each block. Participants proceeded to the experimental trials after completing 10 practice trials.

2.1.3. Analysis

Data were analysed using generalised mixed-effects models (GMMs; Baayen et al., 2017). GMMs are an extension of the general linear models (GLMs) that allow one to specify the distribution family. Since residuals are often positively skewed and heteroscedastic when dealing with nonnegative behavioural data (as, e.g., latency and accuracy), these models are preferred to the classical ANOVAs (Baayen et al., 2008). The GMMs approach allows modelling data for random and fixed effects. Moreover, those methods fit with multiple, crossed-grouping factors and, possibly, unbalanced data sets by stabilising the estimation of parameters (Baayen et al., 2008). Specifically, a Gamma distribution with identity link was used for RT data, and a binomial distribution with logit link was used for accuracy data, which were coded as binary vectors (correct = 1, incorrect = 0) at the trial level.

We adopted a two-step analytic strategy. First, model selection based on the Akaike Information Criterion (AIC) and AIC weights was used to identify the mixed-effects model providing the best balance between goodness-of-fit and parsimony (Wagenmakers & Farrell, 2004). All theoretically plausible models, ranging from the simplest model including only random effects to the most complex model including all main effects and their interactions, were specified a priori and compared simultaneously. The model with the lowest AIC and the highest AIC weight was retained for subsequent inference. Second, inferential statistics from the selected model were inspected to evaluate the effects of the experimental manipulations. Since the critical three-way interaction between Congruency, Fixation, and Target Stimulus was not included in the AIC-selected model, its parameter estimate was obtained from the full model including all main effects and interactions. Final models used for inferential purposes (i.e., those including all fixed effects) were fitted using the Laplace approximation (nAGQ = 1); model comparison for fixed-effects selection was instead based on a faster, non-iterative approximation (nAGQ = 0), a common practice to speed up computation during model selection. To achieve convergence of the final models with nAGQ = 1, RT (originally expressed in milliseconds) was rescaled by a factor of 100 prior to model fitting; all reported fixed-effect estimates and standard errors were subsequently rescaled back to the original millisecond scale for reporting. The same approach was adopted in all subsequent experiments.

2.1.4. Results

Trials in which a missed response was recorded (0.52% of the trials) were discarded and not further analysed. Error trials (4.80%) were excluded from the RT analyses. Nonetheless, they were analysed separately for completeness, although the primary analyses concerned RTs. Correctly-responded trials with a latency smaller than 200 ms or greater than 1300 ms were discarded (0.13% of the trials; see also Cañadas & Lupiáñez, 2012; Marotta et al., 2018). Following data cleaning, each experimental condition had at least 1753 observations, and therefore statistical power was adequate (Brysbaert & Stevens, 2018).

We considered as experimental factors Congruency (Congruent vs. Incongruent), Target Stimulus (Face vs. Arrow) and Fixation (Cross vs. Real-world object), and their interactions as fixed effects. All categorical predictors were entered using sum (deviation) coding (−0.5/+0.5), which allows main effects to be interpreted as effects averaged across the levels of the other factors, rather than conditional on a reference level (Schad et al., 2020). To determine the random-effects structure, we followed the model comparison procedure described by Dalmaso et al. (2023), starting from the maximal model justified by the design and simplifying until convergence was achieved. Models including by-participant random slopes for Target Stimulus, alone or in interaction with Congruency, failed to converge. The final model therefore included a random intercept and a by-participant random slope for Congruency only (correlated random effects structure).

According to the AIC-based model comparison procedure, the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the RT data (AIC weight = 0.70; see Table S1 in the Supplementary Materials). This model outperformed the model additionally including the three-way interaction between Congruency, Target Stimulus, and Fixation (AIC weight = 0.30). Parameter estimates from this model revealed a main effect of Congruency (*Estimate* = 11.71, *SE* = 3.34, *t* = 3.50, $p < .001$; *95% C.I.* [5.16, 18.26]), showing shorter RTs on congruent trials (*M* = 532.60 ms, *SD* = 140.80) than on incongruent trials (*M* = 544.02 ms, *SD* = 131.35). In addition, a main effect of Target Stimulus was found (*Estimate* = 116.06, *SE* = 1.62, *t* = 71.79, $p < .001$; *95% C.I.* [112.89, 119.23]), due to shorter RTs for arrows (*M* = 481.32 ms, *SD* = 127.95) than for faces (*M* = 594.70 ms, *SD* = 119.97). Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = −60.98, *SE* = 3.23, *t* = −18.87, $p < .001$; *95% C.I.* [−67.31, −54.65]). When the target stimulus was the arrows, shorter RTs were found on congruent (*M* = 461.45 ms, *SD* = 122.99) than on incongruent (*M* = 502.43 ms, *SD* = 129.73; *Estimate* = −42.20, *SE* = 3.56, *z* = −11.85; $p < .001$) trials. Instead, when the target stimulus was the face, the reversed pattern was found, with shorter RTs for incongruent (*M* = 584.17 ms, *SD* = 119.92) than for congruent (*M* = 605.17 ms, *SD* = 119.12; *Estimate* = 18.78, *SE* = 3.86, *z* = 4.86; $p < .001$) trials. Finally, the triple interaction between Congruency, Fixation and Target Stimulus, as estimated in the full model, was not significant (*Estimate* = −3.40, *SE* = 6.45, *t* = −0.53, *p* = .60; 95% *C.I.* [−16.04, 9.24]; see also Fig. 2).

As for accuracy, the model comparison revealed that the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the data, outperforming the model additionally including the three-way interaction between Congruency, Fixation, and Target Stimulus ($\chi^2(1) = 1.45$, *p* = .23). Parameter estimates from this model revealed a main effect of Congruency (*Estimate* = −0.81, *SE* = 0.09, *z* = −8.91, $p < .001$; *95% C.I.* [−0.98, −0.63]), showing higher accuracy, on average, for congruent (M = 0.97, SD = 0.17) than for incongruent trials (M = 0.94, SD = 0.23). The main effect of Target Stimulus was not significant (*Estimate* = −0.07, *SE* = 0.09, *z* = −0.78, *p* = .43), with comparable accuracy for arrows (M = 0.95, SD = 0.21) and faces (M = 0.96, SD = 0.20). Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = 1.47, *SE* = 0.18, *z* = 8.16, $p < .001$; *95% C.I.* [1.12, 1.83]). When the target stimulus was the arrows, higher accuracy was found on congruent (M = 0.98, SD = 0.13) than on incongruent (M = 0.92, SD = 0.26; *Estimate* = 1.54, *SE* = 0.14, *z* = 11.25, $p < .001$) trials. Instead, when the target stimulus was the face, no significant difference in accuracy was found between congruent (M = 0.961, SD = 0.19) and incongruent (M = 0.959, SD = 0.20; *Estimate* = 0.07, *SE* = 0.12, *z* = 0.58, *p* = .56) trials. Finally, the triple interaction between Congruency, Fixation, and Target Stimulus was not significant (*Estimate* = −0.44, *SE* = 0.36, *z* = −1.22, *p* = .22; *95% C.I.* [−1.15, 0.27]).

2.1.5. Discussion

Responses were generally slower for gaze stimuli compared to arrow stimuli. This aligns with earlier studies (e.g., Dalmaso et al., 2023; Marotta et al., 2018; Marotta et al., 2019; Román-Caballero et al., 2021a) and may indicate greater perceptual complexity of eye-gaze stimuli (see also for similar results, Hietanen et al., 2006; Vlamings et al., 2005). Moreover, an SCE emerged for arrows, while an RCE emerged for gazes, thus replicating the main pattern that generally emerges when this task is employed (e.g., Dalmaso et al., 2023; Hemmerich et al., 2022; Marotta et al., 2018). Importantly, this pattern was not further qualified by the type of the fixation item, as similar results emerged for both the symbolic cross and the meaningful real-world object fixation items. Therefore, we did not find evidence supporting our main hypothesis based on joint attention. We reasoned that it could be possible that the two types of fixation items have been ineffective in shaping the RCE for gaze because they were intermixed within the trials, thus likely favouring the emergence of possible carry-over effects, which could have masked any potential differences between symbolic crosses and real-world objects. Indeed, it has been suggested that how stimuli of different categories are presented (either in a blocked or intermixed fashion) can deeply impact social orienting responses (e.g., Dalmaso et al., 2023; Pavan et al., 2011; Zhang et al., 2023). In particular, the intermixed presentation of stimuli creates a more dynamic context and promotes continuous comparison between stimuli, unlike the blocked presentation, which lacks this ongoing contrast. For this reason, in Experiment 2, the two fixation items (i.e., the symbolic cross or the real-world object) were presented in two distinct blocks of trials (see Fig. 1), to avoid potential mutual influences between the two item categories.

