Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-025-03119-1

Perceptual grouping of individuals in social triads

Luowei Yan[1] · Clara Colombatto[2] · Jelena Ristic[1]

Accepted: 3 June 2025

Abstract

Human life is built around the need for group membership and social connections. Recent research shows that small interactive groups of two and three individuals (i.e., dyads and triads) are found faster in visual search tasks when group members are facing toward versus away from one another. This 'facing advantage' may reflect the involvement of perceptual grouping processes, with facing groups perceived as a unified whole. Here, we tested this grouping hypothesis by measuring search performance for individuals who were positioned within facing or non-facing groups of three. If facing triads were perceptually grouped, individuation of group members in those triads should be hindered. Participants searched for a target individual, a person raising a fist or a person raising a pointing finger, who was positioned in one of four or eight facing or non-facing triads. The data indicated that while the search for target individuals pointing a finger was overall facilitated, it was specifically hindered when this person was positioned within a facing compared to a non-facing group. These results suggest that the perception of social groups may be attuned to the overall configuration of the group, but also to more sophisticated social communicative signals of individual group members.

Keywords Social groups · Group perception · Visual search · Perceptual grouping

Introduction

Humans have evolved in complex social environments in which group belonging was at a premium for survival. These social contexts have played an important role in shaping our neurocognitive system to support a range of social processes, including interactions, collaboration, and competition (Adolphs, 2009; Caporael, 1997; Colombatto et al., 2025; Malik & Isik, 2023; Vestner et al., 2024). Over the past decade, social perception research, or the study of how people perceive other agents, has shown that the human visual system seems to be specifically attuned to the detection, decoding, and understanding of other people (Quadflieg & Koldewyn, 2017; Ristic & Capozzi, 2022). In this study, we examined how individual group members are detected in small social groups of three (or triads).

✉ Jelena Ristic
jelena.ristic@mcgill.ca

[1] Department of Psychology, McGill University, 2001 Avenue McGill College, Montreal, QC H3A 1G1, Canada

[2] Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

Social information receives prioritized processing in the human visual system (Barzy et al., 2023; McMahon & Isik, 2023; Paparella & Papeo, 2022; Papeo, 2020; Wu et al., 2024). As evidence of the involvement of a phylogenetically old mechanism, visual tuning to social interactions has been observed across species and early in human development. Visually naïve chicks show a preference for face-to-face relative to back-to-back point-light hens (Zanon et al., 2024), while 6-month-old human infants attend to facing groups of two (or dyads) more than non-facing dyads (Goupil et al., 2022). In adults, facing groups are found faster (Goupil et al., 2024; Papeo et al., 2019; Vestner et al., 2021a, b; Yan et al., 2024), remembered better (Ding et al., 2017; Vestner et al., 2019), perceived as spatially closer (Sun et al., 2022; Vestner et al., 2019), and enter conscious awareness earlier (Su et al., 2016). Neuroimaging work suggests that this perceptual specificity may be associated with the working of a visual pathway specialized for the processing of social and interactive visual information (Isik et al., 2017; Pitcher & Ungerleider, 2021).

There are striking parallels between our visual perceptual preferences for social group configurations and real-life preferences for social group interactions. It has been documented that human social interactions most frequently

Published online: 19 June 2025

unfold in small groups of two or three and less commonly in groups of five or more (Caporael, 1997; Dunbar et al., 1995; Peperkoorn et al., 2020). Accordingly, visual search preferences for facing groups have been documented for groups of two, three, and five (Colombatto et al., 2025; Yan et al., 2024), but not for larger groups of six or more (e.g., crowds) where tracking of individual group members becomes more difficult (Yan et al., 2024) and group properties tend to be represented by overall ensemble statistics (Whitney & Yamanashi Leib, 2018). These effects of group size, which relate to social and interactive preferences in life, suggest that at least for small groups, the human visual system may be attuned to not only individual group members, but also to the social context in which those individuals form a group.

There remain outstanding questions regarding the cognitive and perceptual mechanisms underlying the perception of social groups. On the one hand, facing social groups may be detected more efficiently because of familiarity. As small groups like dyads and triads are commonly encountered in life and predominantly in a face-to-face orientation (Colombatto et al., 2025; Hall, 1963; Lu et al., 2023; McMahon & Isik, 2023; Williams & Chakrabarti, 2023), search advantages may reflect those familiar structures. This is consistent with results showing that the search advantage for facing groups is diminished or even abolished for groups presented in inverted orientations, which are rarely (if ever) encountered in life (Abassi & Papeo, 2020; cf. Colombatto et al., 2025; Mersad & Caristan, 2021; Papeo et al., 2019; Vestner et al., 2021a, b; Yan et al., 2024). On the other hand, search advantages for facing groups may also reflect sophisticated social influences on perceptual grouping, whereby the facing configuration may facilitate the perception of a cohesive group and individual members' social connections. Past work has shown that in the context of dyads, observers show greater difficulty identifying an individual positioned within a face-to-face or an interacting dyad compared to an individual positioned in a back-to-back or a non-interacting dyad (Papeo et al., 2019). That is, when participants are asked to search for an individual within a facing dyad, their response times are slower relative to when they are asked to search for an individual within a non-facing dyad. This is thought to reflect the involvement of perceptual grouping, i.e., the ensemble processing of multiple contextually related visual elements (Green & Hummel, 2006; McMains & Kastner, 2010; Papeo et al., 2019). Thus, complex social processes also influence perception, such that individuals who are interacting may be perceived as more unified and/or more socially cohesive than those represented as non-interacting.

