# The joint attention grouping effect: Perceptual binding of observed social interactions.

| | |
|---|---|
| Journal: | *Quarterly Journal of Experimental Psychology* |
| Manuscript ID | QJE-STD-24-290.R2 |
| Manuscript Type: | Standard Article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | McDonough, Katrina; University of East Anglia, Psychology<br>Edwards, S. Gareth; University of East Anglia,<br>Ewing, Louise; University of East Anglia,<br>Bayliss, Andrew; University of East Anglia, Psychology |
| Keywords: | joint attention, gaze perception, social cognition, perceptual grouping |
| | |

# The joint attention grouping effect: Perceptual binding of observed social interactions.

Katrina L. McDonough, S. Gareth Edwards, Louise Ewing & Andrew P. Bayliss
School of Psychology, University of East Anglia, United Kingdom



Keywords: joint attention, gaze perception, social cognition, perceptual grouping
Word Count (main text): 8625
Correspondence regarding this manuscript can be addressed to Andrew Bayliss (Andrew.P.Bayliss@uea.ac.uk)

**Abstract**

The visual system may perceptually process conspecifics more efficiently when they are interacting, versus not, to support social cognitive functions such as group detection. In three experiments, young adult university students were briefly shown dyads (upright or inverted) and made speeded judgments of whether they attended the same location (joint attention) or different locations (non-joint attention). Participants performed worse with inverted stimuli, but this inversion effect was smaller in joint attention conditions. These findings indicate perceptual grouping of joint attention dyads into a single perceptual unit. This *joint attention grouping effect* was evident when dyads looked towards spatial locations (Experiment 1), towards objects (Experiment 2), and for asymmetrically composed stimuli (Experiment 3). The effect was weaker for non-social directional stimuli (Experiment 1). These data support the idea that two interacting individuals are coded as one socially bound perceptual unit, supporting efficient and rapid social cognitive computations.

**The joint attention grouping effect: Perceptual binding of observed social interactions.**

Joint attention describes a social situation whereby two (or more) individuals synchronise their attention towards a common object or spatial location. That is, looking together towards a common focus. Joint attention is typically established by following another person's gaze or pointing gestures, to attend to something together. This shared sociocognitive alignment carries rich social and mentalistic significance and communicative potential (Emery, 2000; Stephenson et al., 2021). It helps individuals to understand each other's intentions, emotions, and interests, which is fundamental for building relationships and participating in social activities. Actively engaging in joint attention is crucial for language acquisition, observational learning, complex mental state inferences, and predicting future behaviour (Mundy et al., 1998). Moreover, those observing other people engaged in joint attention can spontaneously appreciate the shared mental state of others conveyed by joint attention (Calder et al., 2002) and identify opportunities to join their interaction by sharing the same focus of attention, creating a sense of mutual understanding and laying the groundwork for further social engagement.

The advantages conferred by a system that rapidly perceives the nature of an observed social interaction is clear. However, the visual appearance of an observed joint attention episode is relatively complex and variable, which presents a processing challenge. For example, it is relatively straightforward to decipher an observed instance of *mutual gaze*, where two conspecifics are looking at each other due to the perceptual symmetry and the gaze cues that confer mutual attentional facilitation to the two component stimuli (Emery, 2000; Vestner et al., 2019). However, even the simplest configuration of a joint attention episode involves observing two people looking in different directions at a third object or spatial location (e.g. Bayliss & Tipper, 2005; Frischen et al., 2007). Thus, there is great potential for breaks in perceptual symmetry, and the conceptual link between the agents is now mediated by their common locus of regard (i.e. unlike in mutual gaze, it is an indirect relationship). Therefore, a mechanism to facilitate processing of such complex social information into meaningful representations would be highly beneficial for social perception and cognition. This study aimed to investigate whether there is a processing advantage for displays depicting two individuals in a state of joint attention over non-joint attention. Such a demonstration would be evidence for a mechanism that binds social elements into one perceptual unit to facilitate processing.

Utilising the regularities of perceptual inputs can confer significant processing advantages to facilitate efficient interpretation of our complex environment (see Kaiser et al., 2019, for review). Some examples of stimulus inter-relatedness is described classically by Gestalt grouping principles, which have been instrumental for understanding how the perceptual system creates meaningful representations of complex sensory inputs. Applied to non-social stimuli, these grouping principles describe the ways in which individual visual elements are organised or perceptually bound into meaningful wholes (e.g. Coren & Girgus, 1980). The main Gestalt grouping principles include proximity, whereby objects close together in spatial location are perceptually grouped; similarity, whereby grouping is based on the similarity of object features; good continuation, whereby objects are grouped when features are aligned; and experience; whereby perceptual grouping is influenced by how objects have been grouped previously; amongst others. Individual people themselves operate in groups, particularly when engaged in joint attention, and their higher-level social connectedness during such interactions may facilitate Gestalt-like grouping for observers based on their similarity and proximity in space, the alignment of their attention, and the prevalence of such an arrangement in social situations. This Gestalt-like perceptual grouping would therefore provide a mechanism whereby the complexity of joint attention scenes could be simplified by the perceptual binding of the individual participants and their locus of attention into a single cohesive perceptual unit. Such a processing advantage could therefore allow additional processing capacity for further social analysis.

Recent studies have shown processing advantages for more simple social scenes displaying dyadic interactions in the form of mutual gaze (i.e. two people looking at each other) compared with dyads looking away from each other. Using a stimulus categorisation task, Papeo et al. (2017) found that pairs of bodies facing each other were categorised more accurately than non-facing bodies, an effect which was disrupted when the display was inverted. Similarly, Vestner et al. (2019) found that facing dyads were found faster in a visual search task compared with non-facing dyads, which was also disrupted by inversion. Vestner et al. (2019) found that facing dyads were also perceived as closer together and remembered more accurately than non-facing pairs. This effect was termed ‘social binding’ invoking the notion that interacting groups can be perceptually bound together to form a single perceptual unit. Furthermore, these effects were observed when only the head (Strachan et al., 2019; Vestner, et al., 2021) or only the body

was presented (Papeo & Abassi, 2019; Vestner, et al., 2021) and could emerge at a very early stage of processing (Fu et al., 2024). Further work has revealed selectively stronger neural representations for facing pairs in the visual cortex compared with non-facing pairs (Abassi & Papeo, 2020; Mersad & Caristan, 2021).

The current study aimed to investigate whether complex social attention episodes are processed more efficiently by the perceptual system by applying the concept of social binding to displays of joint attention. Specifically, we aimed to test whether this concept of social binding extends from simple displays of mutual gaze (e.g. Papeo et al., 2017; Vestner et al., 2019, 2022b), to more visually and socially complex displays of joint attention. Finding processing advantages for dyads displaying joint attention, compared with dyads not jointly attending, would therefore suggest that individuals engaged in complex social attention episodes are perceptually grouped based on their higher-level social connection.

To test this, we conducted three experiments in which participants made speeded judgements about whether a briefly presented dyad displayed joint-attention or non-joint attention, with presentations appearing either upright or inverted. In Experiment 1 only, stimuli consisted of either human dyads or a non-social dyad that can similarly convey direction: desk lamps (manipulated between groups), where joint attention was conveyed by their orientation towards a common location, compared with different locations for non-joint attention displays. Each subsequent experiment then increased the visual and social complexity of the scene. In Experiment 2, objects were added to the scene at the attended locations so that joint attention was now conveyed by the aligning of attention towards a common object, compared with dyads attending different objects for non-joint attention. The addition of objects not only establishes a more direct line of sight between the individuals and the objects (Lobmaier et al., 2006) but also distinguishes the joint attention display from mere gaze following (Emery, 2000), facilitating attention orienting (Bayliss & Tipper, 2005). Finally, Experiment 3 replicated Experiment 2 but with an asymmetric display, whereby individuals in the dyad were placed at different distances from the objects with different visual angles from the objects. Such an arrangement is not only more realistic in real world social scenes but could help to rule out any lower-level effects based on symmetry of the display.