3. Experiment 2

Everything was identical to Experiment 1, except that in Experiment

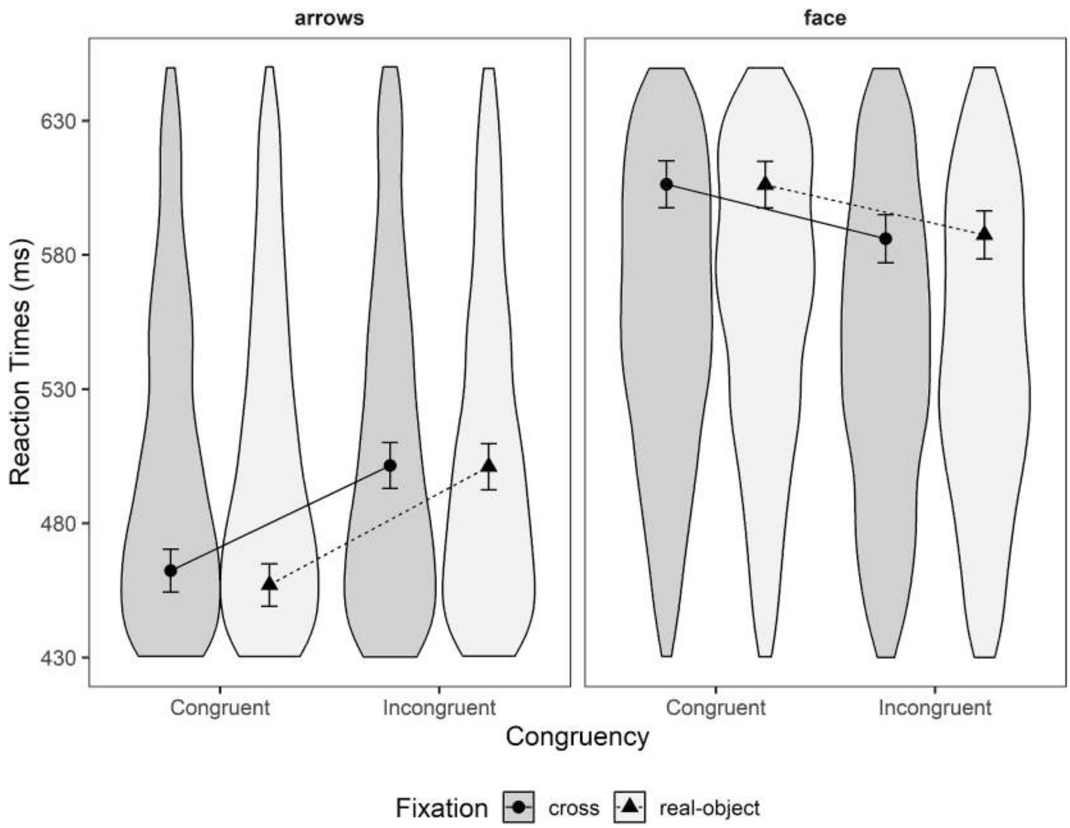

Fig. 2. Model-based estimated marginal means of interaction terms of the selected model for target detection time in milliseconds (Experiment 1). An SCE emerged with arrows, while an RCE was found with faces. Error bars indicate 95% confidence intervals.

2 also the fixation items were presented in two distinct blocks of trials, in order to avoid influence (such as a carry-over effect) between the two categories. As in Experiment 1, we expected a greater RCE for eye-gaze stimuli when presented alongside a real-world object than a symbolic cross.

3.1. Methods, results, discussion

3.1.1. Participants

Given that the experimental design was identical to Experiment 1, we decided to test an identical number of participants. Hence, a new sample of 60 individuals (38 females; *mean age* = 26.47 years, *SD* = 2.61; five left-handed) was recruited using Prolific (as for Experiment 1, with the same monetary compensation). The research protocol was approved by the Ethics Committee for Psychological Research of the University of Padova and was conducted in accordance with the Declaration of Helsinki.

3.1.2. Apparatus, stimuli, and procedure

Everything was identical to Experiment 1, except that both the targets and the fixations were presented in distinct blocks (see also Fig. 1, right panel). Four versions of the task were created to counterbalance the order of presentation of both the target and the fixation stimuli. Each version consisted of four blocks in which both the target and the fixation type were kept constant. The four versions were: face-cross (block 1), face-object (block 2), arrow-cross (block 3), arrow-object (block 4); face-object (block 1), face-cross (block 2), arrow-object (block 3), arrow-cross (block 4); arrow-cross (block 1), arrow-object (block 2), face-cross (block 3), face-object (block 4); arrow-object (block 1), arrow-cross (block 2), face-object (block 3), face-cross (block 4). Fifteen participants were randomly assigned to each of the four versions. As for Experiment 1, the whole experiment included 256 trials (64 for each experimental block) and lasted about 15 min. A short break was granted at the end of each block. Participants proceeded to the experimental trials after completing 10 practice trials.

3.1.3. Analysis

Data were analysed as in Experiment 1.

3.1.4. Results

Trials in which a missed response was recorded (0.79% of the trials) were discarded and not further analysed. Error trials (4.81%) were excluded from RTs. Nonetheless, they were analysed separately for completeness, although the primary analyses concerned RTs. Correctly-responded trials with a latency smaller than 200 ms or greater than 1300 ms were discarded (0.02% of the trials). Following data cleaning, each experimental condition had at least 1738 observations, and therefore statistical power was adequate (Brysbaert & Stevens, 2018).

As for Experiment 1, we considered as experimental factors Congruency (Congruent vs. Incongruent), Target Stimulus (Face vs. Arrow) and Fixation (Cross vs. Real-world object), and their interactions as fixed effects. In addition, the by-participants random slope for the effect of Congruency was chosen.

According to the AIC-based model comparison procedure, the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the RT data (AIC weight = 0.71; see Table S2 in the Supplementary Materials). This model outperformed the model additionally including the three-way interaction between Congruency, Target Stimulus, and Fixation (AIC weight = 0.29). Parameter estimates from this model revealed a main effect of Congruency (*Estimate* = 9.53, *SE* = 3.67, *t* = 2.59, *p* = .01; *95% C.I.* [2.33, 16.73]), showing shorter RTs on congruent trials (*M* = 525.98 ms, *SD* = 129.15) than on incongruent trials (*M* = 536.23 ms, *SD* = 116.69). In addition, a main effect of Target Stimulus was found (*Estimate* = 107.06, *SE* = 1.50, *t* = 71.60, $p < .001$; *95% C.I.* [104.13, 110.00]), due to shorter RTs for arrows (*M* = 477.32 ms, *SD* = 110.91) than for faces (*M* = 584.85 ms, *SD* = 110.93). A main effect of Fixation was also found (*Estimate* = −5.70, *SE* = 1.49, *t* = −3.83, $p < .001$; *95% C.I.* [−8.62, −2.79]), showing slower RTs when the fixation was a cross (*M* = 533.74 ms, *SD* = 122.93), with respect to the real-world object (*M* = 528.31 ms, *SD* = 123.54).

Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = −58.93, *SE* = 2.99, *t* = −19.71, $p < .001$; *95% C.I.* [−64.79, −53.07]). This interaction revealed that when the target stimulus was the arrows, shorter RTs were found on congruent (*M* = 458.81 ms, *SD* = 108.82) than on incongruent (*M* = 496.97 ms, *SD* = 109.74; *Estimate* = −38.99, *SE* = 3.85, *z* = −10.12, $p < .001$) trials. Instead, when the target stimulus was the face, the reversed pattern was found, with shorter RTs for incongruent (*M* = 574.28 ms, *SD* = 110.46) than for congruent (*M* = 595.48 ms, *SD* = 110.40; *Estimate* = 19.93, *SE* = 4.07, *z* = 4.89, $p < .001$) trials. Finally, the interaction between Fixation and Target Stimulus was significant (*Estimate* = 8.01, *SE* = 2.98, *t* = 2.68, *p* = .007; *95% C.I.* [2.16, 13.85]). Even if this interaction was irrelevant to our hypotheses, we further explored it for completeness. When the target stimulus was the arrows, shorter RTs were found for the real-world objects (*M* = 471.98 ms, *SD* = 107.22) than with the cross (*M* = 482.61 ms, *SD* = 114.22; *Estimate* = 9.71, *SE* = 1.88, *z* = 5.17, $p < .001$). Instead, when the target stimulus was the face, no differences were found between the real-world objects (*M* = 584.77 ms, *SD* = 112.56) and the cross (*M* = 584.92 ms, *SD* = 109.30; *Estimate* = 1.70, *SE* = 2.31, *z* = 0.73, *p* = .46). Finally, the triple interaction between Congruency, Fixation and Target Stimulus, as estimated in the full model, was not significant (*Estimate* = −2.80, *SE* = 5.97, *t* = −0.47, *p* = .64; *95% C.I.* [−14.49, 8.90]; see also Fig. 3).

As for accuracy, the model comparison revealed that the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the data, outperforming the model additionally including the three-way interaction between Congruency, Fixation, and Target Stimulus ($\chi^2(1) = 0.05$, *p* = .82). Parameter estimates from this model revealed a main effect of Congruency (*Estimate* = −0.71, *SE* = 0.09, *z* = −7.59, $p < .001$; *95% C.I.* [−0.90, −0.53]), showing higher accuracy, on average, for congruent (M = 0.97, SD = 0.16) than for incongruent trials (M = 0.95, SD = 0.22). The main effect of Target Stimulus was not significant (*Estimate* = −0.18, *SE* = 0.09, *z* = −1.91, *p* = .055), with comparable accuracy for arrows (M = 0.959, SD = 0.20) and faces (M = 0.961, SD = 0.19). Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = 1.61, *SE* = 0.19, *z* = 8.56, $p < .001$; *95% C.I.* [1.24, 1.97]). When the target stimulus was the arrows, higher accuracy was found on congruent (M = 0.98, SD = 0.12) than on incongruent (M = 0.93, SD = 0.25; *Estimate* = 1.52, *SE* = 0.14, *z* = 10.47, $p < .001$) trials. Instead, when the target stimulus was the face, no significant difference in accuracy was found between congruent (M = 0.96, SD = 0.20) and incongruent (M = 0.963, SD = 0.19; *Estimate* = −0.09, *SE* = 0.12, *z* = −0.75, *p* = .45) trials. Finally, the triple interaction between Congruency, Fixation, and Target Stimulus was not significant (*Estimate* = 0.09, *SE* = 0.38, *z* = 0.23, *p* = .82; *95% C.I.* [−0.65, 0.82]).