The present study examined the extent to which facing groups of three or triads were perceived as cohesive units, thus exploring perceptual grouping as a mechanism underlying search advantages found in previous work (Colombatto et al., 2025; Yan et al., 2024). To assess this idea, we measured participants' responses to specific individual members located within facing and non-facing groups. Following the classic Gestalt theory, which shows that when visual elements are perceptually grouped into a whole, the processing of individual elements becomes impaired (Papeo et al., 2019; Wagemans, 2024), we reasoned that if triads were represented as a unified group, the search for individual group members in a grouped triad should be more difficult. We thus expected that, if facing (vs. non-facing) triads are represented as a coherent whole, visual search should be impaired for individuals in facing relative to non-facing groups.

Methods

The study was preregistered on the Open Science Framework at https://osf.io/vhrw6. The custom script for data exclusion and the summarized anonymized data for participants who have consented for data sharing are available at https://osf.io/vgm6h.

Participants

An a priori power analysis run using G*Power 3.1 (Faul et al., 2009) indicated that data from a sample of 34 participants would be needed to obtain a power of 0.80 at $\alpha = 0.05$ to detect a conventionally medium effect size of $d_z = 0.50$. To account for variability in effect size estimates (e.g., Giner-Sorolla et al., 2024; Pek et al., 2024) and ensure adequate statistical power, we increased our target sample size to 70.

A total of 90 participants were recruited from a volunteer undergraduate student pool who received course credits. Data from 73 participants were analyzed (women = 63, men = 10; mean age = 20 years, range = 18–24 years). Data from 17 participants were excluded based on the preregistered criteria, as their overall response accuracy fell below 65%.[1] For the remaining participants, using the same criteria, we further excluded trials in which participants did not respond during the response window (6.11%) and those with a response time 2.5 standard deviations below or above the individual's mean (0.49%). All participants reported normal or corrected-to-normal vision and provided informed consent prior to the experiment. All procedures were approved by the McGill University Research Ethics Board.

[1] Data from 15 participants were excluded based on the preregistered threshold of an observer's overall response accuracy falling below 60%. Data from two additional participants whose overall response accuracy was at 63% and 64% were also excluded, as they were identified as statistical outliers.

Apparatus and stimuli

The experiment was programmed in JavaScript using JsPsych library (de Leeuw et al., 2023) and deployed online on participants' own computers. Figure 1A shows the stimuli, which consisted of greyscale images of groups of three individuals, generated in DAZ Studio 4.22, which were presented against a white background. The width of each stimulus image was adjusted for each participant to 8% of their browser window's width, while the height of each stimulus was scaled to 5/6 of the image's width. Individual human models in each group were positioned at the vertices of an imaginary triangle, with the interpersonal angle and distance between the centres of two adjacent individuals maintained at 120 degrees and 180 units, respectively.

Thirty-six facing triads (Fig. 1A, top panel) were created, with 18 including a target individual displaying a fist raised up (Fig. 1A, left panel) and 18 including a target individual displaying a finger pointed up (Fig. 1A, right panel). The target individual could be placed in the left, center, or right location within the triad. Corresponding 36 non-facing triads (Fig. 1A, bottom panel) were created by rotating each individual 180 degrees around the y-axis such that they faced away from the centre. Target gestures (raised fist and pointed finger) were adapted from previous research (Papeo et al., 2019) and were designed to minimize variation in the physical appearance of the target while providing distinct options for target selection. To add variation to the search display, the posture of non-target individuals within the group varied slightly across six different poses (as illustrated in the Online Supplementary Materials (OSM), Fig. S1). Thus, a total of 72 stimulus triads were generated.

Design

The experiment was a repeated-measures design, with factors *Group* (2: Facing; Non-facing), *Set Size* (2: 4; 8), *Target Type* (2: Raised Fist; Pointed Finger), and *Target Location* (2: Left or Right side of the screen). Group, Set Size, and Target Location were intermixed. Target Type was blocked and randomized for presentation order across participants.