We initially predicted that, across all three experiments, (1) dyads displaying joint attention would be identified faster and more accurately than dyads displaying non-joint

attention, (2) dyads displayed upright would be identified faster and more accurately than inverted dyads, (3) inversion effects will be stronger for dyads displaying joint-attention compared with non-joint attention, and for experiment 1 only, we predicted (4) that these effects would only occur for social stimuli (i.e. human dyads) and would not occur for non-social stimuli (i.e. desk lamps). Our initial hypotheses, particularly those relating to the direction of inversion effects, followed that of previous work that investigated the concept of social binding in simpler social scenes that display mutual gaze configurations (Papeo et al., 2017; 2019; Vestner et al., 2019). These prior studies found stronger inversion effects for dyads displaying mutual gaze, compared with dyads displaying non-mutual gaze, which the authors argue to be driven by a disruption to configural and social processing when mutual gaze displays are inverted. Although our research question examined a different social attention class (joint attention, not mutual gaze), used a different task (social attention judgement vs. visual search), and stimuli (heads vs. whole bodies), we nevertheless initially expected similar effects in our current studies as those noted by previous authors and so pre-registered these hypotheses for experiment 1.

Across all three experiments, and in line with our initial hypotheses, we found processing advantages for displays of joint attention compared with non-joint attention, as well as advantages for upright compared with inverted configurations. Crucially, however, although we found inversion effect differences between joint and non-joint attention conditions across all three experiments, the direction of this effect was reversed compared with prior research. Our findings revealed weaker inversion effects for displays of joint attention compared with displays of non-joint attention, indicating that detection of joint attention episodes is less impaired by inversion compared with episodes of non-joint attention. Although counter to our original naïve hypotheses, we will later propose a mechanism by which our results may demonstrate an advantage for the processing of jointly attending dyads in a subtly – but importantly – different way to a direct mapping from how mutual gaze is processed to how joint attention may be processed. We will argue that our findings suggest that dyads engaged in joint attention are socially bound into one perceptual unit which we describe as the *joint attention grouping effect*. We discuss the initial instance of this finding in the Interim Discussion following Experiment 1, after which we update our hypotheses for the following experiments to reflect these findings.

## Experiment 1

In this pre-registered experiment (https://aspredicted.org/blind.php?x=dm4rz8), we investigated how observed joint attention episodes are processed compared with non-joint attention episodes, when displayed in their canonical, upright orientation and when inverted. To examine if observed effects are selective for social stimuli, another group of participants completed the task with a directional but non-social stimulus category: desk lamps (Vestner et al. 2020; 2022a; 2022b). We hypothesised that (1) dyads displaying joint attention would be identified faster and more accurately than dyads displaying non-joint attention and (2) dyads displayed upright would be identified faster and more accurately than inverted dyads. Crucially, and following prior work, we also initially hypothesised that (3) inversion effects will be stronger for dyads displaying joint-attention compared with non-joint attention. Finally, we predicted (4) that these effects would be selective for social stimuli.

### Method

#### Participants

Sixty-four participants were recruited from the local psychology participant pool for course credit. Twenty-nine were assigned to the faces condition ($M_{age}$=20 years, SD=1.2, 27 women) and 30 were assigned to the lamps condition ($M_{age}$=20 years, SD=1.1, 24 women). Five participants were not included in the analysis; four were excluded in line with our pre-registered criteria (see Results) and one other due to technical failure. A sensitivity power analysis conducted with G*Power (v3.1) revealed that a sample size of 59 (α=.05) provides .90 power (1 - β) to detect effects in either direction with Cohen's *dz*=.43. Prior studies investigating similar effects with whole body dyads facing each other (Papeo et al., 2017; Vestner et al., 2019) and pilot studies from our laboratory yielded effect sizes that were consistently larger (*dz*=.50-1.10). For this and the other 2 experiments reported here, all participants reported normal or corrected-to-normal vision, and the studies were approved by the local Ethics Committee.

#### Stimuli & Apparatus

Face stimuli depicted the head and shoulders of two different female actors, from a total of eight possible identities, displaying a neutral expression and wearing a black t-shirt against a uniform grey background (see Figure 1). One face was presented on the left of the screen and one on the right, at 25% and 75% along the horizontal axis, respectively, and centred vertically.

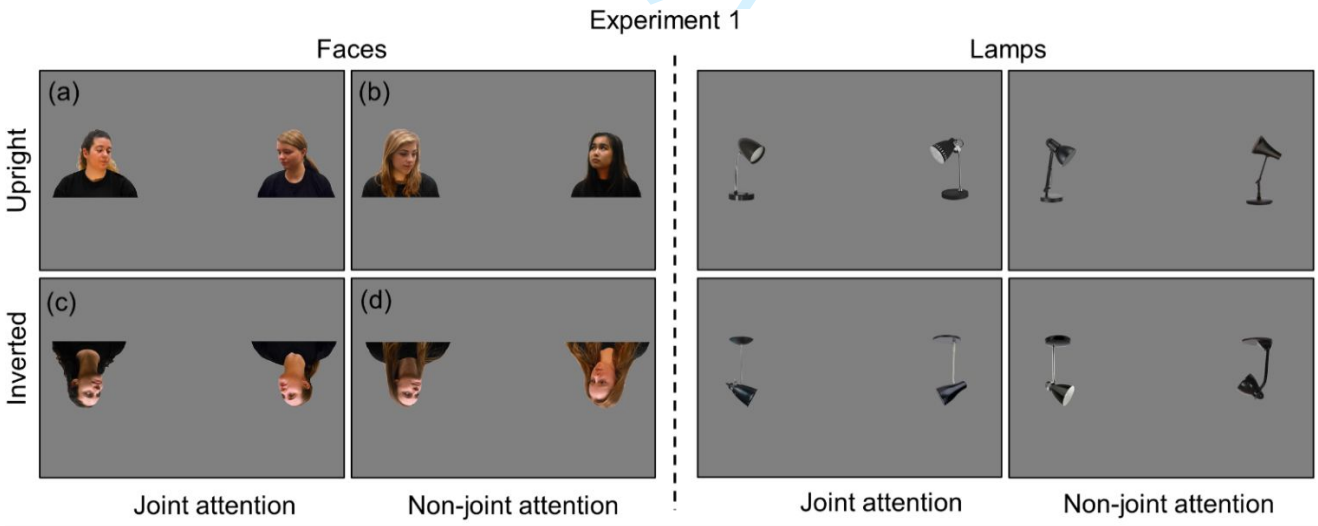
Faces were positioned with the torsos facing forwards and the heads positioned in one of

eight possible arrangements across four conditions. In the upright joint attention condition, both faces were looking towards the top centre of the screen or both faces were looking towards the bottom centre of the screen. In the upright non-joint attention condition, the left face looked towards the top centre of the screen and the right face looked towards the bottom centre of the screen, or vice-versa. These arrangements were mirror-inverted along the horizontal axis to create the stimuli for the inverted joint attention and inverted non-joint attention conditions, respectively. The face stimuli therefore had eight possible arrangements across four conditions.

Lamp stimuli depicted the head, stem, and base of two different desk-style lamps, from a total of eight non-identical lamps. The lamps were black against a uniform grey background and their size and position was consistent with the face stimuli. The lamp stimuli had eight possible arrangements across four conditions, closely matching the face stimuli. This was achieved by altering the angle of the head of the lamp to match the gaze direction of the faces. E-prime 2.0 was used to present the experiment via a HP EliteDisplay (1920x1080) monitor. A standard computer keyboard was used to record participant responses. Adobe Photoshop was used to edit the stimuli.

**Figure 1**

*Example stimuli for each condition in Experiment 1.*




*Note.* Stimuli depicted either faces or lamps (between-subjects). The stimulus conditions for the face stimuli are depicted in the left panel. Each dyad depicted either joint attention towards the same location (a, c) or non-joint attention (b, d), and were presented either upright (a,b) or inverted (c, d). Corresponding conditions were created for the lamp stimuli (right panel).

## Design

The experiment had a mixed design, with dyad gaze (joint vs non-joint attention) and dyad orientation (upright vs inverted) as within-subjects conditions, and stimulus type (faces vs lamps) as the between-subjects condition. The dependent variables were mean accuracy and mean reaction time (RT) of participant responses to whether dyad gaze was directed towards the same space (joint attention) or different spaces (non-joint attention).