3.1.5. Discussion

The results were largely consistent with those reported in Experiment 1, except for a main effect of Fixation, with overall faster responses when fixation was a real-world object, and a significant interaction between Fixation and Target Stimulus, whereby responses to arrows were faster when the fixation item was a real-world object compared to a symbolic cross, while no such difference emerged for gaze stimuli. The theoretical implications of this interaction will be discussed in the

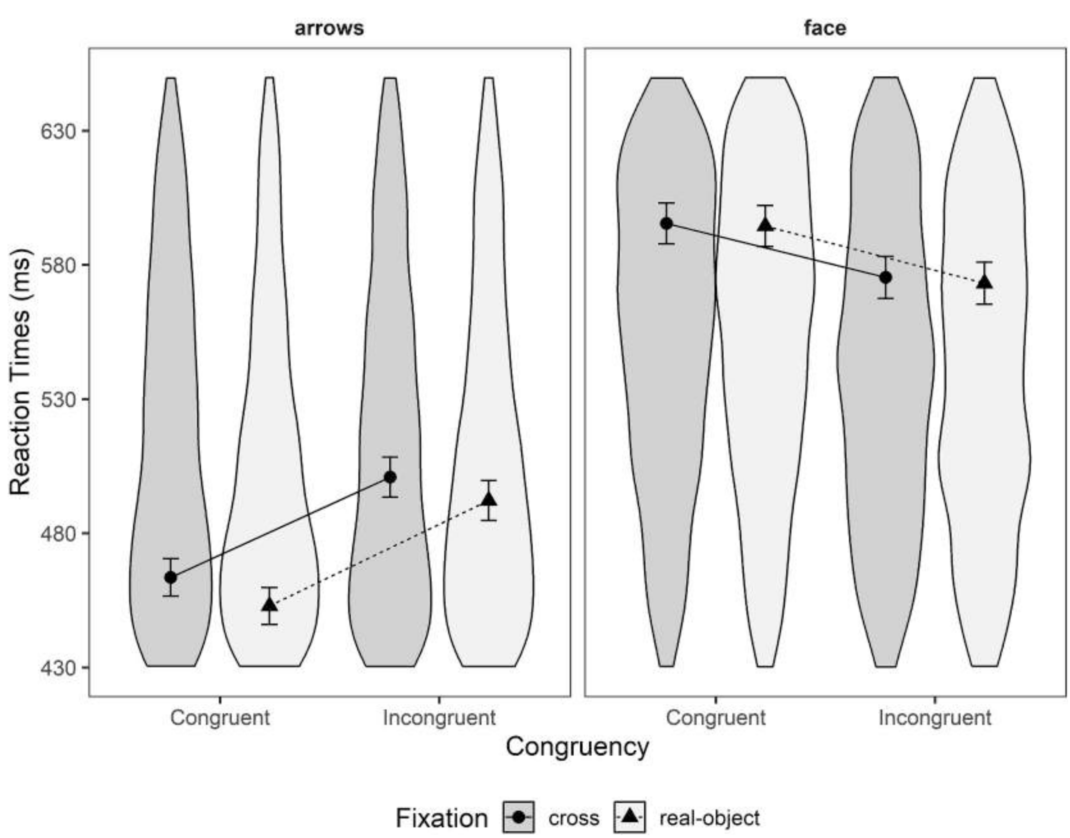

Fig. 3. Model-based estimated marginal means of interaction terms of the selected model for target detection time in milliseconds (Experiment 2). An SCE emerged with arrows, while an RCE was found with faces. Error bars indicate 95% confidence intervals.

General Discussion. Gaze stimuli elicited slower responses than arrow stimuli, an SCE emerged for arrows, whereas an RCE emerged for eye-gaze stimuli (e.g., Dalmaso et al., 2023; Hemmerich et al., 2022; Marotta et al., 2018). Again, these results emerged regardless of the type of fixation (i.e., a symbolic cross or a real-world object), thus suggesting that the blocked presentation was ineffective in shaping this phenomenon.

In Experiment 3, we decided to introduce an oculomotor component to incentivise the emergence of joint attention: at the beginning of each trial, participants were first asked to fixate on the bottom part of the screen and then to move their eyes toward the central fixation item. The initiation of an eye movement toward a potentially relevant item is crucial for the emergence of joint attention, which is, by definition, the movement of one's attention toward the same item looked at by another agent (e.g., Emery, 2000; Pfeiffer-Leßmann et al., 2012; Tomasello & Carpenter, 2007). This is also sustained by prior empirical evidence showing that a particular form of social attention, namely the gaze-leading effect (i.e., the tendency to orient attention toward the face that is looking at the same object we are fixating on; Edwards et al., 2015) emerges only when participants are asked to initiate joint attention toward a real-world item, but it is absent when no initial eye movements are requested of participants. For consistency with previous experiments, we decided to replicate and extend both Experiments 1 and 2 by adopting the same spatial Stroop task and adding an initial oculomotor component.

4. Experiment 3

Everything was identical to Experiment 1, except that participants were first asked to look at a stimulus presented at the bottom part of the screen and then to make an eye movement toward the fixation item. Again, we expected a greater RCE for eye-gaze stimuli when presented alongside a real-world object than a symbolic cross, while no differences were expected for arrows.

4.1. Methods, results, discussion

4.1.1. Participants

Given that the experimental design was identical to Experiments 1 and 2, we decided to test an identical number of participants. Hence, a new sample of 60 individuals (39 females; *mean age* = 22.66 years, *SD* = 2.81; four left-handed) was recruited. All participants were recruited within the student population of the University of Padova. The participants were compensated with 5 €. The research protocol was approved by the Ethics Committee for Psychological Research of the University of Padova and was conducted in accordance with the Declaration of Helsinki.

4.1.2. Apparatus, stimuli, and procedure

To create a dynamic experimental context and to ensure that participants made an eye movement from the bottom to the centre of the screen, Experiment 3 was performed in the laboratory using an EyeLink 1000 Plus (SR Research Ltd., Ottawa, Canada). Participants sat 70 cm away from a 24-in. monitor (1280 × 1024 pixels, 120 Hz), and a chinrest was used to avoid head movements. A display PC running Experiment Builder (SR Research Ltd., Ottawa, Canada) handled timing and stimuli presentation. Eye movements were recorded monocularly at 1000 Hz.

Each experimental session started with a five-point calibration followed by a validation procedure. As for stimuli and procedure, everything was identical to Experiment 1 except that, at the beginning of each trial, the word 'START' appeared at the bottom of the screen. A word stimulus was chosen (instead, for instance, of a pictorial stimulus) to avoid any perceptual similarities with the other stimuli used as central fixation items. Participants were initially asked to look at the word. The trial continued only if participants maintained their eyes on this word for 500 ms. Afterwards, the fixation point appeared in the centre of the screen, and participants were asked to move their eyes from the word to the fixation point (see Fig. 4, left panel). As for Experiment 1, two versions of the task were created to counterbalance the order of target presentation. In one version, participants completed a first block containing only face stimuli followed by a block containing only arrow stimuli. In another version, the order was reversed. Thirty participants were randomly assigned to each of the two versions. The whole experiment included 256 trials (64 for each experimental block) and lasted about 30 min. A short break was granted at the end of each block. Participants proceeded to the experimental trials after completing 10 practice trials.

4.1.3. Analysis

Data were analysed as in Experiments 1 and 2. Eye movement data were not analysed because the experiment was not designed to test oculomotor hypotheses.

4.1.4. Results

Trials in which a missed response was recorded (0.29% of the trials) were discarded and not further analysed. Error trials (6.47%) were excluded from the RT analysis. Nonetheless, they were analysed separately for completeness, although the primary analyses concerned RTs. Correctly-responded trials with a latency smaller than 200 ms or greater than 1300 ms were discarded (2.78% of the trials). Following data cleaning, each experimental condition had at least 1645 observations, ensuring adequate statistical power (Brysbaert & Stevens, 2018).

We considered as experimental factors Congruency (Congruent vs. Incongruent), Target Stimulus (Face vs. Arrow) and Fixation (Cross vs. Real-world object), and their interactions as fixed effects. We followed a model comparison approach with the random effects proposed by Dalmaso et al. (2023) to select the random effects. Due to model convergence issues, the by-participants random slope for the effect of Congruency was chosen.

According to the AIC-based model comparison procedure, the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the RT data (AIC weight = 0.58; see Table S3 in the Supplementary Materials). This model outperformed the model additionally including the three-way interaction between Congruency, Target Stimulus, and Fixation (AIC weight = 0.42). Parameter estimates from this model revealed a main effect of Target Stimulus (*Estimate* = 139.14, *SE* = 1.74, t = 79.94, $p < .001$; *95% C.I.* [135.73, 142.55]), due to shorter RTs for arrows (*M* = 465.65 ms, *SD* = 118.89) than for faces (*M* = 603.37 ms, *SD* = 138.24). The main effect of Congruency was not significant (*Estimate* = 3.48, *SE* = 4.97, t = 0.70, p = .48; *95% C.I.* [−6.27, 13.22]), nor was the main effect of Fixation (*Estimate* = 3.22, *SE* = 1.73, t = 1.87, p = .062; *95% C.I.* [−0.16, 6.60]).

Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = −63.11, *SE* = 3.48, t = −18.14, $p < .001$; *95% C.I.* [−69.93, −56.29]). This interaction revealed that when the target stimulus was the arrows, shorter RTs were found on congruent (*M* = 449.12 ms, *SD* = 120.81) than on incongruent (*M* = 482.97 ms, *SD* = 114.33; *Estimate* = −35.03, *SE* = 5.09, z = −6.88, $p < .001$) trials. Instead, when the target stimulus was the face, the reversed pattern was found, with shorter RTs for incongruent (*M* = 587.77 ms, *SD* = 135.75) than for congruent (*M* = 619.19 ms, *SD* = 138.96; *Estimate* = 28.08, *SE* = 5.44, z = 5.16, $p < .001$) trials. Finally, the triple interaction between Congruency, Fixation and Target Stimulus, as estimated in the full model, failed to reach significance (*Estimate* = −8.08, *SE* = 6.92, t = −1.17, p = .24; *95% C.I.* [−21.64, 5.48]; see also Fig. 5).