Group varied the type of triad as either facing or non-facing. For each search display, half the triads were facing and half were non-facing. *Set Size* varied the number of triads present in the search display between four and eight. *Target Type* varied between an individual showing a raised fist and an individual showing a pointed finger. The target individual was present in one triad in the display, while the remaining distractor triads displayed the individual showing the opposite non-target gesture. Half the blocks presented a target individual with a raised fist and the other half presented a target individual with a pointed finger. The response target was present on each trial. *Target Location* varied the position of the target individual between the left or right side of the screen and was distributed equally across trials.

Figure 1B illustrates a search display with a set size of four. The search display was designed so that triads were positioned along two imaginary ellipses centred on the screen (e.g., Yan et al., 2024). The inner ellipse spanned 25% of the browser window's width and height, and the outer ellipse spanned 40% of the browser window's width and height. The triads could be positioned at one of 16 possible locations along these two imaginary ellipses, set at 30, 60,

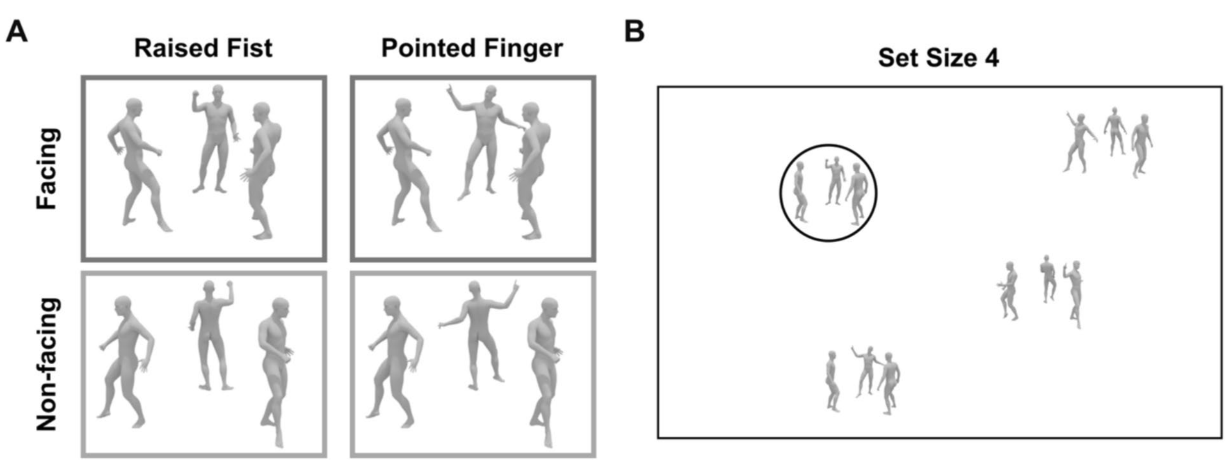

Fig. 1 Example stimuli and search display. **A** Example groups depicting target individuals with a raised fist and a pointed finger within facing (top) and non-facing (bottom) triads. **B** Example search display with a set size of four, including a target individual with a raised fist within a facing group, surrounded by triads containing distractor individuals with pointed fingers. *Note:* Stimuli are not drawn to scale. Target is highlighted for illustration purposes.

120, 150, 210, 240, 300, and 330 degrees. The triad containing the target could appear at one of the eight locations along the inner ellipse, with an equal chance of appearing on the left or right side of the screen. The specific placement within each side was randomly determined. The remaining three (in Set Size four) or seven (in Set Size eight) distractor triads could appear at any of the remaining 15 locations along the two ellipses. The number of triads was equated for each screen quadrant.

Target Placement and *Group Pose* were random variables, such that within each triad, the target individual could be placed in the left, center, or right location, and the pose of the two remaining non-target members was randomly selected from six possible body posture variations as in previous work (e.g.,Papeo et al., 2019; Yan et al., 2024; Fig. S1, OSM).

Procedure

At the beginning of each trial, a black central fixation cross (72 × 72 pixels) was presented for 500 ms. The search array appeared next and remained visible for 3,500 ms or until participants made a response. A 500-ms blank screen served as intertrial interval.

Participants completed 320 trials divided into eight blocks (four for each Target Type) of 40 trials, with eight unique (Group × Set Size × Target Location) combinations repeated five times. Each block started with an instruction to search for a target individual with either a raised fist or a pointed finger. Participants were asked to locate the target quickly and accurately as positioned either on the left or right side of the screen, by pressing 'B' or 'H' key on the keyboard. The assignment of target location to response key was counterbalanced across participants. The experiment lasted around 30 min. Sixteen practice trials were run at the start.

Results

Following the preregistered plan, the first set of analyses examined mean accuracy and correct response time (RT) using two separate two-way repeated-measures ANOVAs with *Group* (Facing; Non-facing) and *Set Size* (4, 8) included as variables. Greenhouse–Geisser corrected degrees of freedom are reported when the assumption of sphericity was violated. Bonferroni-corrected p values are provided for paired-samples two-tailed t-tests. Data exclusion was conducted in RStudio 2024.09.1 + 394. Data analyses were run in SPSS 29. We did not observe a speed-accuracy trade-off ($r(73) = -0.084$, $p = 0.480$).