## Procedure

Participants were tested in a laboratory in groups of up to five at a time and each group was alternately assigned to the faces or lamps condition. Each participant completed eight practice trials then a total of 240 experimental trials, presented in two blocks of 120 randomised trials (each representing all eight stimulus arrangements 15 times). There was a break between blocks. Each trial began with a black central fixation cross on a grey background, displayed for 500ms. Participants were instructed to fixate their gaze on the central fixation cross. This was immediately replaced by the stimulus display for 500ms, followed by a blank grey screen until response. Participants were instructed to indicate whether the faces (or lamps) were looking (or pointing) towards the same space - either both towards the top or both towards the bottom, or towards different spaces - one towards the top and one towards the bottom. They were informed that the stimulus display would appear either upright or inverted. Participants were asked to use the computer keyboard and press the "Z" key with their left hand or the "M" key with their right hand to indicate "same" or "different" (key assignment counterbalanced across participants), as quickly and accurately as possible. The response was followed by a blank grey screen for 2000ms before the next trial began. In practice trials only, participants received onscreen feedback after their responses.

# Results

## Data processing

Data were processed in line with pre-registered criteria (available here: https://aspredicted.org/blind.php?x=dm4rz8) and are available on OSF (McDonough et al., 2023). Individual participants were excluded if they failed to follow instruction (no participants), if their mean accuracy was less than 50% overall (no participants) or in any one condition (two participants excluded), or if their mean accuracy was 3*SD* below the sample mean (sample mean=87%, SD=10%, two participants excluded). Individual trials were removed if RTs were

more than 3*SD*s above or below the individual participant's mean RT (1.27% trials removed). Individual participants were excluded if their mean RT was more than 3SD above or below the sample mean (789ms, *SD*=183ms), no participants excluded. All incorrect trials were removed before RT analysis.

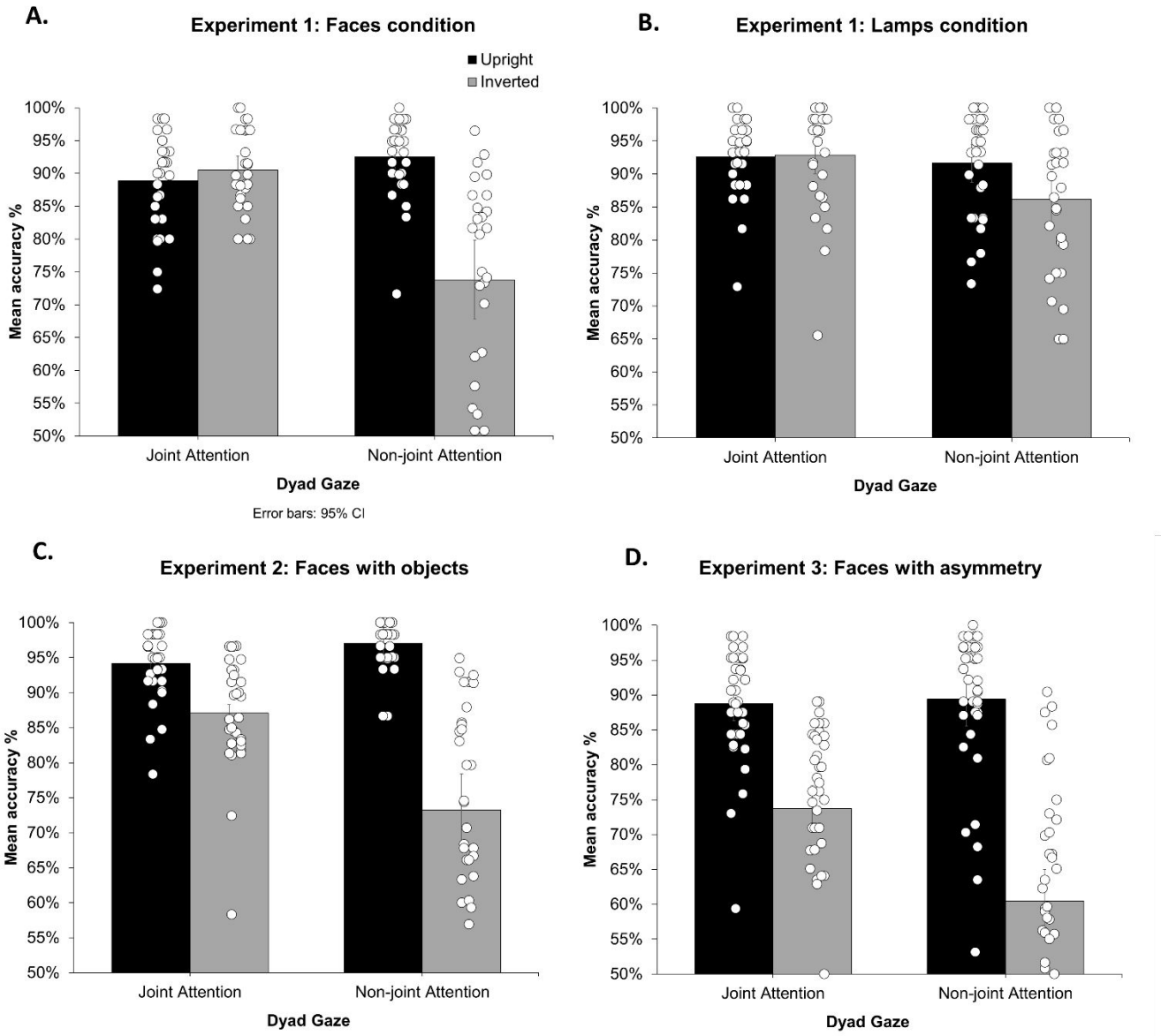
**Data analysis**

***Accuracy.*** Accuracy scores are reported in Figure 2. Percentage accuracy for each participant in each condition were entered into a 2x2x2 mixed ANOVA with dyad gaze (joint vs non-joint attention) and dyad orientation (upright vs inverted) as within-subjects factors and stimulus type (faces vs lamps) as a between-subjects factor. As expected, the analysis revealed a significant main effect of dyad gaze $F(1,57)=29.0$, $p<.001$, $\eta_p^2=.337$, 90%CI [0.17,0.47], with higher accuracy for joint-attention (M = 91%, SD = 6.9%) than non-joint attention (M = 86%, SD = 13%). There was a significant main effect of dyad orientation, $F(1,57)=67.9$, $p<.001$, $\eta_p^2=.544$, 90%CI [0.39,0.64], with higher accuracy for upright (M = 92%, SD = 6.9%) than inverted (M = 86%, SD = 13%) stimuli. There was a significant main effect of stimulus type, $F(1,57)=5.94$, $p=.018$, $\eta_p^2=.042$, with greater accuracy for lamps (M = 91%, SD = 8.6%) compared with faces (M = 86%, SD = 12%). Importantly, there was a significant interaction between dyad gaze and dyad orientation, $F(1,57)=42.4$, $p<.001$, $\eta_p^2=.426$, 90%CI [0.26,0.55], and planned comparisons revealed~~with~~ lower accuracy for inversion for non-joint attention (M(difference) = 12%, SD(-difference) = 12.8%), $t(57)=8.41$, $p<.001$, $d_z=1.1$, but not joint attention (M(difference) =1%, SD(-difference) = 7.0%), $t(57)=.979$, $p=.332$, $d_z=.13$ (see Figure 2). Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025) revealed lower accuracy for non-joint attention (M = 80%, SD = 15%) compared to joint attention conditions (M = 92%, SD = 7.0%) when inverted, $t(57)=6.84$, $p<.001$, $d_z=.98$, but no differences in upright conditions, $t(57)=-1.37$, $p=178$, $d_z=.18$ Furthermore, there was a significant interaction between dyad orientation and stimulus type, $F(1,57)=18.9$, $p<.001$, $\eta_p^2=.249$, 90%CI [0.10,0.39], with larger inversion effects for faces (M(difference) = 9%, SD(difference = 15%) than for lamps (M(difference) = 3%, SD(-difference) = 8.5%), $t(57)=4.35$, $p<.001$, $d=1.13$. The analysis further revealed a significant three-way interaction between dyad gaze, dyad orientation and stimulus type, $F(1,57)=13.7$, $p<.001$, $\eta_p^2=.194$, 90%CI [0.06,0.33], with a larger dyad gaze x dyad orientation interaction for participants in the faces condition than in the lamps condition. The interaction between dyad gaze and stimulus type was not significant,

$F(1,57)=1.92$, $p=.171$, $\eta_p^2=.033$, 90%CI [0,0.14].