As for accuracy, the model comparison revealed that the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the data, outperforming the model additionally including the three-way interaction between Congruency, Fixation, and Target Stimulus ($\chi^2(1)$ = 0.38, p = .53). Parameter

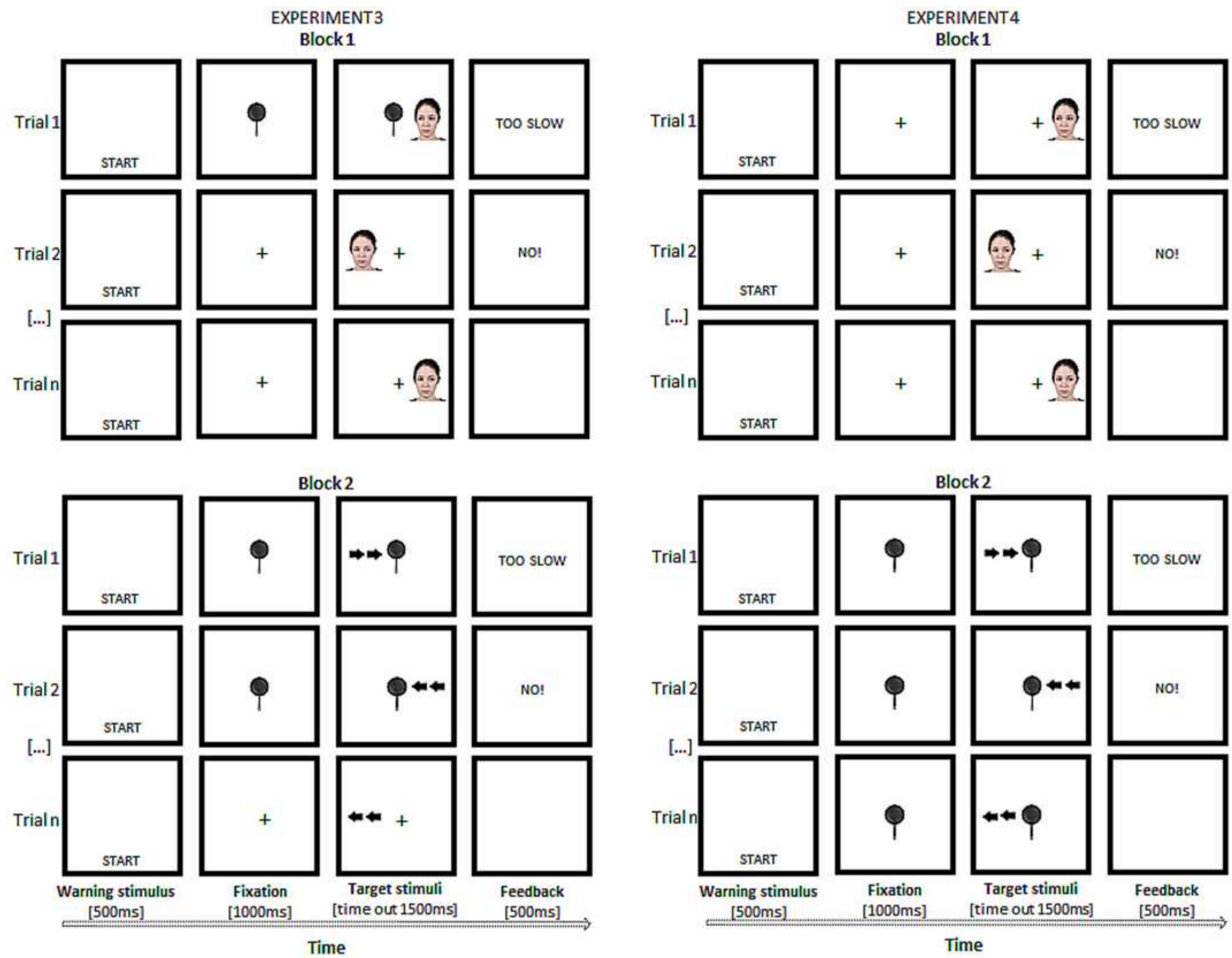

Fig. 4. Examples of trials and stimuli (not drawn to scale) used in the study. The left part of the graph describes the procedure of Experiment 3, while the right part refers to Experiment 4. The face with the averted gaze (Block 1), or arrows (Block 2) appeared either to the left or right of the symbolic cross or the real-world object. While in both experiments, the target stimuli (face or arrows) were presented in different blocks (counterbalanced across participants), the fixation item (symbolic cross or real-world object) was intermixed within the block for Experiment 3 but presented in different blocks for Experiment 4. At the beginning of each trial, the word 'START' (Warning stimulus) appeared at the bottom of the screen. In all blocks, visual feedback was given for slow or incorrect responses, while a blank screen followed correct responses.

estimates from this model revealed a main effect of Congruency (*Estimate* = −0.78, *SE* = 0.09, $z = -8.07$, $p < .001$; *95% C.I.* [−0.97, −0.59]), showing higher accuracy, on average, for congruent (M = 0.95, SD = 0.23) than for incongruent trials (M = 0.93, SD = 0.26). The main effect of Target Stimulus was significant (*Estimate* = −1.54, *SE* = 0.10, $z = -15.82$, $p < .001$; *95% C.I.* [−1.73, −1.35]), with higher accuracy for arrows (M = 0.97, SD = 0.18) than for faces (M = 0.91, SD = 0.30). Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = 1.84, *SE* = 0.19, $z = 9.48$, $p < .001$; *95% C.I.* [1.46, 2.22]). When the target stimulus was the arrows, higher accuracy was found on congruent (M = 0.99, SD = 0.10) than on incongruent (M = 0.95, SD = 0.23; *Estimate* = 1.70, *SE* = 0.18, $z = 9.67$, $p < .001$) trials. Instead, when the target stimulus was the face, no significant difference in accuracy was found between congruent (M = 0.90, SD = 0.31) and incongruent (M = 0.91, SD = 0.29; *Estimate* = −0.14, *SE* = 0.08, $z = -1.68$, $p = .09$) trials. Finally, the triple interaction between Congruency, Fixation, and Target Stimulus was not significant (*Estimate* = −0.25, *SE* = 0.39, $z = -0.62$, $p = .53$; *95% C.I.* [−1.01, 0.52]).

4.1.5. Discussion

The results were largely similar to those reported in Experiments 1 and 2. Gaze stimuli elicited slower responses than arrow stimuli, an SCE emerged for arrows, whereas an RCE emerged for eye-gaze stimuli. These results emerged regardless of the type of fixation (i.e., a symbolic cross or a real-world object), thus suggesting that the oculomotor component had no role in impacting the RCE for gaze. Notably, the main effect of Congruency, which was significant in Experiments 1 and 2, did not reach significance in the present experiment; nevertheless, the critical Congruency × Target Stimulus interaction remained significant and followed the same pattern observed in the previous experiments.

In the next and final Experiment 4, we replicated and extended Experiment 2 to explore whether the possible impact of joint attention in the RCE for gaze could emerge in a blocked design. Hence, the two fixation items (i.e., the symbolic cross or the real-world object) were presented in two blocks of trials (as for Experiment 2). Moreover, as in Experiment 3, at the beginning of each trial, participants were first asked to fixate on a stimulus placed at the bottom part of the screen and then to make an eye movement toward the central fixation item (see Fig. 4, right panel).

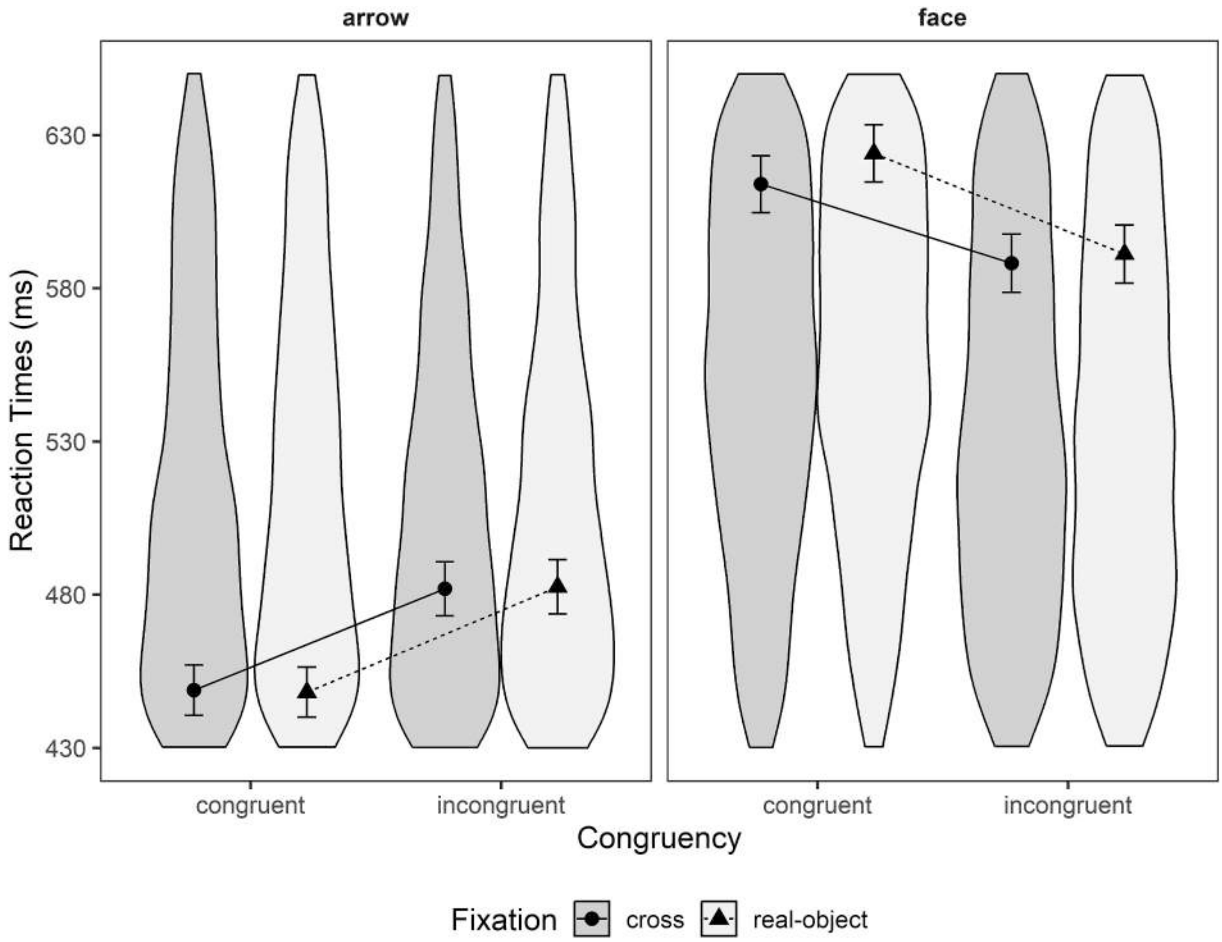

Fig. 5. Model-based estimated marginal means of interaction terms of the selected model for target detection time in milliseconds (Experiment 3). An SCE emerged with arrows, while an RCE was found with faces. Error bars indicate 95% confidence intervals.

