



Tracking multiple perspectives: Spontaneous computation of what individuals in high entitative groups see

Xiaoyan He^{1,2} · Yingqiao Yang^{1,2} · Lan Wang^{1,2} · Jun Yin^{1,2}

Accepted: 4 December 2020 / Published online: 19 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

Perspective-taking ability is crucial for supporting social interactions. It has been widely suggested that the calculation of an individual's perspective is spontaneous. Nevertheless, people typically engage with more than one individual, and computing what individuals in a crowd see is important. The current study explored whether people spontaneously compute the perspectives of individuals displayed in a crowd. The classic visual perspective-taking task was adopted, but the picture of the room was presented with four human avatars facing two walls. The results showed that if the crowd of individuals was treated as a high entitative group, when none of the perspectives of the individuals contained the same number of discs as that from the perspective of the participant, the judgment of the participant's perspective was slower than when a proportion of the perspectives of the individuals displayed in the crowd were consistent with the participant's perspective, even if the perspectives of the multiple individuals in a crowd were not explicitly noticed. This altercentric intrusion effect was not present when the crowd had low entitativity. These findings were replicated by using different methods to operationalize group entitativity. Hence, this study demonstrates that spontaneously tracking the perspectives of individuals displayed in a crowd has a boundary condition and that people can spontaneously compute what individuals in high entitative groups see.

Keywords Perspective taking · Groups · Entitativity · Spontaneous computation

Introduction

In everyday life, what others see is usually different from our view of the surroundings; therefore, the ability to compute and determine someone else's perspective is fundamental to ensuring success in social interactions (Sebanz, Bekkering, & Knoblich, 2006). This visual perspective-taking ability has been suggested to be present from very early development (Kovács, Téglás, & Endress, 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Ward, Ganis, & Bach, 2019), and deficits in this ability lead to serious problems in adulthood (Hamilton, Brindley, & Frith, 2009; Todd & Simpson, 2016). Hence, how perspective taking works has attracted much attention from researchers.

Previous research has demonstrated through a wide range of procedures that the calculation of an individual's perspective is implicit and automatic rather than cognitively demanding (Apperly & Butterfill, 2009; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Kovács et al., 2010; Qureshi et al., 2010; Samson et al., 2010; Ward et al., 2019). One of these procedures, referred to as the visual perspective-taking task, has been used in a systematic way to examine spontaneous computation of what others can see (i.e., Level 1 perspective taking; this contrasts with Level 2 perspective taking, which involves processing how someone else sees a particular stimulus; Mattan, Quinn, Apperly, Sui, & Rotshtein, 2015; Qureshi et al., 2010; Samson et al., 2010; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). In this task, participants view a picture of a room with an individual avatar facing one of the walls that has discs displayed on the walls. The number of discs from the avatar's perspective may or may not be consistent with the participant's perspective. Using this task, it was found that even when the avatar's perspective was not explicitly emphasized, the avatar's perspective interfered with participants' explicit judgments about the relevant perspective (i.e., altercentric intrusion effect;

✉ Jun Yin
yinjun1@nbu.edu.cn

¹ Department of Psychology, Ningbo University, No. 616 Fenghua Rd., Ningbo 315211, China

² Center of Group Behavior and Social Psychological Service, Ningbo University, Ningbo, China

slower response times in the inconsistent compared to the consistent condition), indicating that computing another's perspective is spontaneous. Importantly, in a dual-task paradigm, the secondary task (e.g., making a response when hearing two auditory tones) did not affect the calculation of other individuals' perspectives (Qureshi et al., 2010), and the time pressure did not disrupt the Level-1 visual perspective calculation (Todd, Simpson, & Cameron, 2019). Further evidence has suggested that this visual perspective-taking task could reflect social processes instead of low-level stimulus feature effects or associations, as participants are sensitive to whether the avatar was able to see the discs (Furlanetto, Becchio, Samson, & Apperly, 2015). Specifically, the altercentric intrusion effect was found only when the goggles of an avatar were apparent but not when they were opaque.