**Figure 2**

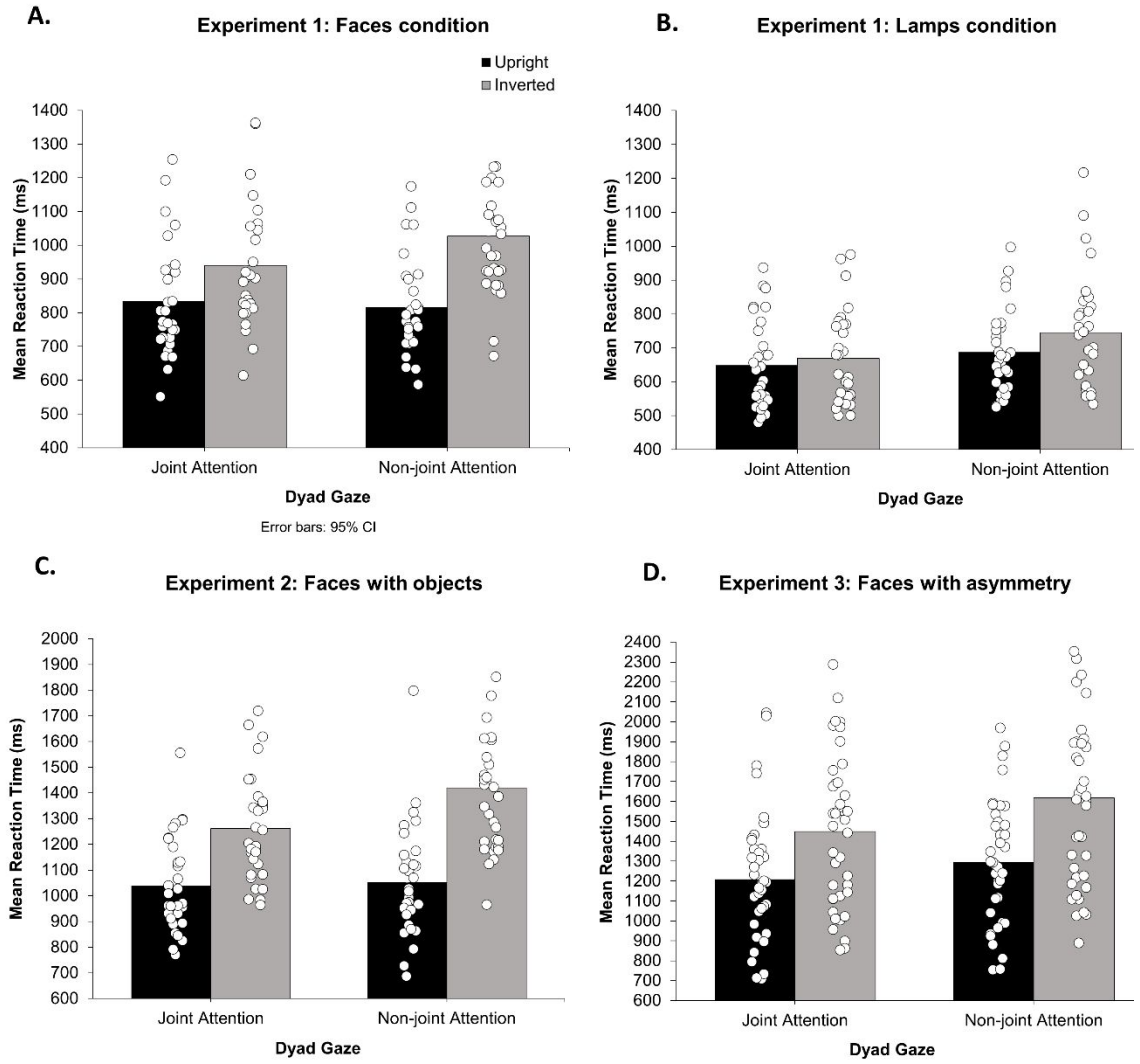
*Mean accuracy (%) for each Experiment.*




*Note.* Panel A: Mean accuracy (%) for the faces condition in Experiment 1. Panel B: Mean accuracy (%) for the lamps condition in Experiment 1. Panel C & D: Mean accuracy (%) for Experiment 2 and 3, respectively. Error bars are 95% confidence intervals.

**Figure 3**

*Mean reaction times for all experiments.*




*Note.* Panel A: Mean reaction times (ms) for the faces condition in Experiment 1. Panel B: Mean reaction times for the lamps condition in Experiment 1. Panel C & D: Mean reaction time for Experiment 2 and 3, respectively. Error bars are 95% confidence intervals.

***Reaction Time.*** RTs for each participant in each condition were entered into a 2x2x2 mixed ANOVA model. As expected, this revealed a significant main effect of dyad gaze, $F(1,57)=45.8$, $p<.001$, $\eta_p^2=.446$, 90%CI [0.28,0.56], with faster RTs for joint attention (M = 771ms, SD = 196ms) than non-joint attention (M = 817ms, SD = 202ms). There was a significant

main effect of dyad orientation, $F(1,57)=161.8$, $p<.001$, $\eta_p^2=.739$, 90%CI [0.63,0.80], with faster RTs for upright (M = 745ms, SD = 162ms) than inverted (M = 843ms, SD = 222ms) stimuli. There was a significant main effect of stimulus type, $F(1,57)=30.4$, $p<.001$, $\eta_p^2=.348$, with faster RTs for lamps (M = 688ms, SD = 143ms) compared with faces (M = 904ms, SD = 191ms). Importantly, there was a significant interaction between dyad gaze and dyad orientation, $F(1,57)=52.0$, $p<.001$, $\eta_p^2=.477$, 90%CI [0.31,0.59], and planned comparisons~~which~~ revealed a larger increase in RT with inversion (inversion effect) for non-joint attention (133ms inversion effect, SD = 108ms), $t(58)=13.5$, $p<.001$, $d_z=.70$, than joint attention (62ms inversion effect, SD = 78ms), $t(58)=7.46$, $p<.001$, $d_z=.35$. Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025), revealed faster RTs for joint-attention (M = 802ms, SD = 212ms) than non-joint attention conditions (M = 884ms, 226ms) when inverted, $t(57)=8.67$, $p<.001$, $d_z=.36$, but no differences in upright conditions, $t(57)=-1.39$, $p=.170$, $d_z=.06$. There was also a significant dyad orientation by stimulus type interaction, $F(1,57)=60.4$, $p<.001$, $\eta_p^2=.515$, 90%CI [0.36,0.62], with larger inversion effects for faces (M = 159ms, SD = 96ms) than for lamps (M = 38ms, SD = 61ms), $t(57)=7.77$, $p<.001$, $d=2.02$.The analysis further revealed a significant three-way interaction between dyad gaze, orientation and stimulus type, $F(1,57)=11.4$, $p=.001$, $\eta_p^2=.166$, 90%CI [0.04, 0.30], with a larger dyad gaze x dyad orientation interaction in the faces condition than in the lamps condition. The interaction between dyad gaze and stimulus type was not significant, $F(1,57)=2.84$, $p=.098$, $\eta_p^2=.047$, 90%CI [0, 0.16] (see Figure 3).

**Interim Discussion**

Experiment 1 revealed strikingly clear differences in the processing of joint compared with non-joint attention displays, with greater accuracy and faster reaction times overall for detecting episodes of joint attention compared with non-joint attention. Furthermore, participants were significantly impaired in accuracy and RTs when judging inverted compared with upright dyads ~~(a dual-face inversion effect)~~. Importantly, although we did indeed find inversion effect differences between joint and non-joint attention conditions for both accuracy and RT measures, these differences were driven by the inverted conditions, and more importantly, the direction of these effects were contrary to our initial pre-registered hypotheses: participants showed weaker inversion effects for displays of joint attention than non-joint attention.

We initially expected to find stronger inversion effects for joint attention displays compared with non-joint attention displays, following prior studies that used a similar general approach but with different stimuli classes and tasks to investigate the concept of social binding in simpler social scenes that display mutual gaze configurations (Papeo et al., 2017; 2019; Vestner et al., 2019). In light of our current contradictory findings, and after incorporating research published after Experiment 1 was conceived, we discuss below some potential reasons as to why our pattern of results may have emerged.