Overall, task performance was high with an average response accuracy of 88.16%. The ANOVA on accuracy returned a significant main effect of Set Size ($F(1, 72) = 69.52$, $p < 0.001$, $\eta_p^2 = 0.491$), with more accurate responses in displays with a set size of four (M = 90.10%, SE = 0.73%) than those with a set size of eight (M = 86.23%, SE = 0.92%). No other effects were significant (Group, $F(1, 72) = 0.04$, $p = 0.851$, $\eta_p^2 = 0.000$; Group × Set Size, $F(1, 72) = 1.20$, $p = 0.278$, $\eta_p^2 = 0.016$).

The ANOVA on mean correct RTs also indicated a significant main effect of Set Size ($F(1, 72) = 1253.41$, $p < 0.001$, $\eta_p^2 = 0.946$), with target individuals located faster in displays with a set size of four (M = 1,552.28 ms, SE = 22.05 ms) than those with a set size of eight (M = 2,016.53 ms, SE = 24.56 ms). Neither the main effect of Group ($F(1, 72) = 0.85$, $p = 0.359$, $\eta_p^2 = 0.012$) nor the interaction between Group and Set Size were significant ($F(1, 72) = 0.16$, $p = 0.687$, $\eta_p^2 = 0.002$).

While this study was not specifically designed to investigate differences between gesture types as these were chosen for the purposes of target selection, a raised fist and a pointing finger differ in a socially meaningful way, given that pointing a finger typically involves a communicative function, more so than raising a fist. As a result, we examined the data in preregistered exploratory analyses involving *Target Type* as an additional variable. Two three-way repeated-measures ANOVAs examined mean accuracy and mean correct RT with *Group* (Facing; Non-facing), *Set Size* (4, 8), and *Target Type* (Raised Fist, Pointed Finger) included as variables.

Figure 2A plots mean accuracy as a function of Group and Target Type. In addition to a significant main effect of Set Size ($F(1, 72) = 70.22$, $p < 0.001$, $\eta_p^2 = 0.494$), this analysis also indicated a significant main effect of Target Type ($F(1, 72) = 12.99$, $p < 0.001$, $\eta_p^2 = 0.153$), with target individuals pointing a finger overall located more accurately (M = 89.90%, SE = 0.78%) than those raising a fist (M = 86.30%, SE = 1.08%). No other effects were significant (all Fs < 2.24, ps ≥ 0.139).

Figure 2B shows mean correct RT as a function of Group and Target Type. The RT analysis once again returned a significant main effect of Set Size ($F(1, 72) = 1151.93$, $p < 0.001$, $\eta_p^2 = 0.941$), with targets located faster in displays with a set size of four (M = 1,555.02 ms, SE = 22.26 ms) than those with a set size of eight (M = 2,019.23 ms, SE = 24.48 ms). There was also a significant main effect of Target Type ($F(1, 72) = 52.54$, $p < 0.001$, $\eta_p^2 = 0.422$), with target individuals pointing a finger (M = 1,730.54 ms, SE = 22.28 ms) located overall faster than those raising a fist (M = 1,843.70 ms, SE = 25.03 ms). Importantly, there was also a significant interaction between Group and Target Type ($F(1, 72) = 4.28$, $p = 0.042$, $\eta_p^2 = 0.056$), showing that target individuals pointing a finger who were placed within facing triads were found slower than target individuals pointing a finger who were placed within non-facing triads (Pointed

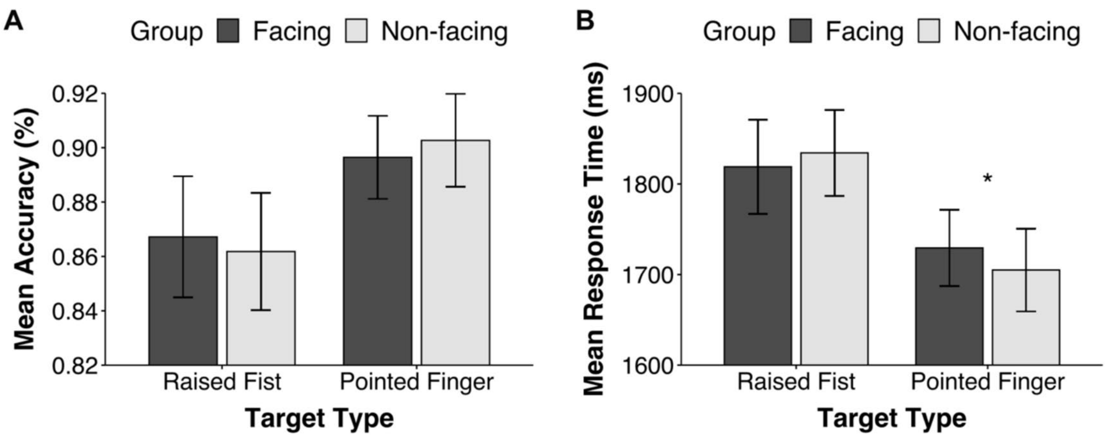

Fig. 2 Results. **A** Mean accuracy as a function of Group and Target Type. **B** Mean correct response time as a function of Group and Target Type. *$p < 0.05$. Error bars are 95% confidence intervals.