5. Experiment 4

Everything was identical to Experiment 3, except that (like in Experiment 2) fixation items were presented within two blocks of trials to avoid any influence (e.g., carry-over effects) between the two categories. As in previous experiments, we expected a greater RCE for eye-gaze stimuli when presented alongside a real-world object than a symbolic cross.

5.1. Methods, results, discussion

5.1.1. Participants

A new sample of 60 individuals (47 females; *mean age* = 20.95 years, *SD* = 1.88; nine left-handed) was recruited. All participants were recruited within the student population of the University of Padova. The participants were compensated with 5 €. The research protocol was approved by the Ethics Committee for Psychological Research of the University of Padova and was conducted in accordance with the Declaration of Helsinki.

5.1.2. Apparatus, stimuli, and procedure

Everything was identical to Experiment 3, except that (like in Experiment 2) both the targets and the fixations were presented in distinct blocks (see Fig. 4, right panel).

5.1.3. Analysis

Data were analysed as in previous experiments.

5.1.4. Results

Trials in which a missed response was recorded (0.43% of the trials) were discarded and not further analysed. Error trials (5.67%) were excluded from RTs analysis. Nonetheless, they were analysed separately for completeness, although the primary analyses concerned RTs. Correctly-responded trials with a latency smaller than 200 ms or greater than 1300 ms were discarded (2.73% of the trials). Following data cleaning, each experimental condition had at least 1627 observations, ensuring adequate statistical power (Brysbaert & Stevens, 2018).

We considered as experimental factors Congruency (Congruent vs. Incongruent), Target Stimulus (Face vs. Arrow) and Fixation (Cross vs. Real-world object), and their interactions as fixed effects. Due to model convergence issues, the by-participants random slope for the effect of Congruency was chosen.

According to the AIC-based model comparison procedure, the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the RT data (AIC weight = 0.72; see Table S4 in the Supplementary Materials). This model outperformed the model additionally including the three-way interaction between Congruency, Target Stimulus, and Fixation (AIC weight = 0.28). Parameter estimates from this model revealed a main effect of Target Stimulus (*Estimate* = 152.88, *SE* = 1.66, $t = 91.95$, $p < .001$; *95% C.I.* [149.62, 156.14]), due to shorter RTs for arrows (*M* = 455.43 ms, *SD* = 100.76) than for faces (*M* = 607.86 ms, *SD* = 134.44). A main effect of Fixation was also found (*Estimate* = 6.83, *SE* = 1.65, $t = 4.13$, $p < .001$; *95% C.I.* [3.59, 10.07]), showing faster RTs when the fixation was a cross (*M* = 524.51 ms, *SD* = 137.9), with respect to the real-world object (*M* =

530.12 ms, *SD* = 142.59). The main effect of Congruency was not significant (*Estimate* = 1.33, *SE* = 4.54, *t* = 0.29, *p* = .77; *95% C.I.* [−7.56, 10.22]).

Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = −67.42, *SE* = 3.32, *t* = −20.29, *p* < .001; *95% C.I.* [−73.93, −60.91]). This interaction revealed that when the target stimulus was the arrows, shorter RTs were found on congruent (*M* = 438.42 ms, *SD* = 100.53) than on incongruent (*M* = 472.86 ms, *SD* = 97.99; *Estimate* = −35.04, *SE* = 4.64, *z* = −7.55, *p* < .001) trials. Instead, when the target stimulus was the face, the reversed pattern was found, with shorter RTs for incongruent (*M* = 590.67 ms, *SD* = 131.97) than for congruent (*M* = 624.97 ms, *SD* = 134.71; *Estimate* = 32.38, *SE* = 5.02, *z* = 6.46, *p* < .001) trials. Finally, the interaction between Fixation and Target Stimulus was significant (*Estimate* = 13.28, *SE* = 3.31, *t* = 4.01, *p* < .001; *95% C.I.* [6.79, 19.78]). When the target stimulus was the arrows, no differences were found between the real-world objects (*M* = 456.00 ms, *SD* = 100.85) and the cross (*M* = 454.87 ms, *SD* = 100.68; *Estimate* = −0.19, *SE* = 1.91, *z* = −0.10, *p* = .92). Instead, when the target stimulus was the face, shorter RTs were found with the cross (*M* = 601.85 ms, *SD* = 132.34) than with the real-world objects (*M* = 614.00 ms, *SD* = 136.3; *Estimate* = −13.47, *SE* = 2.71, *z* = −4.97, *p* < .001). Finally, the triple interaction between Congruency, Fixation and Target Stimulus, as estimated in the full model, was not significant (*Estimate* = −1.65, *SE* = 6.63, *t* = −0.25, *p* = .80; *95% C.I.* [−14.64, 11.34]; see also Fig. 6).

As for accuracy, the model comparison revealed that the model including the two-way interactions between Congruency and Fixation, Congruency and Target Stimulus, and Fixation and Target Stimulus provided the best overall fit to the data, outperforming the model additionally including the three-way interaction between Congruency, Fixation, and Target Stimulus ($\chi^2(1) = 0.72$, *p* = .39). Parameter estimates from this model revealed a main effect of Congruency (*Estimate* = −0.62, *SE* = 0.10, *z* = −6.45, *p* < .001; *95% C.I.* [−0.81, −0.43]), showing higher accuracy, on average, for congruent (M = 0.96, SD = 0.22) than for incongruent trials (M = 0.94, SD = 0.25). The main effect of Target Stimulus was significant (*Estimate* = −1.55, *SE* = 0.10, *z* = −16.04, *p* < .001; *95% C.I.* [−1.74, −1.36]), with higher accuracy for arrows (M = 0.98, SD = 0.15) than for faces (M = 0.92, SD = 0.29). The main effect of Fixation was also significant (*Estimate* = −0.22, *SE* = 0.09, *z* = −2.51, *p* = .01; *95% C.I.* [−0.40, −0.05]), with higher accuracy for cross (M = 0.953, SD = 0.22) than for real-objects (M = 0.946, SD = 0.25). Critically, the interaction between Congruency and Target Stimulus was significant (*Estimate* = 1.08, *SE* = 0.19, *z* = 5.61, *p* < .001; *95% C.I.* [0.70, 1.45]). When the target stimulus was the arrows, higher accuracy was found on congruent (M = 0.99, SD = 0.11) than on incongruent (M = 0.96, SD = 0.19; *Estimate* = 1.16, *SE* = 0.17, *z* = 6.72, *p* < .001) trials. Instead, when the target stimulus was the face, no significant difference in accuracy was found between congruent (M = 0.92, SD = 0.29) and incongruent (M = 0.92, SD = 0.30; *Estimate* = 0.08, *SE* = 0.08, *z* = 0.97, *p* = .33) trials. Finally, the triple interaction between Congruency, Fixation, and Target Stimulus was not significant (*Estimate* = −0.33, *SE* = 0.39, *z* = −0.86, *p* = .39; *95% C.I.* [−1.10, 0.43]).

5.1.5. Discussion

The results of Experiment 4 were largely consistent with those reported in previous experiments, further replicating the Fixation × Target