Previous studies investigating processing advantages for more simple social scenes displaying mutual gaze configurations (e.g. Papeo et al., 2017; 2019; Vestner et al., 2019) initially provided evidence that such scenes provide strong cues to higher-level social groupings and are processed more efficiently by the visual system compared with displays of non-mutual gaze, in the same way that low-level perceptual features are grouped. They ~~areguedargued~~ that inversion disrupts these social and configural processes, resulting in stronger inversion effects for mutual gaze compared with non-mutual gaze displays. However, recent findings have brought into question the high-level social nature of these processing advantages. It has been argued that mutual gaze displays may be afforded a processing advantage, not due to social grouping, but due to strong directional gaze cues towards task-relevant pairings. In mutual gaze displays, where pairs are facing each other, attention is cued back and forth from one face to the other, creating an "attentional hotspot" towards the centre of the task-relevant facing pair, whereas for non-facing pairs attention is cued away from each face and outwards of the social scene. This would allow an advantage for task-relevant mutual gaze displays based on spatial attention boosting the perceptual representation of mutual gaze displays (Flavell et al., 2022), rather than due to 'social' grouping. Indeed, Vestner et al. (2020; 2022a; 2022b) have since reported that their higher-level 'social binding' effects may in fact be driven by this domain general, attentional cueing effect, revealing similar grouping effects for non-social stimuli such as desk lamps and desk fans that also direct visual-spatial attention. This therefore suggests that inversion effects found in these prior studies may reflect disruptions to attentional cueing processes, rather than disruptions in social processing.

In the current study, episodes of joint attention are not only visually and socially complex but also afford a physical arrangement of interaction partners where the alignment of their gaze is never directed towards one another. Instead, gaze is directed away from one another (in all

conditions) and towards a common object or spatial location (in the joint attention condition). Therefore, if anything, attention is cued away from the task-relevant interacting pair and outwards towards the common object/location. We therefore argue that the findings of the current study are less likely to be influenced by the domain general distributions of spatial attention that Vestner and ~~colleauges~~colleagues now take as a strong potential underpinning of their effects. Clearly, attentional processes are engaged in our displays, but the specific pattern of attention distribution that is thought to lead to their results is absent. It is therefore clear that our research question, stimuli and task is not likely to produce the same pattern of results as those found in ~~mutal~~mutual gaze paradigms.

Thus far, we have been able to offer an ~~explaination~~explanation as to why our findings differ from previous work (and from our original hypothesis). But we also need to explain the pattern of results we did obtain. To approach this, it is also important to note that although our paradigm involves the inversion of a display containing faces, face inversion disrupts configural processing of the internal features of a face (Farah et al., 1995; Maurer et al., 2002) and while many studies report that inversion disrupts gaze perception (see Frischen et al., 2007, for review), other studies show preserved and sophisticated utilization of gaze information in ~~misaligned~~ displays of faces rotated 90 or 180 degrees ~~displays~~ in the picture place (Bayliss et al., 2004; Tipples, 2005; and see Frizchen et al., 2007). Furthermore, we highlight that our task did not examine face processing per se, but rather the ability of the perceptual system to register the relationship between the gaze direction of two separate faces. Therefore, we argue that participants were able to determine the direction of gaze in our inverted displays, albeit to a lesser extent than upright displays, as evident in the overall processing advantages found for upright compared with inverted displays. These findings are consistent with more recent research that revealed processing advantages for triadic interactions between three people that persisted in inverted displays (Colombatto et al., 2025).

Applying the above work to our own, we can propose that our findings may suggest~~indicate~~ an advantage for displays of individuals bound into a common joint attention episode, but via a different mechanism than originally conceived. Specifically, we suggest that dyads engaged in joint attention are perceptually grouped and processed as a single entity, resulting in more efficient processing compared with the processing of two separate entities in non-joint attention conditions. This is demonstrated by noting that when displays are inverted

processing of joint attention are ~~preserved~~ impacted less because they require only one operation to resolve inversion to process the overall perceptual unit (jointly gazing faces are bound as a single unit). In contrast, displays of non-joint attention lack this cohesive grouping, therefore requiring the resolution of two representations from inversion to process each individual separately, relatively impairing performance in inverted displays. This interpretation is consistent with our findings that there is no overall advantage to processing of upright jointly attending dyads, instead the advantage is revealed only when the system is disrupted by display inversion.

This new interpretation of our findings is consistent with how the visual system processes multi-object arrangements, especially when the perceptual system is challenged, where any processing advantages are indicative of "integrative processing" of multiple objects into a single representation, compared to the individual processing of multiple individual objects (for review, see Kaiser et al., 2019). Experiment 1 also ~~demonstrated~~ showed a much larger effect with social stimuli than with non-social directional stimuli, though gross performance ~~differnces~~differences between the groups limit the interpretability of these data (i.e. the lamps were simpler stimuli than the faces). Furthermore, it is also important to note that we did find significant inversion effects in both joint and non-joint conditions in our RT data, a hallmark of high-level processing (Kaiser et al., 2019) although we did not find a significant inversion effect of the joint attention condition in our accuracy data, likely due to ceiling effects.

~~Taken t~~Together, these findings therefore support the idea that the perceptual system efficiently processes complex social information by perceptually grouping individuals based on their high-level social connection such as engagement in joint attention, when compared with how the perceptual system tries to resolve the non-alignment of two faces looking in different directions to one another.

We therefore conclude that the findings from experiment 1 provide evidence for a *joint attention grouping effect*: that when the perceptual system is challenged, observed episodes of joint attention are processed more efficiently that non-joint attention episodes through social binding mechanisms ~~that persist even in inverted displays~~. To support this interpretation, we aimed to replicate these findings in Experiment 2.

## Experiment 2

Experiment 2 replicated and extended Experiment 1 (faces condition only) by inserting objects into attended locations, therefore adding perceptual complexity and increasing the social

relevance of the interaction. Objects are important for joint attention: distinguishing joint attention from mere gaze following (Emery, 2000); facilitating attention orienting of the interacting pair (Bayliss & Tipper, 2005), and establishing direct line of sight (Lobmaier et al., 2006). If dyads engaged in joint attention are perceptually grouped based on their social interaction status, then the additional perceptual complexity of the included objects should not disrupt the *joint attention grouping effect* that we found in experiment 1. We therefore hypothesised that we would replicate the results of Experiment 1.

## Method

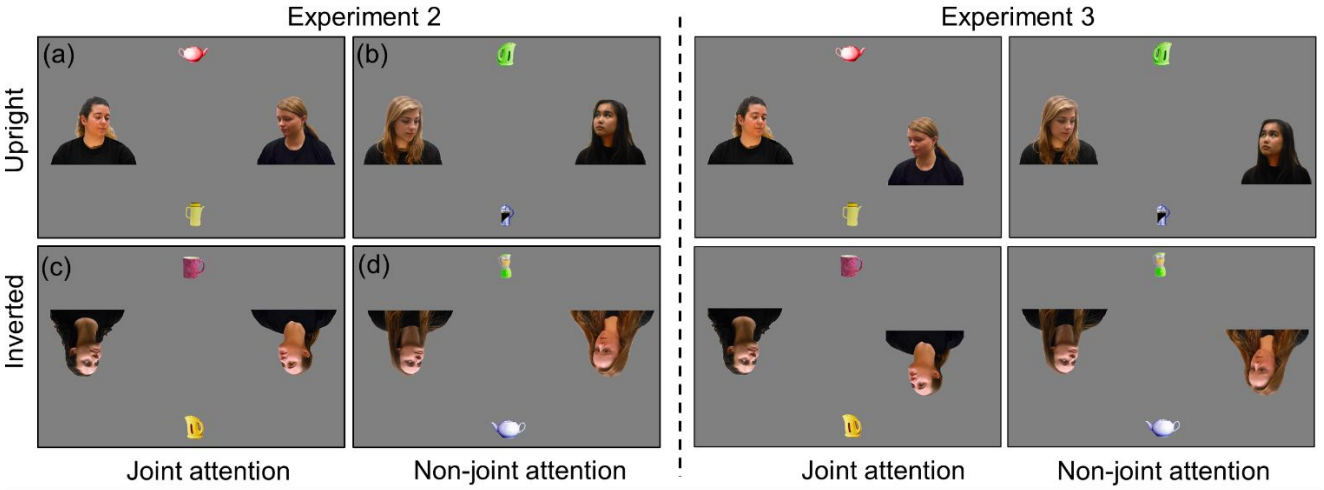
The methodological approach was the same as in Experiment 1 except where noted.