Finger in Facing, M = 1,743.15 ms, SE = 21.99 ms; in Non-facing, M = 1,717.94 ms, SE = 23.86 ms; $t(72) = 2.30$, $p = 0.024$, Cohen's $d_z = 0.270$). There was no significant difference in responses for target individuals raising a fist located in facing relative to non-facing triads (Raised Fist in Facing, M = 1,835.86 ms, SE = 27.13 ms; in Non-facing, M = 1,851.54 ms, SE = 24.99 ms; $t(72) = -1.07$, $p = 0.288$, Cohen's $d_z = -0.125$). Neither the main effect of Group ($F(1, 72) = 0.33$, $p = 0.570$, $\eta_p^2 = 0.005$) nor the three-way interaction between Set Size, Target Type, and Group were significant ($F(1, 72) = 2.24$, $p = 0.139$, $\eta_p^2 = 0.030$).

Thus, while no significant overall difference emerged between responses to targets in facing versus non-facing groups, further exploratory analyses examining the effects of specific target type revealed that only individuals pointing a finger were found slower in facing relative to non-facing triads, whereas no significant difference was found for individuals raising a fist.[2]

Discussion

Recent research on the perception of social groups has shown that observers are faster to detect social groups when group members are facing toward each other versus away from each other (Colombatto et al., 2025; Papeo et al., 2019; Yan et al., 2024). Here we examined whether such search preference reflected an increased grouping of individuals in facing groups of three, which is one of the interactive social group sizes most frequently encountered in life (Caporael, 1997; Dunbar et al., 1995; Peperkoorn et al., 2020; Ristic & Capozzi, 2022). Participants searched for a target individual who displayed either a raised fist or a pointed finger. The target was located within either a facing or a non-facing triad, and in a search display composed of four or eight groups. Results indicated that while search speed decreased with set size, grouping of facing triads varied with individual target type. Specifically, targets with a pointed finger were found slower when embedded in facing relative to non-facing triads. In contrast, no significant difference was found in the search performance for targets with a raised fist positioned in facing relative to non-facing triads.

These results suggest that the search advantage for facing groups is attuned to both the group's overall social structure (i.e., facing vs. non-facing groups) and individual group members' social signals (i.e., raising a fist vs. pointing a finger), as target individuation was hindered in facing triads for members displaying a pointed finger but not for those displaying a raised fist. This may be due to the perception of the interactive nature of the group, as a pointed finger and a raised fist differ in their social meaning, and certain individuals may be perceived as strengthening the groups' interactive social context. That is, an index finger pointed upward could be interpreted as a cue to draw others' attention (Pease, 1991; Saitz & Cervenka, 2019), while a fist raised upward could be associated with strength (Saitz & Cervenka, 2019). Further, a pointed finger is often described as a deictic or declarative gesture, and can be used to direct attention to objects or individuals within the environment (Cartmill & Goldin-Meadow, 2016). In this way, such a visual social signal could serve to strengthen the visual

[2] We report an additional experiment in the Online Supplementary Materials, which examined the role of overall group context. This experiment showed that the differences between individual target types were no longer significantly different when comparing responses to pointed-finger and raised-fist targets placed in groups in which all members were facing one another (i.e., all-facing triad) versus those in which only the target individual was turned away from an otherwise facing group (i.e., disrupted triad).

relationship between individual items, resulting in stronger perceptual grouping. Therefore, this finding suggests that in addition to the overall group structure (e.g., Colombatto et al., 2025), perceptual grouping may also be modulated by the social characteristics of individual group members.

Beyond the socio-communicative signals, these results may also reflect differences in the visual aspects of the two target individuals. Indeed, detecting a target with a pointed finger involves detecting the presence of a visual feature (an extended finger) as opposed to detecting the absence of such a feature (a raised fist). Our results support this general processing difference, with overall faster responses for target individuals with a pointed finger. However, our results also suggest that this general processing difference likely did not account for the individuation difference observed between the two target types, since hindered search individuation was present only for targets with a pointed finger and specifically when they were located in facing but not in non-facing groups. Thus, the visual difference appears meaningful only within a specific group context (i.e., facing triad).