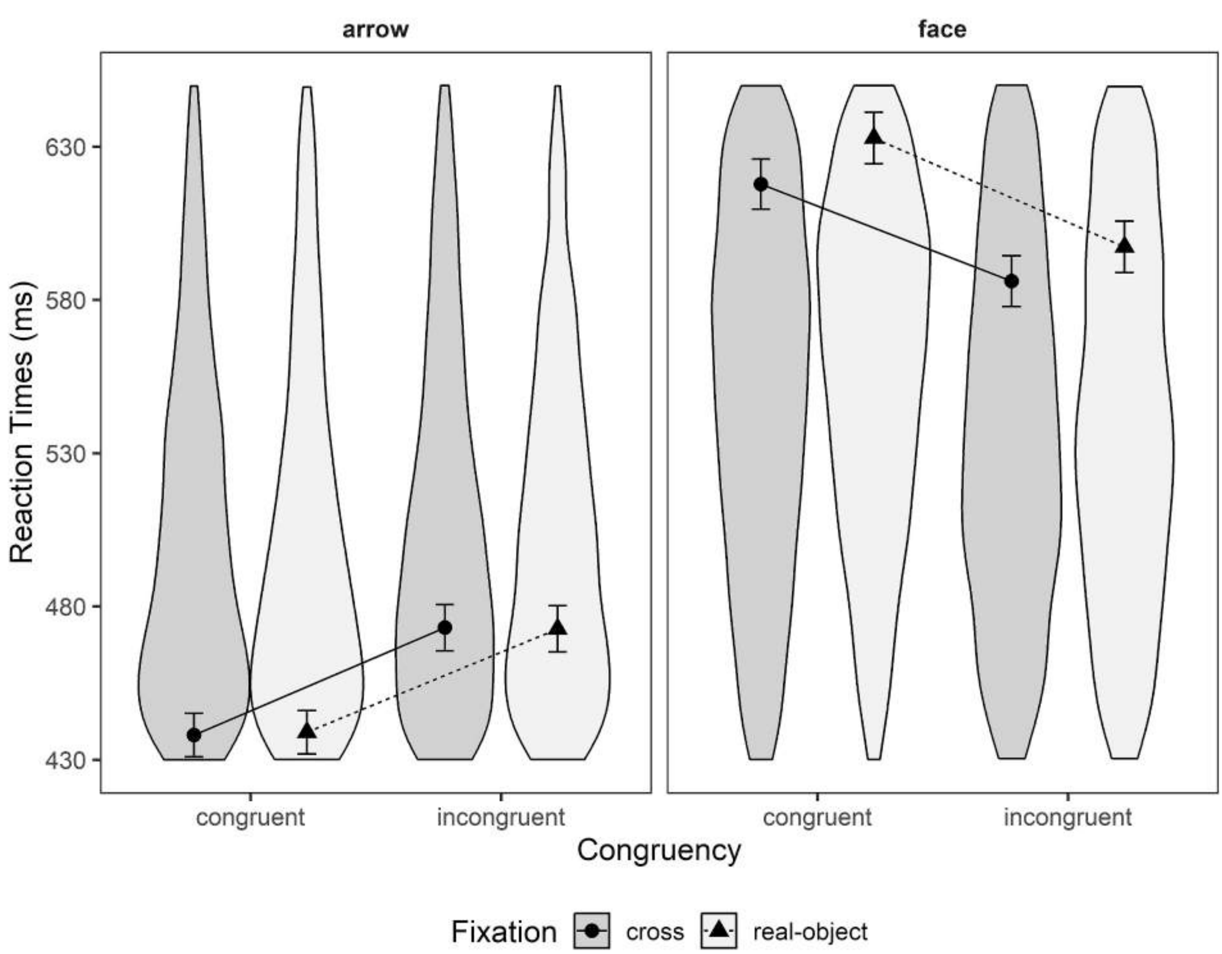

Fig. 6. Model-based estimated marginal means of interaction terms of the selected model for target detection time in milliseconds (Experiment 4). An SCE emerged with arrows, while an RCE was found with faces. Error bars indicate 95% confidence intervals.

Stimulus interaction. Gaze stimuli elicited slower responses than arrow stimuli, an SCE emerged for arrows, whereas an RCE emerged for eye-gaze stimuli. Again, these results emerged regardless of the type of fixation (i.e., a symbolic cross or a real-world object), thus suggesting that the beginning of an eye movement toward a potentially relevant item and the blocked presentation of such stimuli were irrelevant for shaping this phenomenon.

6. Cross-experiment meta-analysis

To provide a quantitative synthesis of the primary finding across the four experiments, we conducted a meta-analysis on the parameter estimates of the critical Congruency × Fixation × Target Stimulus interaction, as obtained from the full mixed-effects models fitted separately for each experiment. Because these estimates and their standard errors already account for the clustering of trials within participants (via the by-participant random effects structure), they were meta-analysed directly, rather than being converted into standardised effect sizes based on the number of trials, which would have ignored this clustering and yielded artificially narrow confidence intervals. The interaction estimates were small and non-significant in every experiment (all $|z| < 1.2$, all $p > .24$; Exp. 1: *Estimate* = −3.40 ms, *SE* = 6.45; Exp. 2: *Estimate* = −2.80 ms, *SE* = 5.97; Exp. 3: *Estimate* = −8.08 ms, *SE* = 6.92; Exp. 4: *Estimate* = −1.65 ms, *SE* = 6.63), and consistently smaller than half the average RCE magnitude (±13 ms; see equivalence test below). A fixed-effects meta-analysis yielded a combined estimate of −3.83 ms (*95% C.I.* [−10.16, 2.51], $z = -1.18$, $p = .236$), with no evidence of heterogeneity across experiments ($Q(3) = 0.52$, $p = .915$, $I^2 = 0\%$). This result provides converging evidence that fixation type did not modulate the RCE across the four experiments (see also Fig. 7).

To formally evaluate whether the null effect obtained from the meta-analysis reflected genuine equivalence to zero rather than mere non-significance, we conducted a two one-sided test (TOST) equivalence procedure (Lakens, 2017) on the meta-analytic estimate, using the TOSTER package (see Lakens et al., 2018) in R. The equivalence bound was set to ±13 ms, corresponding to half the average magnitude of the RCE observed across the four experiments ($M_{exp} = 26.98$ ms), reflecting the minimum modulation that would be theoretically meaningful given the size of the effect under investigation. Both one-sided tests were significant ($z = -5.21$, $p < .001$; $z = 2.84$, $p = .002$), indicating that the combined estimate fell statistically within the equivalence bounds (*90% C.I.* [−9.14, 1.49]), while the conventional null-hypothesis test remained non-significant ($z = -1.18$, $p = .236$). Together, these results provide direct statistical evidence that fixation type does not meaningfully modulate the RCE, beyond the mere absence of a significant effect.

7. General discussion

When eye-gaze and arrow stimuli are employed in a spatial Stroop task, an SCE is generally observed for arrows and an RCE for gaze (Chacón-Candia et al., 2023; Hemmerich et al., 2022; Ishikawa et al., 2021, 2025b; Tanaka et al., 2025). One theoretical account that may explain this dissociation is the 'joint attention' hypothesis (Edwards et al., 2020), according to which, on incongruent trials involving gaze cues, the eyes of the face are directed toward the central fixation point, namely the same spatial location gazed at by the participant. This is, by definition, an episode of joint attention, which cannot be applied to arrows as they lack any social intentionality (e.g., Emery, 2000).

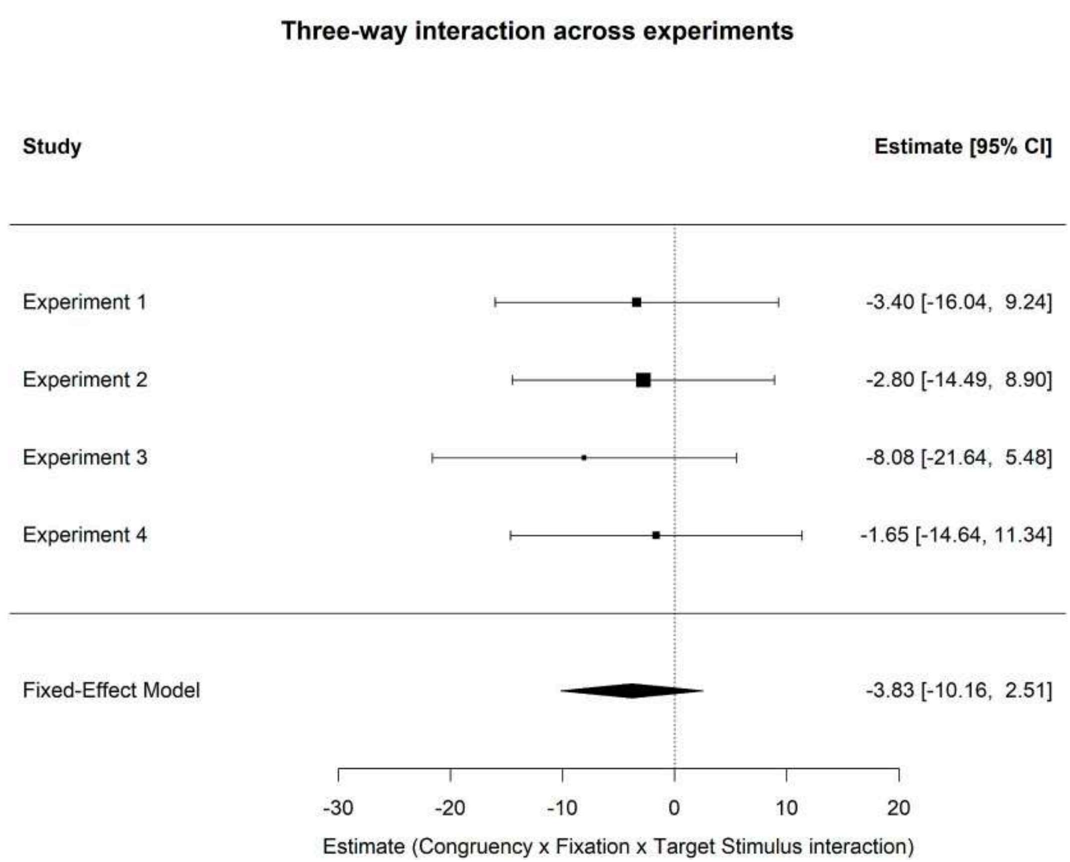

Fig. 7. Forest plot of the parameter estimates for the Congruency × Fixation × Target Stimulus interaction across the four experiments, obtained from the mixed-effects models fitted separately for each experiment. Each square represents the interaction estimate (in ms) for a single experiment, with horizontal lines indicating 95% confidence intervals. The diamond at the bottom represents the combined estimate from the fixed-effects meta-analytic model. The dashed vertical line indicates zero effect. All estimates are close to zero and confidence intervals cross zero, providing converging evidence for the absence of a modulation of the RCE by fixation type across experiments.

The present study aimed to further investigate the role of joint attention in the RCE for gaze through a systematic approach. We conducted four experiments in which we manipulated the nature of the fixation point, which could either be a real-world object (which is the typical item involved in joint attention episodes; e.g., Emery, 2000) or a symbolic one (i.e., a black cross). In addition, we also varied the context through which these two categories were presented (i.e., intermixed in Experiments 1 and 3, blocked in Experiments 2 and 4), and we also inserted the request to make an eye movement toward the central item (Experiments 3 and 4).