### Participants

Thirty-five adults participated, of whom 32 were included in the analysis ($M_{age}$=19 years, *SD*=1.3, 26 women). A sensitivity power analysis conducted with G*Power (Version 3.1) revealed that a sample size of 32 (α=.05) provides .90 power (1-β) to detect effects in the predicted direction with Cohen's $d_z$=.53 and effects in either direction with Cohen's *dz*=.59. In Experiment 1, effect sizes for the dyad gaze x dyad orientation interaction in the face condition were larger (accuracy $d_z$=1.12, RT $d_z$=1.32).

**Figure 4**

*Example stimuli for each condition in Experiment 2 and 3.*




*Note.* The stimulus conditions for Experiment 2 are depicted in the left panel. Each dyad was arranged symmetrically about the horizontal-axis and displayed either joint attention towards the same object (a, c) or non-joint attention (b, d), and were presented either upright (a, b) or inverted (c, d). The stimulus conditions for Experiment 3 are depicted in the right panel. Here,

dyads were arranged in an asymmetric display.

### Stimuli & Apparatus

Face stimuli were the same as in Experiment 1 and the stimuli also included household objects positioned at the top centre and bottom centre of the screen, in line with dyad gaze (see Figure 4). The objects were chosen on each trial randomly from a set of 48 images, representing six different kitchen objects (blender, teapot, mug, kettle, jug, cafetiere), in four different colours (red, yellow, blue, green), facing either to the left or to the right taken from Bayliss et al. (2006).

### Design & Procedure

The experiment had a within-subjects design with dyad gaze (joint vs non-joint attention) and dyad orientation (upright vs inverted) as conditions. Unlike Experiment 1, only face stimuli were presented. Participants were asked to judge whether the faces were looking towards the same *object* or different *objects*.

## Results

### Data processing

The same exclusion criteria from Experiment 1 were applied. Two participants were excluded due to <50% accuracy in at least one condition, and one participant was excluded with mean RT >3SDs above or below sample mean (M=1194ms, *SD*=226ms). Individual trials >3*SD*s above or below the individual participant's mean RT were removed (1.37%).

### Data analysis

***Accuracy.*** Accuracy scores are reported in Figure 2. Percent accuracy scores for each participant in each condition were entered into a 2x2 repeated-measures ANOVA with dyad gaze (joint vs non-joint attention) and dyad orientation (upright vs inverted) as within-subjects factors. The analysis revealed a significant main effect of dyad gaze $F(1,31)=11.4$, $p=.002$, $\eta_p^2=.269$, 90%CI [0.07,0.44], with higher accuracy for joint-gaze (M = 91%, SD = 7.5%) than non-joint attention (M = 85%, SD = 16%). There was a significant main effect of dyad orientation, $F(1,31)=128.7$, $p<.001$, $\eta_p^2=.806$, 90%CI [0.68,0.86], with higher accuracy for upright (M = 96%, SD = 4.6%) than inverted (M = 80%, SD = 14%) stimuli. Most importantly, there was once again a significant interaction between dyad gaze and dyad orientation, $F(1,31)=30.2$, $p<.001$, $\eta_p^2=.494$, 90% CI [0.27, 0.63], and planned comparisons revealed~~with~~ a larger inversion effect for non-joint attention (M(difference) = 24%, SD(difference) = 14%), $t(31)=9.70$, $p<.001$,

$d_z$=2.2, than joint attention (M(difference) = 7%, SD(-difference) = 8.6%, $t$(31)=4.67, $p$<.001, $d_z$=1.06. Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025), revealed lower accuracy for non-joint attention (M = 73%, SD = 15%) compared to joint attention conditions (M = 87%, SD = 7.9%) when inverted, $t$(31)=4.66, $p$<.001, $d_z$=1.2, and lower accuracy for joint-attention (M = 94%, SD = 5.1%) compared to non-joint attention (M = 97%, SD = 3.4%) when upright, $t$(31)=-2.79, $p$=.009, $d_z$=.66.

***Reaction Time.*** RTs for each participant in each condition were entered into a 2x2 repeated-measures ANOVA model and results mirrored those observed for accuracy (see Figure 3). There was a significant main effect of dyad gaze, $F$(1,31)=25.3, $p$<.001, $\eta_p^2$=.449, 90% CI [0.22, 0.59], with faster RTs for joint attention (M = 1150ms, SD = 221ms) than non-joint attention (M = 1236ms, SD = 305ms). There was a significant main effect of dyad orientation, $F$(1,31)=143.2, $p$<.001, $\eta_p^2$=.822, 90%CI [0.71, 0.87], with faster RTs for upright (M = 1045ms, SD = 197ms) than inverted stimuli (M = 1341ms, SD = 250ms). Again, critically, there was a significant interaction between dyad gaze and dyad orientation, $F$(1,31)=21.2, $p$<.001, $\eta_p^2$=.406, 90%CI [0.18,0.56], and planned comparisons revealed~~with~~ a larger inversion effect for non-joint attention (M = 368ms, SD = 198ms), $t$(31)=10.5, $p$<.001, $d_z$=1.5, than joint attention (M = 224ms, SD = 124ms), $t$(31)=10.2, $p$<.001, $d_z$=1.1. Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025), revealed faster RTs for joint-attention (M = 1262ms, SD = 202ms) than non-joint attention conditions (M = 1420ms, SD = 271ms) when inverted, $t$(31)=5.78, $p$<.001, $d_z$=.66, but no differences in upright conditions, $t$(31)=.80, $p$=.432, $d_z$=.07.

## Discussion

Experiment 2 replicated the effects from Experiment 1, this time with more visually and socially complex scenes. Specifically, Experiment 2 showed greater accuracy and faster reaction times for upright displays compared with inverted displays, and, importantly, weaker inversion effects for both accuracy and reaction time data for jointly attending dyads compared with non-jointly attending dyads. This provides further evidence~~confirms~~ that when the perceptual system is challenged, jointly attending dyads are processed more efficiently than non-jointly attending dyads, even~~the perceptual system efficiently processes jointly attending dyads~~ in more naturalistic displays with the added visual complexity of images of objects in the scene.

There was however one different finding in the additional post hoc tests compared with Experiment 1. Here, the additional post-hoc tests revealed a significant difference between upright conditions for the accuracy data in experiment 2, with higher accuracy for non-joint attention compared with joint-attention conditions. While this is in the opposite direction to what would be expected, we treat this finding with caution given the ceiling effects for accuracy in these conditions, and the likelihood that this difference is driven by one or three individual data points in the joint-attention upright condition (see Figure 2, panel C).

## Experiment 3

Although face dyads in Experiment 1 and 2 had different identities they nevertheless held significant perceptual symmetry, particularly in the joint attention display. This could have contributed to the efficient processing of inverted displays of joint attention. Here, in this pre-registered experiment (https://aspredicted.org/blind.php?x=vp2vx9), we tested directly whether our *joint attention grouping effect* replicates when displays are not symmetrical (see Vestner et al., 2019). Following the replication of finding the opposite pattern of data compared with our registered hypotheses in Experiment 2, here we employed a revised pre-registered hypothesis in the same direction as the data pattern in Experiments 1 and 2.

### Method

The main difference in methodological approach compared with Experiment 2 is that this experiment was conducted online instead of in the laboratory.

#### Participants

Fifty adults participated in exchange for course credit. Forty were included in the analysis ($M_{age}$=23 years, *SD*=7.8 years, 26 women). A sensitivity power analysis again confirmed that 40 participants (α=.05) provide .90 power (1-β) to detect effects in the predicted direction with Cohen's $d_z$=.47 and effects in either direction with Cohen's $d_z$=.53. In Experiment 2, effect sizes for the dyad gaze x dyad orientation interaction were larger (accuracy $d_z$=1.0, RT $d_z$=.81). All participants reported normal or corrected-to-normal vision.

#### Stimuli & Apparatus

Stimuli in Experiment 3 were comparable to those in Experiment 2 but presented in an asymmetric display (see Figure 4). This was achieved by repositioning only one of the faces (left face or right face) per trial, either 5% closer to or 5% further from the gazed-at object, along the line of sight, therefore creating an additional four stimuli arrangements. This method for

achieving asymmetry was chosen to ensure that (1) the line of sight from the face to the gazed object was maintained, (2) the distance between the face and the gazed object was counterbalanced across all conditions, and (3) the faces never shared the same position along the y-axis, therefore achieving asymmetry in all trials. This therefore created 32 possible arrangements (the original eight arrangements multiplied by the four additional arrangements. Inquisit (Millisecond) software was used to present the experiment online via the SONA Systems recruitment platform.