In the above efforts, an isolated individual was the target of the computation in the visual perspective taking task. However, individuals are typically embedded in crowds, and crowds largely influence our interactive behaviors (Cracco & Cooper, 2019; Gallup et al., 2012; Moussaïd et al., 2016). When faced with multiple individuals, each individual holds a discrete perspective, and multiple competing perspectives may be involved, especially if some individuals see a target, while the other individuals do not. There have been few investigations of how people compute these divergent perspectives displayed by a crowd of individuals. For example, do people spontaneously compute (some of) the perspectives of the individuals displayed in a crowd?

Intuitively, when considering the spontaneous computation of a single individual's perspective, it might be expected that multiple perspectives of individuals displayed in a crowd could also be calculated automatically (here, we do not refer to all individuals' perspectives, but some perspectives are possible, such as those of the individuals in a crowd who see a target). This intuitive hypothesis was consistent with recent studies. Though rarely about perspective taking, experiments have suggested that individuals in a crowd are automatically processed with regard to their actions, gaze orientations, and so on (Capozzi, Bayliss, & Ristic, 2018; Cracco, & Cooper, 2019; Cracco, de Coster, Andres, & Brass, 2016; Sun, Yu, Zhou, & Shen, 2017). For instance, observing multiple actions activated mirror neuron systems more than a single action (Cracco et al., 2016), and when multiple gaze cues diverge in a crowd, the observers followed the gaze orientation displayed by most individuals (Capozzi et al., 2018). However, understanding actions and following gazes do not mean that what individuals see has been computed, as the perspective has its referred content, mostly located in the surroundings (Apperly & Butterfill, 2009). Importantly, one study directly explored perspective taking for multiple individuals. Capozzi et al. (2014) modified the classic perspective-taking task by including two avatars. In their experimental setting, the participant and the avatar would

sometimes see the same number of discs (consistent perspective) and sometimes see a different number of discs (the avatars were unable to see some of the discs visible to the participant; inconsistent perspective). It was found that when these two avatars had conflicting perspectives (e.g., one of them looking at discs but the other looking at an empty wall), the altercentric intrusion effect disappeared. Hence, the individuals' discrepant perspectives may not be spontaneously tracked, which was not consistent with the intuitive hypothesis. Therefore, caution should be taken before concluding that spontaneously calculating one individual's perspective extends to calculating multiple perspectives.

With regard to individuals in a crowd, the efficiency of processing has been largely suggested to be dependent upon the group entitativity formed by them, that is, the degree to which a crowd of individuals is perceived as a unified group (Crawford, Sherman, & Hamilton, 2002; Ding, Gao, & Shen, 2017; Hamilton & Sherman, 1996; Yin, Ding, Zhou, Shui, Li, & Shen, 2013; Yin, Xu, Duan, & Shen, 2018). For example, the transference of traits from one group member to other group members was stronger in high-entitative groups than in low-entitative groups (Crawford et al., 2002). Importantly, even when one of the members in a group was asked to be selected, the other was automatically selected as well (Yin et al., 2018), that is, binding individuals into groups leads to faster processing and enhanced memory of interacting partners (Ding et al., 2017; Vestner, Tipper, Hartley, Over, & Rueschemeyer, 2019). Hence, the efficiency of processing individuals' features can be enhanced when these individuals are treated as an actual group, which is a process similar to perceptual grouping (Vestner et al., 2019; Yin et al., 2018). In this case, given that group entitativity determines the "groupness" of individuals and perspectives are considered social properties of individuals, we hypothesized that group entitativity would modulate whether people spontaneously calculate the perspective of individuals displayed in a crowd. Specifically, when a crowd of individuals has high group entitativity and given that individuals in this group can be simultaneously selected and efficiently processed, the perspective of at least some of the individuals displayed in the crowd would be spontaneously computed and subject to showing the altercentric intrusion effect (i.e., some perspectives of the individuals displayed in a crowd interfere with the participant's perspective). However, when a crowd of individuals has low group entitativity, given that the individuals in this group are more likely to be independently selected, the perspectives of some individuals displayed in this crowd would be not simultaneously tracked and, thus, show none of the altercentric intrusion effects.