Across all experiments, the results were remarkably consistent, revealing a robust SCE for arrow cues and a robust RCE for gaze cues, independently of the type of fixation point, the context, and the presence of the oculomotor component. Our results do not provide evidence supporting our initial hypothesis that the RCE could be due (or at least influenced) to joint attention, at least when operationalised as in the current context. This conclusion is also supported by developmental evidence from Aranda-Martín et al. (2022), who found no evidence of the RCE up to 12 years of age, although the basic abilities underlying joint attention emerge within the first year of life (see Mundy, 2018). This suggests that the mechanisms underlying the RCE emerge later in development, making a direct role for joint attention in the RCE less likely. Our results, therefore, appear to diverge from the joint attention hypothesis proposed by Edwards et al. (2020). This apparent discrepancy may be tentatively resolved by considering a key methodological difference between the two paradigms. Indeed, while Edwards et al. (2020) manipulated both central and peripheral fixation locations, we exclusively presented objects at the centre of the display. Our choice was mainly motivated by the will to more closely align with previous literature on the RCE for eye-gaze stimuli (e.g., Cañadas & Lupiáñez, 2012; Dalmaso et al., 2023; Marotta et al., 2018; Tanaka et al., 2025) and to ensure compatibility with the oculomotor components (introduced here in Experiments 3 and 4) derived from the gaze-leading effect (Edwards et al., 2015). We can tentatively speculate that centrally positioned objects, with eye-gaze stimuli directed toward them, may simultaneously engage mechanisms of both joint attention (Edwards et al., 2020) and eye contact (Marotta et al., 2018), which has recently received further empirical support (Marotta & Chacón-Candia, 2026), thus resulting in overlapping effects. In turn, such an overlap could have masked the contribution of joint attention in our paradigm, making the RCE primarily driven by mechanisms related to the processing of eye-gaze stimuli themselves, rather than by the nature of the fixation object. By contrast, when objects are placed peripherally and are looked at by the face (such as in Edwards et al., 2020), the eye contact component is missing, potentially allowing joint attention effects to emerge more clearly. Nevertheless, this interpretation remains speculative and should be interpreted with caution. Future studies could therefore operationalise joint attention in the same way as in the present work, while also introducing peripheral object positions to better isolate the specific contribution of joint attention to the RCE. Finally, although the Fixation × Target Stimulus interaction was not of primary theoretical interest, its presence in two out of four experiments deserves brief consideration. Notably, however, the pattern was not consistent across experiments, either in terms of which stimulus type was affected (arrows in Experiment 2, gaze in Experiment 4) or in the direction of the effect (in Experiment 2, responses to arrows were faster with real-world objects than with the cross, whereas in Experiment 4, responses to gaze stimuli were faster with the cross than with real-world objects). This inconsistency makes it unlikely that a single, straightforward mechanism (e.g., differences in perceptual complexity between fixation types) can account for the effect, and we therefore refrain from drawing firm conclusions about its origin. Future studies specifically designed to investigate this interaction will be needed to clarify whether it reflects a reliable phenomenon.

It is worth recalling that, other than the three main social accounts proposed to explain the possible nature of the RCE for eye-gaze stimuli (i.e., eye contact, Marotta et al., 2018; joint attention, Edwards et al., 2020; joint distraction, Hemmerich et al., 2022), recent works also assessed the possible role of more low-level perceptual mechanisms. In particular, the dual-stage hypothesis (Román-Caballero et al., 2021a; Tanaka et al., 2025, 2026; see also Ponce et al., 2025a, 2025b; Ponce et al., 2026), proposes that the RCE may arise from difficulties in extracting the target location in complex visual contexts, especially when gaze is embedded in whole faces. To directly test this latter hypothesis, Tanaka et al. (2025, 2026) investigated the temporal dynamics of the RCE by examining RTs distributions in a spatial Stroop task. Using a distributional analysis approach (for a review, see Proctor et al., 2011), reaction times from congruent and incongruent trials were rank ordered from fastest to slowest for each participant and then divided into quartile bins (i.e., four). This method allows for a more fine-grained understanding of how the RCE unfolds over time, revealing whether fast, automatic processes drive the effect or whether it emerges later due to more controlled mechanisms. These values were then visualised as a delta plot, a visualization that maps the magnitude of the congruency effect across time bins. Their findings supported the dual-stage hypothesis: The SCE for arrows decreased with increasing RTs, consistent with the temporal decay of location-based interference, whereas the RCE for gaze stimuli increased over time, reflecting the gradual build-up of selective inhibition.

To better understand the temporal dynamics of cognitive interference and control of the RCE and SCE, we adopted the RT-based distributional approach proposed by Ponce et al. (2025b) and Tanaka et al. (2025, 2026), conducting the analysis separately for each fixation type. The results demonstrate that the standard congruency effect for arrow stimuli decreased as reaction times increased, consistent with the idea that location-based interference weakens over time. This pattern was observed in all experiments except for Experiment 1. In contrast, for the RCE, the pattern of results suggests an initial increase with delayed responses, followed by a stabilisation phase, and then a renewed increase. This was particularly evident in Experiment 3 (with cross fixation) and Experiment 4 (with real-world fixation). Importantly, this analysis offers additional insight into the patterns of RCE and SCE as a function of RTs and fixation type (see Supplemental Material, Fig. S1). Crucially for our hypotheses, across all the experiments, the distributional analyses revealed a similar RCE whether fixation was a real-world object or a symbolic cross. In other words, fixation type did not modulate the RCE, indicating that the joint attention hypothesis (as operationalised here) was not supported, even when considering the temporal dynamics of the effect. Importantly, this temporal dissociation between the SCE and the RCE replicates the distributional patterns reported by Tanaka et al. (2025, 2026) and Ponce et al. (2025b), while extending them by demonstrating that such dynamics are robust across different fixation types.

Finally, some limitations of the current set of experiments could be addressed in future studies. First, although the present set of experiments was designed to increase ecological validity, especially in Experiments 3 and 4 where participants actively engaged with the task, the paradigm still remained relatively artificial. Future studies should further increase its ecological validity by using more realistic settings, for example through virtual reality (e.g., Rubo & Gamer, 2021). Such approaches may provide a more robust operationalisation of joint attention, which is traditionally defined as a dynamic and reciprocal social process (e.g., Emery, 2000). It will therefore be important for future studies to determine whether the present findings reflect a limited role of joint attention in the RCE or limitations in how joint attention has been experimentally operationalised. Moreover, although eye movements were recorded in Experiments 3 and 4, they were not specifically designed to collect reliable oculomotor measures (e.g., saccadic latency or direction). Given that eye movement measures provide a more direct and sensitive index of attentional allocation than manual responses (e.g., Pereira et al., 2022), future studies investigating the nature of the RCE for gaze stimuli should also consider focusing on these measures.

7.1. Conclusions

Across four experiments, our findings provide no evidence supporting the hypothesis that joint attention, as operationalised in the present study, underlies the RCE for gaze. Rather, the results suggest that other mechanisms, both social (e.g., Marotta et al., 2018) and/or perceptual (e.g., Tanaka et al., 2025, 2026), may have played a more prominent role in shaping the observed pattern. While the present study focused specifically on testing the joint attention hypothesis, future work could aim to directly contrast social and perceptual explanations of the RCE. For instance, manipulating the morphological features of the eyes (e.g., inverting sclera/pupil contrast; see, for instance, Ricciardelli et al., 2000) or using faces as target objects may provide a more fine-grained understanding of whether the effect arises from high-level social mechanisms or from low-level perceptual factors. Clarifying the relative contribution of these mechanisms, and their potential interaction, remains an important goal for future research aiming to better understand how gaze cues are processed in complex social environments.

CRediT authorship contribution statement

Anna Lorenzoni: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mario Dalmaso:** Writing – review & editing, Validation, Supervision, Software, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the 'Ministero dell'Università e della Ricerca' (MUR) under the grant 'PRIN 2022 2022H4WZKN'.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2026.104101.

Data availability

Data, experiment codes, and stimuli associated with this work can be found here: https://osf.io/h9f4v/

References

Adams, R. B., Jr., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science, 14*(6), 644–647.
Aranda-Martín, B., Ballesteros-Duperón, M.Á., & Lupiáñez, J. (2022). What gaze adds to arrows: Changes in attentional response to gaze versus arrows in childhood and adolescence. *British Journal of Psychology, 113*(3), 718–738.
Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206–234.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.
Baron-Cohen, S. (1996). Reading the mind in the face: A cross-cultural and developmental study. *Visual Cognition, 3*(1), 39–60.
Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., & Becker, S. I. (2013). 'Gaze leading': Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General, 142*(1), Article 76.
Blair, C. D., Capozzi, F., & Ristic, J. (2017). Where is your attention? Assessing individual instances of covert attentional orienting in response to gaze and arrow cues. *Vision, 1*(3), Article 19.
Böckler, A., van der Wel, R. P., & Welsh, T. N. (2014). Catching eyes: Effects of social and nonsocial cues on attention capture. *Psychological Science, 25*(3), 720–727.
Bonventre, S., & Marotta, A. (2023). Is the reversed congruency effect unique to the eye-gaze? Investigating the effects of finger pointing, eye-gaze and arrows stimuli on spatial interference. *Frontiers in Cognition, 2*, Article 1135435.
Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, Article e9414.
Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), Article 9.
Callejas, A., Shulman, G. L., & Corbetta, M. (2014). Dorsal and ventral attention systems underlie social and symbolic cueing. *Journal of Cognitive Neuroscience, 26*(1), 63–80.
Cañadas, E., & Lupiáñez, J. (2012). Spatial interference between gaze direction and gaze location: A study on the eye contact effect. *Quarterly Journal of Experimental Psychology, 65*(8), 1586–1598.
Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i–174.
Chacón-Candia, J. A., Lupiáñez, J., Casagrande, M., & Marotta, A. (2020). Sex differences in attentional selection following gaze and arrow cues. *Frontiers in Psychology, 11*, Article 95.
Chacón-Candia, J. A., Román-Caballero, R., Aranda-Martín, B., Casagrande, M., Lupiáñez, 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 & Biobehavioral Reviews, 144*, Article 104993.
Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review, 26*(6), 1787–1802.
Dalmaso, M., Castelli, L., & Galfano, G. (2020a). Social modulators of gaze-mediated orienting of attention: A review. *Psychonomic Bulletin & Review, 27*(5), 833–855.
Dalmaso, M., Castelli, L., & Galfano, G. (2020b). Early saccade planning cannot override oculomotor interference elicited by gaze and arrow distractors. *Psychonomic Bulletin & Review, 27*(5), 990–997.
Dalmaso, M., Castelli, L., & Galfano, G. (2021). Increased gaze cueing of attention during COVID-19 lockdown. *iScience, 24*(11).
Dalmaso, M., Galfano, G., & Castelli, L. (2023). Are eyes special? Gaze, but not pointing gestures, elicits a reversed congruency effect in a spatial Stroop task. *Attention, Perception, & Psychophysics, 85*(8), 2547–2557.
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.
Edwards, S. G., Seibert, N., & Bayliss, A. P. (2020). Joint attention facilitates observed gaze direction discrimination. *Quarterly Journal of Experimental Psychology, 73*(1), 80–90.
Edwards, S. G., Stephenson, L. J., Dalmaso, M., & Bayliss, A. P. (2015). Social orienting in gaze leading: A mechanism for shared attention. *Proceedings of the Royal Society B: Biological Sciences, 282*(1812), Article 20151141.
Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604.
Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition, 7*(6), 705–718.
Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490–495.
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), Article 694.
Gibson, B. S., & Sztybel, P. (2015). What laboratory studies of symbolic spatial cues reveal about the control of attention in everyday life. In J. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention*. Cambridge, MA: MIT Press.
Gregory, S. E., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(2), Article 237.
Hamilton, A. F. D. C. (2016). Gazing at me: The importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*(1686), Article 20150080.
Hemmerich, K., Narganes-Pineda, C., Marotta, A., Martín-Arévalo, E., Jiménez, L., & Lupiáñez, J. (2022). Gaze elicits social and nonsocial attentional orienting: An interplay of shared and unique conflict processing mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 48*(8), Article 824.
Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9*(2), 131–143.
Hietanen, J. K., Myllyneva, A., Helminen, T. M., & Lyyra, P. (2016). The effects of genuine eye contact on visuospatial and selective attention. *Journal of Experimental Psychology: General, 145*(9), Article 1102.
Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *NeuroImage, 33*(1), 406–413.
Ishikawa, K., Dalmaso, M., Tanaka, Y., Oyama, T., & Okubo, M. (2025a). Ethnicity does not matter: Comparable reversed congruency effects for gaze stimuli from same-and other-ethnicity faces. *Psychological Research, 89*(6), 1–11.
Ishikawa, K., Oyama, T., & Okubo, M. (2021). The malfunction of domain-specific attentional process in social anxiety: Attentional process of social and non-social stimuli. *Cognition and Emotion, 35*(6), 1163–1174.
Ishikawa, K., Oyama, T., Tanaka, Y., & Okubo, M. (2025b). Perceiving social gaze produces the reversed congruency effect. *Quarterly Journal of Experimental Psychology, 78*(1), 9–21.
Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1–2), 1–19.