Further, visual differences between the two response target individuals could have also arisen from slightly different body poses of the non-target members, which varied randomly across six different poses to provide variation in the display structure. To examine if group pose affected our main results, we ran an exploratory analysis on mean correct RTs in which *Group Pose* (six variations, depicted in Fig. S1, OSM) was included as an additional variable in the repeated measures ANOVA with *Group* (Facing, Non-Facing), *Set Size* (4, 8), and *Target Type* (Raised Fist, Pointed Finger). The results yielded one significant interaction with Group Pose that was not central to the current focus. An interaction between Target Type and Group Pose ($F(5, 310) = 2.44$, $p = 0.034$, $\eta_p^2 = 0.038$) showed that while targets with a pointed finger were found overall faster than those with a raised fist across all group poses, this difference was statistically smaller for group poses 4 and 5 (Group Pose 4, $t(72) = 3.38$, $p = 0.002$, Cohen's $d_z = 0.396$; Group Pose 5, $t(72) = 2.41$, $p = 0.028$, Cohen's $d_z = 0.282$; all other ts ≥ 4.40, ps < 0.001). No other effects involving Group Pose were significant (all Fs < 1.34, ps ≥ 0.254). Thus, although non-target group members varied in their poses, this variation in group structure did not significantly alter the main result which demonstrated slower individuation of individuals with a pointed finger in facing triads.

In an additional experiment (presented in the OSM), we examined the influence of the shared group context on individuation performance, by comparing individuation performance for triads in which all individuals were facing each other to triads in which only the target was non-facing, and other individuals remained facing the group. The results of this experiment suggested that group context may also influence the strength of grouping. That is, although hindered individuation was found for targets placed in facing triads in the main experiment, this effect was not statistically significant when only the target individual was turned away from the group. Thus, a person turned away from a facing group (i.e., disrupting an all-facing triad) appears to be processed differently than a person turned away from a non-facing group (i.e., aligning with an all-non-facing group). Future experiments are needed to examine if this finding depends on the similarity among group members, the overall social context of the group, potential subgrouping effects in the remaining non-target dyads, and/or the nature of individual group members' communicative gestures.

There are also several other future avenues worth pursuing. Cross-cultural variations exist both in the use and the meaning of gestures (Kendon, 1997; Kita, 2009). The 'OK' sign, made by forming a ring with the thumb and index finger, means 'good' in the UK, for example, but can be considered offensive in Greece. In Italy, for instance, index-finger pointing with the palm down pinpoints a distinct referent at the center of the conversation or discussion, whereas with a vertical palm it highlights a relevant but less central referent instead. Hence, it is possible that different cultural and social contexts may render different gestures socially relevant for some social groups and irrelevant for others. Future research may test a wider range of social actions and communicative gestures (e.g., individual with a finger pointing up vs. a finger pointing down) to explore their impact on social perceptual grouping. Examining cross-cultural or individual differences in visual experiences may also provide further insight into the mechanisms of social perceptual grouping. Visual experiences are documented to vary across cultures and research has shown that cultural background can influence visual perception (Blais et al., 2008; Estéphan et al., 2018). Hence, it is possible that individuals from collectivistic cultures relative to those from individualistic cultures may perceive groups and the corresponding group members differently, and as such the relationship between overall group structure and the influence of individual members on perception may vary as a function of larger societal values.

More generally within the context of visual processes, the results of this work also highlight the role of both the whole and the parts in visual social perception, and as such are well contextualized within the pioneering contributions of Mary A. Peterson, whose research has demonstrated the influence of both holistic and configural processes in visual object perception (Curby et al., 2024; Hochberg & Peterson, 1987; Peterson & Rhodes, 2003). An interplay between holistic and part-based processes in the perception of faces, objects, and scenes is particularly pertinent, indicating that the processes underlying object perception may also influence the processing of scenes (Peterson & Rhodes, 2003). Furthermore, Peterson's research on figure-ground organization highlights social groups as an interesting case for examining how individual members are perceived either as part of an integrated whole (the group) or as elements within a broader background (the crowd; Alt & Phillips, 2022; Peterson & Skow, 2008).

In summary, the current results build on existing research on the perception of facing groups (Colombatto et al., 2025; Papeo et al., 2019; Vestner et al., 2021a, b; Yan et al., 2024) to show that social grouping is influenced by group context and the characteristics of individual group members. As such, this work sheds light on the role of social signals, at both individual and group levels, as an important source of prioritization in human perception.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13414-025-03119-1.

Acknowledgements The authors acknowledge the contribution of Francesca Capozzi and Victoria Fratino to earlier versions of this project.

Author's contribution Conceptualization, methodology: all authors; data curation, formal analysis, investigation, visualization, writing – original draft: L.Y.; software and validation: C.C.; funding acquisition: all authors; project administration: L.Y. and J.R.; writing – review and editing: all authors; resources and supervision: J.R.

Funding This research was supported by fellowships from the Natural Sciences and Engineering Council of Canada (NSERC) and the Fonds de recherche du Quebec (FRQNT) to LY, grants from NSERC to CC and JR, and Social Sciences and Humanities Research Council (SSHRC), William Dawson, and GW Stairs funds to JR.