To test the above hypothesis, this study adopted a similar visual perspective-taking task to that of Capozzi et al. (2014), and presented a crowd of human avatars facing different walls of a room (see Fig. 1). The condition of

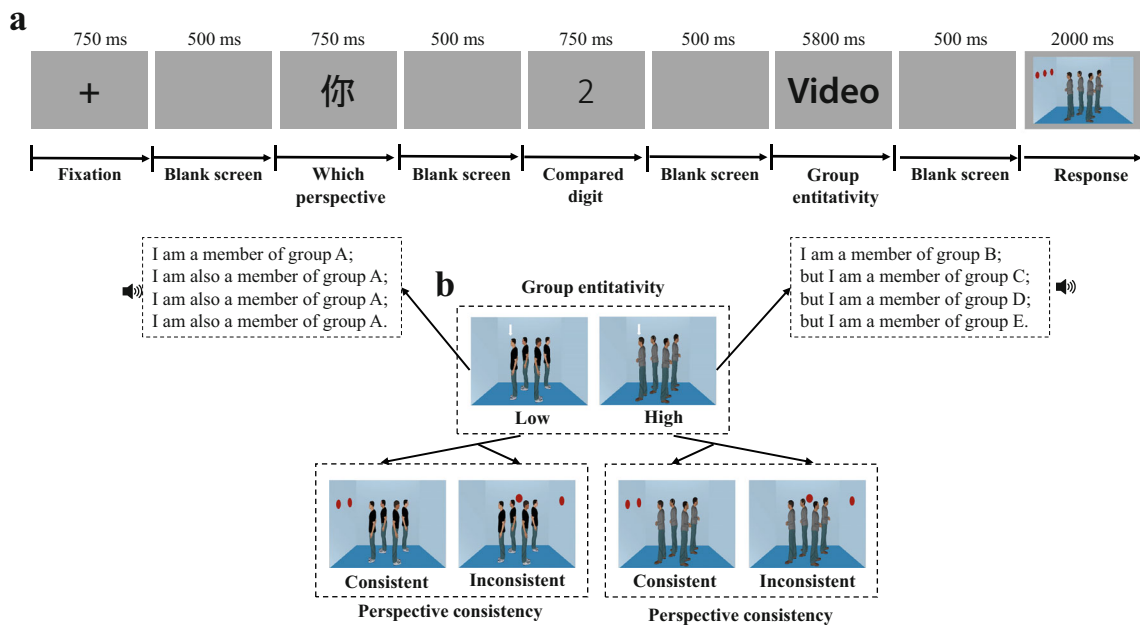


Fig. 1 Overview of the experimental procedure and conditions where (a) represents an example trial sequence of experiment, and (b) shows how the group entitativity was manipulated and how each condition was conceptualized in Experiment 1

consistent perspective was set such that some of the multiple individual perspectives and the participant's perspective had the same number of discs; the condition of inconsistent perspective was set such that none of the perspectives from the individuals contained the same number of discs as that from the perspective of the participant. To ensure the replicability of the conclusion, we used two methods – language labels (Experiment 1) and visual cues (Experiment 2) – to manipulate group entitativity. The spontaneous computation of what individuals in a crowd see would lead to a slower response under the inconsistent perspective condition than under the consistent perspective condition.

Experiment 1

This experiment followed the same visual perspective-taking task as Capozzi et al. (2014), with the exception that the picture of the room was presented with four human avatars facing two of the walls. The reason why we adopted this setting was that at least three persons were usually used to manipulate the group information (Gaertner & Insko, 2000; Petersen & Blank, 2003). With two avatars facing each side wall (the back wall was not used as the avatar's perspective was not clear from participant's view), different views can be formed and have equal salience regarding the number of avatars with different perspectives; otherwise, participants' attention would be oriented to the perspective that involved more persons.