Jones, S. (2015). The mediating effects of facial expression on spatial interference between gaze direction and gaze location. *The Journal of General Psychology, 142*(2), 106–117.
Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics, 71*(2), 314–327.
Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269.
Lorenzoni, A., Calignano, G., Dalmaso, M., & Navarrete, E. (2023). Linguistic identity as a modulator of gaze cueing of attention. *Scientific Reports, 13*(1), Article 10829.
Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*(2), 174–207.
Makris, S., Hadar, A. A., & Yarrow, K. (2013). Are object affordances fully automatic? A case of covert attention. *Behavioral Neuroscience, 127*(5), Article 797.
Mareschal, I., Calder, A. J., & Clifford, C. W. (2013). Humans have an expectation that gaze is directed toward them. *Current Biology, 23*(8), 717–721.
Marotta, A., Aranda-Martín, B., De Cono, M., Ballesteros-Duperón, M.Á., Casagrande, M., & Lupiáñez, J. (2022). Integration of facial expression and gaze direction in individuals with a high level of autistic traits. *International Journal of Environmental Research and Public Health, 19*(5), Article 2798.
Marotta, A., & Chacón-Candia, J. (2026). When gaze conflicts with space: Implicit eye contact and the reversed congruency effect. *Psychonomic Bulletin & Review, 33*(4), Article 122.
Marotta, A., Lupiáñez, J., Martella, D., & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 326.
Marotta, A., Lupiáñez, J., Román-Caballero, R., Narganes-Pineda, C., & Martín-Arévalo, E. (2019). Are eyes special? Electrophysiological and behavioural evidence for a dissociation between eye-gaze and arrows attentional mechanisms. *Neuropsychologia, 129*, 146–152.
Marotta, A., Román-Caballero, R., & Lupiáñez, J. (2018). Arrows don't look at you: Qualitatively different attentional mechanisms triggered by gaze and arrows. *Psychonomic Bulletin & Review, 25*(6), 2254–2259.
Moore, C., & Dunham, P. J. (Eds.). (1995). *Joint attention: Its origins and role in development*. Hove, UK: Lawrence Erlbaum.
Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience, 47*(6), 497–514.
Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science, 16*(5), 269–274.
Oyama, T., & Okubo, M. (2022). Face and eyes: Critical analysis of the cue-target spatial-conflict account for gaze cueing effect. *Japanese Journal of Psychology, 92*(6), 511–521.
Pavan, G., Dalmaso, M., Galfano, G., & Castelli, L. (2011). Racial group membership is associated to gaze-mediated orienting in Italy. *PLoS One, 6*(10), Article e25608.
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.
Pereira, E. J., Birmingham, E., & Ristic, J. (2022). Infrequent faces bias social attention differently in manual and oculomotor measures. *Attention, Perception, & Psychophysics, 84*(3), 829–842.
Pfeiffer-Leßmann, N., Pfeiffer, T., & Wachsmuth, I. (2012). An operational model of joint attention–timing of the initiate-act in interactions with a virtual human. In *Proceedings of KogWis 2012* (p. 96).
Ponce, R., Lupiáñez, J., González-García, C., Casagrande, M., & Marotta, A. (2025a). Exploring the spatial interference effects elicited by social and non-social targets: A conditional accuracy function approach. *British Journal of Psychology, 116*(1), 69–88.
Ponce, R., Lupiáñez, J., González-García, C., Casagrande, M., & Marotta, A. (2025b). Investigating the gaze-driven reversed congruency effect in the spatial Stroop task: A distributional approach. *British Journal of Psychology, 116*(4), 1107–1127.
Ponce, R., Marotta, A., Román-Caballero, R., Tanaka, Y., Akçeşme, Ş., Chacón-Candia, J. A., & Lupiáñez, J. (2026). When arrows behave like eyes: Reversal of spatial Stroop interference by visual masking. *Cognition, 275*, Article 106623.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.
Proctor, R. W., Miles, J. D., & Baroni, G. (2011). Reaction time distribution analysis of spatial correspondence effects. *Psychonomic Bulletin & Review, 18*(2), 242–266.
Ricciardelli, P., Baylis, G., & Driver, J. (2000). The positive and negative of human expertise in gaze perception. *Cognition, 77*(1), B1–B14.
Ridderinkhof, K. R., Van den Wildenberg, W. P., Wijnen, J., & Burle, B. (2004). Response inhibition in conflict tasks is revealed in delta plots. In M. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 369–377). New York, NY: Guilford Press.
Ridderinkhof, R. K. (2002a). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz, & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 494–519). Oxford University Press.
Ridderinkhof, R. K. (2002b). Micro-and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research, 66*(4), 312–323.
Román-Caballero, R., Marotta, A., & Lupiáñez, J. (2021a). Target–background segregation in a spatial interference paradigm reveals shared and specific attentional mechanisms triggered by gaze and arrows. *Journal of Experimental Psychology: Human Perception and Performance, 47*(11), Article 1561.
Román-Caballero, R., Marotta, A., & Lupiáñez, J. (2021b). Spatial interference triggered by gaze and arrows. The role of target background on spatial interference. *Psicológica Journal, 42*(2), 192–209.
Rubo, M., & Gamer, M. (2021). Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. *British Journal of Psychology, 112*(1), 301–314.
Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, Article 104038.
Shultz, S., Klin, A., & Jones, W. (2018). Neonatal transitions in social behavior and their implications for autism. *Trends in Cognitive Sciences, 22*(5), 452–469.
Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods, 48*, 1197–1204.
Tanaka, Y., Ishikawa, K., Oyama, T., & Okubo, M. (2023). Face inversion does not affect the reversed congruency effect of gaze. *Psychonomic Bulletin & Review, 30*(3), 974–982.
Tanaka, Y., Ishikawa, K., Oyama, T., & Okubo, M. (2024). Eye gaze is not unique: The reversed congruency effect on gaze and tongue targets. *Quarterly Journal of Experimental Psychology, 77*(7), 1485–1497.
Tanaka, Y., Oyama, T., Ishikawa, K., & Okubo, M. (2025). Temporal dynamics of activation and suppression in a spatial Stroop task: A distribution analysis on gaze and arrow targets. *Attention, Perception, & Psychophysics, 87*(2), 367–383.
Tanaka, Y., Oyama, T., Ishikawa, K., & Okubo, M. (2026). Trial-by-trial modulations in a spatial Stroop task: A distribution analysis on social and non-social targets. *Quarterly Journal of Experimental Psychology, 79*(1).
Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics, 72*(3), 695–705.
Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37*(4), 571–590.
Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. *Memory & Cognition, 16*, 64–70.
Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review, 9*(2), 314–318.
Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Erlbaum.
Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*(1), 121–125.
Torres-Marín, J., Carretero-Dios, H., Acosta, A., & Lupiáñez, J. (2017). Eye contact and fear of being laughed at in a gaze discrimination task. *Frontiers in Psychology, 8*, Article 1954.
Vlamings, P. H., Stauder, J. E., van Son, I. A., & Mottron, A. L. (2005). Atypical visual orienting to gaze-and arrow-cues in adults with high functioning autism. *Journal of Autism and Developmental Disorders, 35*(3), 267–277.
Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196.
Zhang, X., Dalmaso, M., Castelli, L., Fiorese, A., Lan, Y., Sun, B., & Galfano, G. (2021). Social attention across borders: A cross-cultural investigation of gaze cueing elicited by same-and other-ethnicity faces. *British Journal of Psychology, 112*(3), 741–762.
Zhang, X., Dalmaso, M., Galfano, G., & Castelli, L. (2023). Tuning social modulations of gaze cueing via contextual factors. *Psychonomic Bulletin & Review, 30*(3), 1004–1010.