Data availability Behavioral data are available via the Open Science Framework at https://osf.io/vgm6h (Experiment 1) and https://osf.io/hy36m (Experiment 2).

Code availability The custom script for participant exclusions is available via the Open Science Framework at https://osf.io/vgm6h (Experiment 1) and https://osf.io/hy36m (Experiment 2).

Declarations

Ethics approval All procedures were approved by the McGill University Research Ethics Board.

Consent to participate Participants provided informed consent prior to the experiment.

Consent for publication Not applicable.

Conflicts of interest The authors declare no conflicts of interest.

References

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience, 40*(4), 852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60*(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Alt, N. P., & Phillips, L. T. (2022). Person perception, meet people perception: Exploring the social vision of groups. *Perspectives on Psychological Science, 17*(3), 768–787. https://doi.org/10.1177/17456916211017858

Barzy, M., Morgan, R., Cook, R., & Gray, K. L. (2023). Are social interactions preferentially attended in real-world scenes? Evidence from change blindness. *Quarterly Journal of Experimental Psychology*, 17470218231161044. https://doi.org/10.1177/17470218231161044

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), Article e3022. https://doi.org/10.1371/journal.pone.0003022

Caporael, L. R. (1997). The evolution of truly social cognition: The core configurations model. *Personality and Social Psychology Review, 1*(4), 276–298. https://doi.org/10.1207/s15327957pspr0104_1

Cartmill, E. A., spsampsps Goldin-Meadow, S. (2016). Gesture. In D. Matsumoto, H. C. Hwang, spsampsps M. G. Frank (Eds.), *APA handbook of nonverbal communication.* (pp. 307–333). American Psychological Association. https://doi.org/10.1037/14669-012

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

Curby, K. M., Teichmann, L., Peterson, M. A., & Shomstein, S. S. (2024). Holistic processing is modulated by the probability that parts contain task-congruent information. *Attention, Perception, & Psychophysics, 86*(2), 471–481. https://doi.org/10.3758/s13414-023-02738-w

de Leeuw, J. R., Gilbert, G. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351.

Ding, X., Gao, Z., & Shen, M. (2017). Two equals one: Two human actions during social interaction are grouped as one unit in working memory. *Psychological Science, 28*(9), 1311–1320. https://doi.org/10.1177/0956797617707318

Dunbar, R. I. M., Duncan, N. D. C., & Nettle, D. (1995). Size and structure of freely forming conversational groups. *Human Nature, 6*(1), 67–78. https://doi.org/10.1007/BF02734136

Estéphan, A., Fiset, D., Saumure, C., Plouffe-Demers, M.-P., Zhang, Y., Sun, D., & Blais, C. (2018). Time course of cultural differences in spatial frequency use for face identification. *Scientific Reports, 8*(1), 1816. https://doi.org/10.1038/s41598-018-19971-1

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

Giner-Sorolla, R., Montoya, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Aberson, C. L., Bostyn, D. H., Conrique, B. G., Ng, B. W., Schoemann, A. M., & Soderberg, C. (2024). Power to detect what? Considerations for planning and evaluating sample size. *Personality and Social Psychology Review*, 10888683241228328. https://doi.org/10.1177/10888683241228328

Goupil, N., Papeo, L., & Hochmann, J.-R. (2022). Visual perception grounding of social cognition in preverbal infants. *Infancy, 27*(2), 210–231. https://doi.org/10.1111/infa.12453

Goupil, N., Rayson, H., Serraille, É., Massera, A., Ferrari, P. F., Hochmann, J.-R., & Papeo, L. (2024). Visual preference for socially relevant spatial relations in humans and monkeys. *Psychological Science, 35*(6), 681–693. https://doi.org/10.1177/09567976241242995

Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1107–1119. https://doi.org/10.1037/0096-1523.32.5.1107

Hall, E. T. (1963). A system for the notation of proxemic behavior. *American Anthropologist, 65*(5), 1003–1026.

Hochberg, J., & Peterson, M. A. (1987). Piecemeal organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General, 116*(4), 370–380. https://doi.org/10.1037/0096-3445.116.4.370

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences, 114*(43), E9145–E9152. https://doi.org/10.1073/pnas.1714471114

Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, *26*(Volume 26, 1997), 109–128. https://doi.org/10.1146/annurev.anthro.26.1.109

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145–167. https://doi.org/10.1080/01690960802586188

Lu, X., Li, X., Hu, C., Deng, J., Sheng, W., & Zhu, L. (2023). A two-branch deep learning with spatial and pose constraints for social group detection. *Engineering Applications of Artificial Intelligence, 124*, Article 106583. https://doi.org/10.1016/j.engappai.2023.106583

Malik, M., & Isik, L. (2023). Relational visual representations underlie human social interaction recognition. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-43156-8