Methods

Participants Twenty-four participants (10 males and 14 females) between 17 and 25 years of age ($M = 20.04$, $SD = 2.05$) were recruited on campus and paid approximately \$3 to participate in this experiment. The sample size was determined by a power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). We ran a pilot study in which only the condition of the high entitative group was included by presenting four similar static avatars, recruited 24 valid participants, and obtained a medium effect size ($\eta_p^2 = 0.22$ in SPSS; consistent perspective vs. inconsistent perspective). To be conservative, a medium effect size as recommended by Cohen (1988; $f = 0.25$, $\eta_p^2 = 0.15$ with the option “as in SPSS” in G*Power) was expected. With this expected effect size, an alpha level of .05, power of .80, and a 2×2 within-subjects design, the suggested sample size was 24 individuals. This sample size was also close to that of a previous study that explored perspective taking for multiple individuals, with a sample size of 22 (Capozzi et al., 2014). As both experiments conducted a 2×2 within-subjects design, 24 valid individuals in each experiment were recruited. All participants were right-handed, had normal or corrected-to-normal vision, and were naïve about the purpose of the experiment. All participants received information sheets about the experimental procedure and signed informed consent forms after learning the purpose and procedure of the experiment. The experimental protocol was approved by the Institutional Review Board at the Department of Psychology of the authors' university. In all experiments, we report all measures,

manipulations, and exclusions. All materials and data can be accessed at https://osf.io/fp49e/?view_only=474d334ecd11433e95cab94d16f64c7c. None of the experiments was preregistered.

Apparatus and stimuli The stimuli were all presented on a gray background using a 19-in. cathode ray tube monitor (1,024 × 768 pixel resolution; 100-Hz refresh rate) at a 60-cm viewing distance using PsychoPy (Peirce et al., 2019).

The stimuli consisted of a picture showing a lateral view into a room with the left, back, and right walls visible and with red discs displayed on walls (each subtending $1^\circ \times 1^\circ$). The three-dimensional (3D) human avatars were created with Poser 11 software and had different faces and postures but were the same gender and wore the same clothes (see Fig. 1). The avatars were positioned at the center of the room and equally distanced from each other. They stood back-to-back in pairs, such that two avatars faced the left wall and two faced the right wall. Female participants were presented with female avatars, and male participants were presented with male avatars.

To achieve a consistent perspective condition, all discs were placed on either the left or right wall, i.e., two avatars saw the same discs as the participants, and two avatars saw nothing. Under the inconsistent perspective condition, one disc was placed at the back wall, and accordingly, none of the avatars could see the disc, which resulted in none of the avatars seeing the same discs as the participants. To match this setting under the condition of an inconsistent perspective, for the remaining discs, two avatars could see the discs, and two avatars could not. The position of the avatars was kept constant across trials, but the position and number of discs changed. The pairing of discs with the avatars' perspectives is shown in Figure S1.

Procedure and design Each trial started with the presentation of a fixation cross for 750 ms (see Fig. 1). After 500 ms, the word “你” (i.e., you) was presented for 750 ms, instructing the participants to judge how many discs could be seen from their own perspective. Following this display, a digit (0–3) appeared for 750 ms, which specified the perspective content for the participant to verify. Then, an animation lasting 5,800 ms was presented to show different manipulations of group entitativity, similar to the procedure of Ting, He, and Baillargeon (Ting, He, & Baillargeon, 2019). After an interval of 500 ms, the picture of the room appeared and remained on the screen until a response was made within 2 s. The participants were asked to report whether the number of discs from their own perspective matched the digit previously presented as quickly as possible by pressing the “Left Arrow” or “Right Arrow” button on a standard keyboard. The next trial was initiated after a 1.5- to 2.5-s intertrial interval.

In 50% of the trials, the animation contained a high entitativity group, in which four avatars were pointed at by an arrow one by one, accompanied by identical language labels (i.e., “I am a member of group A; I am also a member of group A; I am also a member of group A; I am also a member of group A”). In the other 50% of trials, the animation contained a low entitativity group, in which four avatars were pointed at by an arrow in sequence but accompanied by different language labels (i.e., “I am a member of group B; but I am a member of group C; but I am a member of group D; but I am a member of group E”).