**Design & Procedure**

The design structure and procedure closely replicated Experiment 2. Participants completed a total of 256 experimental trials (each representing all 32 stimulus arrangements 8 times), across three blocks. To mitigate additional challenges of online data collection, participants were asked to provide feedback of their testing experience, reporting any reason to suggest that they did not complete the experiment appropriately (e.g. technical difficulties, disruptions during the task, etc.). Feedback responses formed part of the pre-registered exclusion criteria for online data collection.

## Results

**Data processing**

Data were processed in line with a pre-registered criteria (available here: https://aspredicted.org/blind.php?x=vp2vx9) that replicated Experiments 1 and 2, with the addition of participant feedback responses. Based on this feedback, one participant was removed for reporting periods of random responding and failure to pay attention. Four participants were excluded due to <50% accuracy overall, and five excluded for <50% accuracy in at least one condition. Individual trials >3SDs above or below the individual participant's mean RT were removed (1.46%).

**Data analysis**

***Accuracy.*** Accuracy scores are reported in Figure 2. There was a significant main effect of dyad gaze $F(1,39)=14.7$, $p<.001$, $\eta_p^2=.274$, 90%CI [0.09,0.43], with higher accuracy for joint gaze (M = 81%, SD = 13%) than non-joint attention (M = 75%, SD = 19%). There was a significant main effect of dyad orientation, $F(1,39)=329.6$, $p<.001$, $\eta_p^2=.894$, 90%CI [0.83,0.92], with higher accuracy for upright (M = 89%, SD = 9.3%) than inverted (M = 67%, SD = 15%) stimuli. Importantly, there was a significant interaction between dyad gaze and dyad orientation,

$F(1,39)=14.4$, $p<.001$, $\eta_p^2=.269$, 90% CI [0.09, 0.43], and planned comparisons revealed~~with~~ a larger inversion effect for non-joint attention (M(difference) = 29%, SD(-difference) = 15%, $t(39)=12.0$, $p<.001$, $d_z=2.3$, than joint attention (M(difference) = 15%, SD(-difference) = 12%, $t(39)=7.75$, $p<.001$, $d_z=1.4$. Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025), revealed lower accuracy for non-joint attention (M = 61%, SD = 15%) compared to joint attention conditions (M = 74%, SD = 13%) when inverted, $t(39)=4.14$, $p<.001$, $d_z=.97$, but no differences in upright conditions, $t(39)=-.479$, $p=.635$, $d_z=.07$.

***Reaction Time.*** There was a significant main effect of dyad gaze, $F(1,39)=25.2$, $p<.001$, $\eta_p^2=.393$, 90%CI [0.19,0.54], with faster RTs for joint attention (M = 1328ms, SD = 370ms) than non-joint attention (M = 1457ms, SD = 418ms), $t(39)=5.02$, $p<.001$, d=.79. There was a significant main effect of dyad orientation, $F(1,39)=58.6$, $p<.001$, $\eta_p^2=.600$, 90%CI [0.42,0.70], with faster RTs for upright (M = 1252ms, SD = 313ms) than inverted stimuli (M = 1533ms, SD = 427ms). Importantly, there was a significant interaction between dyad gaze and dyad orientation, $F(1,39)=4.34$, $p=.044$, $\eta_p^2=.100$, 90%CI [0.01,0.26], and planned comparisons revealed~~with~~ a larger inversion effect for non-joint attention (M = 322ms, SD = 274ms), $t(39)=7.43$, $p<.001$, $d_z=.83$, than joint attention (M = 240ms, SD = 252ms), $t(39)=6.01$, $p<.001$, $d_z=.68$. Two further post-hoc tests, Bonferroni corrected for multiple comparisons (adjusted alpha = .025), revealed faster RTs for joint-attention (M = 1448ms, SD = 381ms) than non-joint attention conditions (M = 1618ms, SD = 457ms) when inverted, $t(39)=4.43$, $p<.001$, $d_z=.40$, and also faster RTs for joint-attention (M = 1208ms, SD = 320ms) than non-joint attention conditions (M = 1295ms, SD = 303ms) when upright, $t(39)=3.49$, $p=.001$, $d_z=.28$.

## Discussion

Experiment 3 replicated our *joint attention grouping effect* using an asymmetrical display, again showing a weaker inversion effect for joint attention episodes in both accuracy and RT data. This suggests that low-level perceptual symmetry of the dyadic display in our prior experiments cannot fully explain this effect, supporting the notion that socially interacting individuals are instead grouped together based on their higher-level social interaction status. Furthermore, it is important to note here that, contrary to our previous experiments, the reaction time data for experiment 3 shows that the weaker inversion effect for joint attention episodes is not only driven by differences in the inverted conditions, but also now differences in the upright conditions, with faster reactions to upright joint attention displays compared with upright non-

joint attention displays. It is likely that since the stimuli in experiment 3 poses the greatest challenge to the perceptual system compared to experiment 1 and 2, performance is no longer at ceiling, allowing the joint attention grouping effect to be revealed in both upright and inverted displays.

## General discussion

Over three experiments, we found data consistent with the notion that pairs of jointly attending individuals are perceptually bound together as a single social unit. This conclusion is supported by our finding of a *joint attention grouping effect* due to a weaker inversion effect for jointly attending dyads compared with non-jointly attending dyads. These effects were found in both measures of accuracy as well as RT measures. These findings suggest that observed joint attention signals a social interaction between two individuals and engages 'social binding' processes that unify the perceptual processing of the observed individuals (e.g. Vestner et al., 2019). This binding, akin to gestalt grouping, affords a processing advantage to jointly attending dyads that is not present for pairs of faces that are looking in different directions and therefore not treated as a single unit (Kaiser et al., 2019). Therefore – in the latter non-joint attention case - when the display is inverted, the processing costs are applied two-fold since the system must overcome the processing disruption of two unrelated ~~imnverted~~inverted perceptual elements (two faces). However, when joint attention displays are inverted, the disruption of inversion is applied to the single perceptual unit, which can be more efficiently resolved than two unrelated inverted faces. This is consistent with our observation that ~~there is~~we mostly find no overall advantage to processing of upright jointly attending dyads per se, especially when accuracy is at ceiling; we show consistently that the advantage is revealed only when the system is disrupted by display inversion. When the perceptual system is challenged further (and accuracy is no longer at ceiling) in experiment 3, representing more naturalistic social scenes, here we reveal a 'social binding' effect for joint attention displays both when upright, and when inverted.

Secondly, we observed significantly weaker grouping effects for non-biological objects, specifically desk lamps that can be oriented by external force. This contrasts somewhat with findings by Vestner et al. (2022b), who showed that binding between dyads displaying *mutual* gaze extended to non-biological objects including desk lamps that direct visual attention in the same way as mutual gaze displays. Therefore, our much weaker binding effect for lamps suggests that our observed effect may be somewhat selective for social stimuli. However, we

acknowledge that this finding could be influenced by the overall relative processing ease of the lamps compared with faces, which tempers any strong claims regarding the domain specificity of our results. Finally, our findings suggest that the observed effects operate at a relatively high perceptual level. In Experiment 3 we showed that our *joint attention grouping effect* is preserved even when the symmetry of the display is disrupted. Furthermore, in all experiments, we do find inversion effects for both joint and non-joint attention conditions (albeit to different degrees), a hallmark of higher-level perceptual processing (Kaiser et al., 2019). The findings are therefore primarily driven by the social significance of joint attention, rather than lower-level stimulus features.