McMahon, E., & Isik, L. (2023). Seeing social interactions. *Trends in Cognitive Sciences, 27*(12), 1165–1179. https://doi.org/10.1016/j.tics.2023.09.001

McMains, S. A., & Kastner, S. (2010). Defining the units of competition: Influences of perceptual organization on competitive interactions in human visual cortex. *Journal of Cognitive Neuroscience, 22*(11), 2417–2426. https://doi.org/10.1162/jocn.2009.21391

Mersad, K., & Caristan, C. (2021). Blending into the crowd: Electrophysiological evidence of gestalt perception of a human dyad. *Neuropsychologia, 160*, Article 107967. https://doi.org/10.1016/j.neuropsychologia.2021.107967

Paparella, I., & Papeo, L. (2022). Chunking by social relationship in working memory. *Visual Cognition, 30*(5), 354–370. https://doi.org/10.1080/13506285.2022.2064950

Papeo, L. (2020). Twos in human visual perception. *Cortex, 132*, 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science, 30*(10), 1483–1496. https://doi.org/10.1177/0956797619867295

Pease, A. (1991). *Body language: How to read others' thoughts by their gestures* (16. impression). Sheldon Press.

Pek, J., Pitt, M. A., & Wegener, D. T. (2024). Uncertainty limits the use of power analysis. *Journal of Experimental Psychology: General, 153*(4), 1139–1151. https://doi.org/10.1037/xge0001273

Peperkoorn, L. S., Becker, D. V., Balliet, D., Columbus, S., Molho, C., & Lange, P. A. M. V. (2020). The prevalence of dyads in social life. *PLoS ONE, 15*(12), Article e0244188. https://doi.org/10.1371/journal.pone.0244188

Peterson, M. A., & Rhodes, G. (2003). *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*. Oxford University Press.

Peterson, M. A., & Skow, E. (2008). Inhibitory competition between shape properties in figure-ground perception. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 251–267. https://doi.org/10.1037/0096-1523.34.2.251

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences, 25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences, 1396*(1), 166–182. https://doi.org/10.1111/nyas.13331

Ristic, J., & Capozzi, F. (2022). Mechanisms for individual, group-based and crowd-based attention to social information. *Nature Reviews Psychology*, *1*(12), Article 12. https://doi.org/10.1038/s44159-022-00118-z

Saitz, R. L., & Cervenka, E. J. (2019). *Handbook of gestures: Colombia and the United States* (Vol. 31). Walter de Gruyter GmbH & Co KG. https://doi.org/10.1515/9783110810325-003

Su, J., van Boxtel, J. J. A., & Lu, H. (2016). Social interactions receive priority to conscious perception. *PLoS ONE, 11*(8), Article e0160468. https://doi.org/10.1371/journal.pone.0160468

Sun, Z., Ye, C., Sun, T., Yu, W., & Li, X. (2022). The spatial distance compression effect is due to social interaction and not mere configuration. *Psychonomic Bulletin & Review, 29*(3), 828–836. https://doi.org/10.3758/s13423-021-02045-1

Vestner, T., Gehdu, B. K., Gray, K. L. H., & Cook, R. (2024). Autistic adults exhibit a typical search advantage for facing dyads. *Autism Research, 17*(12), 2572–2578. https://doi.org/10.1002/aur.3265

Vestner, T., Gray, K. L. H., & Cook, R. (2021a). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition, 208*, Article 104550. https://doi.org/10.1016/j.cognition.2020.104550

Vestner, T., Over, H., Gray, K. L. H., Tipper, S. P., & Cook, R. (2021b). Searching for people: Non-facing distractor pairs hinder the visual search of social scenes more than facing distractor pairs. *Cognition, 214*, Article 104737. https://doi.org/10.1016/j.cognition.2021.104737

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. https://doi.org/10.1037/xge0000545

Wagemans, J. (2024). Perceptual Organization. *In Oxford Research Encyclopedia of Psychology*. https://doi.org/10.1093/acrefore/9780190236557.013.877

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Williams, E., & Chakrabarti, B. (2023). The integration of head and body cues during the perception of social interactions. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218231181001

Wu, J., Guo, Y., Chen, Z., Shen, M., & Gao, Z. (2024). Dual routes of chunking social interaction: Insights from grouping two agent actions in working memory. *Journal of Experimental Psychology: General.* https://doi.org/10.1037/xge0001539

Yan, L., Colombatto, C., & Ristic, J. (2024). Search advantages for facing social groups reflect optimal interactive group sizes. *iScience*, *27*(11), 111105. https://doi.org/10.1016/j.isci.2024.111105

Zanon, M., Lemaire, B. S., Papeo, L., & Vallortigara, G. (2024). Innate sensitivity to face-to-face biological motion. *iScience*, *27*(2), 108793. https://doi.org/10.1016/j.isci.2024.108793