In each type of group entitativity, in 50% of the trials, some of the multiple perspectives can see the same number of discs as that from the perspective of participant (i.e., consistent perspective condition); in 50% of the trials, none of multiple perspectives had the same number of discs as that from the perspective of the participant (i.e., inconsistent perspective condition). On matching trials, the digit specifying the perspective content always corresponded with the number of discs seen from the participant's perspective. On mismatching trials, the number of discs seen from the participant's perspective was different from the digit presented. For each group entitativity condition, there were 48 matching trials with 24 consistent perspective trials and 24 inconsistent perspective trials. There were 48 mismatching trials with 24 consistent perspective trials and 24 inconsistent perspective trials. Based on previous studies (Samson et al., 2010; Santiesteban, et al., 2014), we also added 16 filler trials where no discs were displayed on the wall so that “0” would also sometimes be the correct answer; each group entitativity condition had 8 filler trials. These filler trials included an equal number of consistent and inconsistent trials and matching and mismatching trials. The trials were divided into four blocks of 52 test trials (48 test trials and 4 filler trials). The order of the trials within a block was completely randomized. The formal experimental trials were preceded by a block of 26 practice trials. Participants completed the whole experiment within 30 min.

Considering that mismatching trials were included to balance the trials, as all previously published studies emphasized (Samson et al., 2010; Santiesteban, et al., 2014; Ramsey, Hansen, Apperly, & Samson, 2013), we analyzed performance only during the matching trials. In this case, the experiment formed a 2 (group entitativity: low vs. high) × 2 (perspective consistency: consistent vs. inconsistent) within-subjects design. The response time was analyzed with repeated-measures analysis of variance (ANOVA) after excluding trials with wrong responses and response times more than 2.5 standard deviations (SDs) from the mean (exclusion rate = 7.29%). This exclusion method was based on previous studies (Furlanetto et al., 2015; Surtees, Apperly, & Samson, 2013). See the supplementary information for an analysis of the accuracy of the data. In addition to the traditional null

hypothesis significance testing procedure, we conducted Bayesian analyses (Cumming, 2014) to compute the ratio of the likelihood probability of two competing hypotheses. We computed the Bayes Factor (BF_{10} , $H1/H0$ as computed here) using JASP software (JASP Team, 2020). In the Bayesian analysis, $BF_{10} < 0.33$ suggests substantial evidence against between-condition differences, while $BF_{10} > 3.00$ suggests substantial evidence supporting between-condition differences. Values between 0.33 and 3.00 are considered inconclusive (Dienes, 2014).

To verify the manipulation of group entitativity, after completing the perspective-taking task, the participants were asked to watch the animations presenting different levels of entitativity and evaluate whether the avatars looked like a unified group (i.e., group entitativity; Campbell, 1958) using three items. These three items were established according to the definition and characteristics of a social group (Morewedge et al., 2013). Generally, a social group consists of at least three persons who are similar to each other, have a collective interest, and depend on each other. Hence, the participants were asked to evaluate the group entitativity based on whether the four avatars in the animation (1) were similar to each other, (2) had a collective interest, and (3) depended on each other; the evaluation was rated on a 7-point scale (1 = *not at all*, 7 = *very much*). The order of presentation of the different animations was counterbalanced across participants. It was found that the group's entitativity was evaluated as greater in the condition of high entitativity ($M = 4.89$, $SD = 0.77$) than in the condition of low entitativity ($M = 3.88$, $SD = 0.93$), $t(23) = 4.26$, $p < 0.001$, Cohen's $d = 0.87$, 95% CI of mean difference = [0.52 1.51], $BF_{10} = 101.94$.