While our overall finding of a processing advantage for complex social scenes appears consistent with prior findings that showed similar advantages for displays of mutual gaze, the patterns of results that lead us to arrive at these conclusions appear to diverge from prior findings, and therefore our initial hypotheses. For example, Vestner et al. (2019) found a visual search advantage for mutually gazing bodies, while we only showed a performance advantage for observed joint attention displays when the faces were inverted, with relatively similar performance between joint and non-joint attentional displays for upright faces. Even more striking, in Papeo et al's (2017) limited exposure stimulus categorisation task, there was a stronger inversion effect for mutually gazing whole bodies, while we note a consistently weaker inversion effect for jointly attending faces. However, it is important to highlight some key differences between the experiments presented here and those from the prior literature which may have some explanatory value. Firstly, episodes of joint attention are more visually, and socially complex forms of social interaction compared with mutual gaze. This added complexity may explain why we did not find processing differences for joint attention displays when presented upright, but only when these displays were challenged by inversion. Secondly, it has since been shown that displays of mutual gaze (i.e. face-to-face pairs) also direct visual spatial attention inwards towards the task-relevant interacting pair, while displays of non-mutual gaze (i.e. back-to-back pairs) direct visual spatial attention outwards and away from the task-relevant pair (Vestner et al., 2022). Therefore, the processing advantages for mutually gazing pairs may be solely explained by this attentional cueing effect, which is disrupted by inversion. In contrast, displays of both joint and non-~~dis~~joint attention would, if anything, direct visual spatial attention away from the task-relevant face stimuli either towards a common locus of regard or in a more

diffuse manner in the case of non-~~dis~~joint attention. Thus, while the attention cueing account may explain the advantage for mutually gazing individuals, it does not easily explain our data.

An alternative attentional account is worthy of further consideration, however. While none of our conditions featured stimuli that directed attention towards sources of information relevant to the task (i.e. the faces never looked at each other), attentional processes certainly will be involved and deployed differently in the various conditions. For example, while not directly cueing one another, the faces in the joint attention displays could – by gazing at a common location – provide a stable representation via spatial attention shifts among the three elements (face 1, face 2, and common locus of regard) that is absent in non-joint conditions. In this way, attention orienting may indeed provide the 'structure' across the three separate units upon which the socially bound percept is built (see Stephenson et al., 2021).

Finally, we assert that direct comparison of the present studies with those investigating mutual gaze is not straightforward because we used a different task, with very different stimuli. More broadly, both our studies and those investigating mutual gaze are limited by the fact that here joint attention displays are only compared with non-joint attention, and the studies of those examining mutual gaze quite naturally and properly compare mutual vs facing away displays, making a straightforward comparison challenging at best. Our aim was to examine how inversion affects the discrimination of joint vs. non joint attention displays. We interpret our present results as suggesting that in joint attention episodes, the two individuals are perceptually unified into a single processing element which therefore renders the effects of inversion less disruptive since only one representation needs to be resolved. For non-joint attention displays, the individuals are instead attending to their own separate loci of regard, creating two sets of person-object perceptual units with spatially relevant arrangements, leading to two representations requiring resolution when inverted, which - relative to the joint attention displays - impairs performance. Indeed, similar results have now been shown in studies investigating triadic social interactions between three people, revealing processing advantages for socially interacting triads compared with non-interacting triads that persisted in inverted displays (Colombatto, et al., 2025). Therefore, while it is indeed important to note some differences in stimuli, tasks, and data patterns in our study compared with those examining mutual gaze processing, and rightly caution against drawing direct comparison, there are some conclusions that we can draw confidently. Firstly, our data pattern is highly consistent across experiments, and secondly, where we observe

any performance benefit, it is for a 'joint attention' display. We therefore propose that some form of higher-level perceptual grouping of related individual exemplars occurs, supporting processing of observed joint attention displays, particularly when the system is challenged further through inversion of the display. In this way, we argue that although we were indeed inspired by work on perception of mutual gaze to examine joint attention, and based our initial hypotheses on this prior work, direct comparisons are difficult to make since the very different stimuli necessitated a different task and overall approach that those studying mutual gaze were able to implement.

We acknowledge that the generalizability of our findings may be limited. Our study focused on investigating the possibility of a perceptual grouping effect in a convenience sample of educated predominantly Western, female, young, healthy adults. Moreover, face stimuli used in our study primarily represented the gender, age, and skin tone of our participant group. Considering differences in face processing among groups with diverse backgrounds and visual expertise (Tanaka et al., 2004), and differences in joint attention between men and women (Bayliss et al., 2005), future research exploring group and individual differences in the processing of observed gaze-based social interactions holds significant potential.

In conclusion, ~~we demonstrate~~our findings suggest that observed episodes of joint attention are processed particularly efficiently by the perceptual system, especially when the perceptual system is challenged. Our findings support the idea that socially interacting individuals are bound as a single perceptual unit based on high-level cues to social connectedness. Importantly, our results overcome alternative explanations based on direct attentional cueing between the social stimuli and demonstrate that these perceptual groupings are stronger for social stimuli and persist even in visually and socially complex scenes. By grouping individuals into socially connected pairs or groups, perceptual binding facilitates efficient processing of complex social scenes, enhances understanding of others' actions and interactions, and helps to identify opportunities for social interaction with others.

## Author Contributions

KM, SGE and APB conceived the ideas and designed the experiments. KM and SGE created the stimuli. KM programmed and conducted the experiments. KM analysed the data. All authors interpreted the data. KM and APB wrote the manuscript. All authors provided critical revisions throughout.

## Acknowledgements

We thank Blessing Chikota, Anisya Namyra Andi Djohar, Anthony Cheng, and Natalie Williams for data collection assistance and Charlotte Grove, Theodora Karadaki and Lisa J. Stephenson for assistance in developing early versions of stimuli and procedures.

## Funding

This project was funded by a Leverhulme Project Grants RPG-2016-173 and RPG-2023-106 awarded to APB and by a BIAL Foundation grant 147/18 to APB, LE and Lisa J. Stephenson.

## Declaration of conflicting interests

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

## References

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, *40*(4), 852-863.

Bayliss, A. P., Di Pellegrino, G., & Tipper, S. P. (2004). Orienting of attention via observed eye gaze is head-centred. *Cognition*, *94*(1), B1-B10.

Bayliss, A. P., Pellegrino, G. D., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology*, *58*(4), 631-650.

Bayliss, A. P., & Tipper, S. P. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology*, *96*(1), 95-114.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40*(8), 1129-1138.

Colombatto, C., Capozzi, F., Fratino, V., & Ristic, J. (2025). A perceptual advantage for social groups in interactive configurations. *Visual Cognition*, 1–12. https://doi.org/10.1080/13506285.2025.2462043

Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance, 6*(3), 404-412. doi:10.1037//0096-1523.6.3.404

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 628-634.

Flavell, J. C., Over, H., Vestner, T., Cook, R., & Tipper, S. P. (2022). Rapid detection of social interactions is the result of domain general attentional processes. *Plos one*, *17*(1), e0258832.

Fu, Y., Zhou, M., Zhou, J., Shen, M., & Chen, H. (2024). Unconscious prioritization for face-to-face people. *Journal of Experimental Psychology: General*, *153*(5), 1268

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, *133*(4), 694.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in cognitive sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Lobmaier, J. S., Fischer, M. H., & Schwaninger, A. (2006). Objects capture perceived gaze direction. *Experimental Psychology*, *53*(2), 117.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255-260.

McDonough, K. L., Edwards, S. G., Ewing, L., & Bayliss, A. P. (2023, July 12). *The joint attention grouping effect: Perceptual binding of observed social interactions*. OSF. osf.io/e7s4g. doi: 10.17605/OSF.IO/E7S4G.

Mersad, K., & Caristan, C. (2021). Blending into the crowd: electrophysiological evidence of gestalt perception of a human dyad. *Neuropsychologia*, *160*, 107967.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant behavior and development*, *21*(3), 469-482.

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human–human interactions. *Journal of Experimental Psychology: Human Perception and Performance, 45*(7), 877.

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, *28*(3), 369-379.

Strachan, J. W., Sebanz, N., & Knoblich, G. (2019). The role of emotion in the dyad inversion effect. *PloS one*, *14*(7), e0219185.

Stephenson, L. J., Edwards, S. G., & Bayliss, A. P. (2021). From gaze perception to social cognition: The shared-attention system. *Perspectives on Psychological Science*, *16*(3), 553-576.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*(1), B1-B9.

Tipples, J. (2005). Orienting to eye gaze and face processing. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 843.

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S. A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, *148*(7), 1251-1268.

Vestner, T., Gray, K. L., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, *200*, 104270.

Vestner, T., Gray, K. L., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, *208*, 104550.

Vestner, T., Gray, K. L., & Cook, R. (2022a). Sensitivity to orientation is not unique to social attention cueing. *Scientific reports, 12*(1), 5059.

Vestner, T., Over, H., Gray, K. L., & Cook, R. (2022b). Objects that direct visuospatial attention produce the search advantage for facing dyads. *Journal of Experimental Psychology: General*, *151*(1), 161.