Results

The dependent variable was the reaction time for judging the number of discs from one's own perspective (Fig. 2). We analyzed these reaction times using 2×2 repeated-measures ANOVA with group entitativity (low vs. high) and perspective consistency (consistent vs. inconsistent) as within-subject variables. It was revealed that both the main effect of perspective consistency ($F(1, 23) = 4.99$, $p = 0.035$, $\eta_p^2 = 0.18$, $BF_{10} = 2.27$) and the interaction effect between the two factors ($F(1, 23) = 11.02$, $p = 0.003$, $\eta_p^2 = 0.32$, $BF_{10} = 3.76$) were significant, but the main effect of group entitativity was not significant, $F(1, 23) = 0.03$, $p = 0.872$, $\eta_p^2 < 0.01$, $BF_{10} = 0.24$. The analysis of simple effects showed that when the group entitativity was high, if none of the multiple perspectives was consistent with the participant's perspective, the participants were slower ($M = 587$ ms, $SD = 82$) to confirm how many discs they (i.e., the participants) observed than if some of the multiple perspectives (i.e., at least one) were consistent with the participant's perspective ($M = 566$ ms, $SD = 64$), $t(23) = 3.79$, $p = 0.001$, Cohen's $d = 0.77$, 95% CI of mean

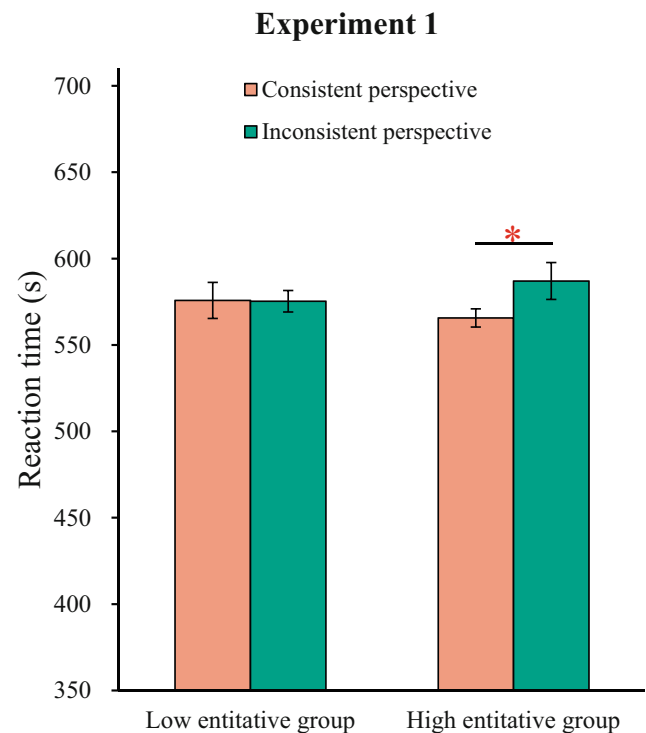


Fig. 2 Mean reaction times in Experiment 1 as a function of group entitativity and perspective consistency. Error bars indicate within-subject 95% confidence intervals. The asterisk represents a significant difference in reaction times between the conditions

difference = [10, 33], $BF_{10} = 36.39$. However, no such effect was found when the group entitativity was low (consistent perspective: $M = 576$ ms, $SD = 64$; inconsistent perspective: $M = 576$ ms, $SD = 70$), $t(23) = 0.08$, $p = 0.938$, Cohen's $d = 0.02$, 95% CI of mean difference = [-12, 11], $BF_{10} = 0.22$. Hence, the altercentric intrusion effect was found for the irrelevant perspectives of individuals from the high-entitative group, suggesting that multiple perspectives in high-entitative groups are spontaneously computed.

Experiment 2

To further examine the replicability of the findings in Experiment 1, we used a new method to operationalize group entitativity by varying the visual information indicating entitativity, such as common fate among individuals (Campbell, 1958; Lakens, 2010).

Methods

Twenty-four new students (9 males and 15 females) between 18 and 24 years of age ($M = 20.50$, $SD = 2.28$) who did not participate in Experiment 1 were paid approximately \$3 to participate in this experiment. The stimuli and the procedure were the same as those of Experiment 1, with the exception of

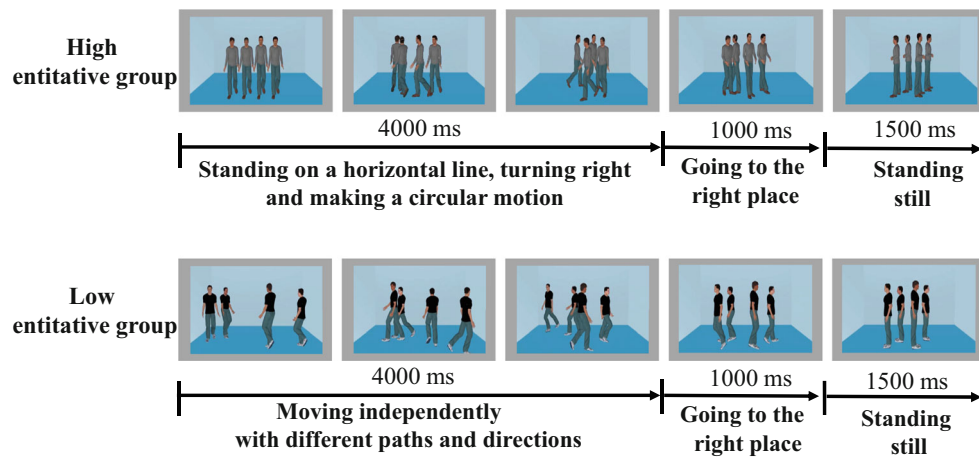


Fig. 3 How the group entitativity was manipulated in Experiment 2

the animations used to operationalize group entitativity and their durations (see Fig. 3). These animations lasted 6,500 ms presented different visual cues about group entitativity, similar to those of Powell and Spelke (2013). In 50% of the trials, an animation was shown with a high entitative group, in which four avatars first turned to the front of the screen and stood on a horizontal line. Then, all of them turned to the right and made synchronized movements around a circular path. In the other 50% of the trials, an animation was shown with a low entitative group in which the four avatars moved independently in different paths and directions. Hence, this experiment was also a 2 (group entitativity: low vs. high) \times 2 (perspective consistency: consistent vs. inconsistent) within-subjects design. The same method as Experiment 1 was adopted to exclude trials before computing response times, and 7.03% of trials were finally excluded.

After completing the perspective-taking task, the participants were asked to evaluate group entitativity for the four avatars in the animations displaying different levels of group entitativity, as was done in Experiment 1. The analysis revealed that the participants evaluated the group entitativity higher in the condition of high group entitativity ($M = 5.04$, $SD = 0.92$) than in the condition of low group entitativity ($M = 2.39$, $SD = 0.89$), $t(23) = 11.61$, $p < 0.001$, Cohen's $d = 2.37$, 95% CI of mean difference = [2.18, 3.13], $BF_{10} = 2.15 \times 10^8$. Hence, this manipulation was valid.

Results

We analyzed the reaction times using 2×2 repeated-measures ANOVA with group entitativity (low vs. high) and perspective consistency (consistent vs. inconsistent) as within-subject variables. Similar to Experiment 1 (Fig. 4), we found that both the main effect of perspective consistency ($F(1, 23) = 7.21$, $p = 0.013$, $\eta_p^2 = 0.24$, $BF_{10} = 4.52$) and the interaction effect between the two factors ($F(1, 23) = 13.36$, $p = 0.001$, $\eta_p^2 = 0.37$, $BF_{10} = 12.57$) were significant, but the main effect of

group entitativity was not significant, $F(1, 23) = 0.04$, $p = 0.850$, $\eta_p^2 < 0.01$, $BF_{10} = 0.23$. The analysis of simple effects showed that when the group entitativity was high, if the number of discs from any of the perspectives of the individuals displayed in this crowd was inconsistent with the participant's perspective, the participants were slower ($M = 576$ ms, $SD = 81$) to confirm how many discs they (i.e., participants) observed than if the number of discs from some individuals' perspectives was consistent ($M = 559$ ms, $SD = 63$), $t(23) = 3.88$, $p = 0.001$, Cohen's $d = 0.79$, 95% CI of mean difference = [8, 30], $BF_{10} = 44.35$. Meanwhile, no such effect was found

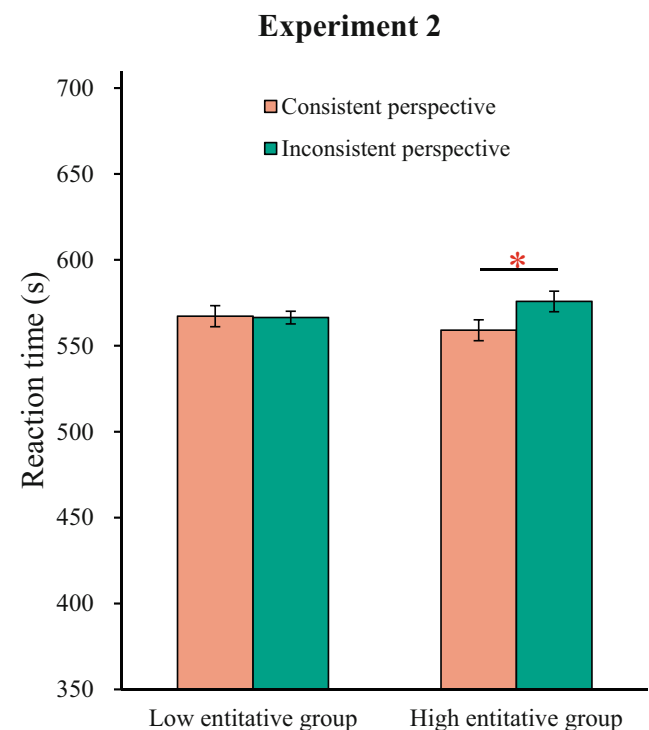


Fig. 4 Mean reaction times in Experiment 2 as a function of group entitativity and perspective consistency. Error bars indicate within-subject 95% confidence intervals. The asterisk represents a significant difference in reaction times between the conditions

when the group entitativity was low (consistent perspective: $M = 567$ ms, $SD = 74$; inconsistent perspective: $M = 567$ ms, $SD = 77$), $t(23) = 0.26$, $p = 0.801$, Cohen's $d = 0.05$, 95% CI of mean difference = $[-8, 6]$, $BF_{10} = 0.22$. Again, the findings suggested that the participants spontaneously computed what the individuals displayed in the high entitative group see.

Discussion

The current study explored whether people spontaneously compute the perspective of (some) individuals displayed in a crowd. It was found that if the crowd of individuals was treated as a high entitative group, when the number of discs from any of the perspectives of the individuals displayed in this crowd was inconsistent with that of the participant's perspective, the judgment of the participant's perspective was slower than that when a proportion of the perspectives of the individuals displayed in this crowd were consistent, even if the multiple perspectives were not explicitly noticed. This altercentric intrusion effect was not present when the crowd had low entitativity. Furthermore, these findings were replicable when using different methods to operationalize group entitativity. Hence, this finding demonstrates that spontaneously computing the perspectives of individuals displayed in a crowd has a boundary condition and that people can spontaneously compute what individuals in high entitative groups see.

Nevertheless, it is unclear whether our findings fully support our conclusion. First, because the disc on the back wall was always placed in the inconsistent condition (see Fig. 1) and never in the consistent condition, it may be argued that this perceptual cue led to the critical difference between these two conditions. The same perceptual differences were used in low- and high-entitativity conditions. However, only the perspectives displayed by the individuals in the high-entitative group, who were likely to be treated as a unified group, were spontaneously computed, thereby excluding the perceptual explanation. Second, this finding may simply indicate a failure to replicate the basic effect of perspective taking. However, in the trials mixed with the conditions in which the crowd had high entitativity, it was rarely probable to only focus on the latter trials and, thus, fail to compute what the others see only when the crowd had low entitativity. Moreover, the reaction times did not significantly differ across the different group entitativity conditions, which, to some extent, suggests that both settings have the same salience. Hence, the different effects between the low-entitativity and high-entitativity conditions largely reflect that computing the perspectives of individuals displayed in a crowd is modulated by the group entitativity. Finally, it could be argued that only one individual facing the discs was tracked from his/her perspective. It is possible because this computed output can also interfere with the participant's perspective. However, this argument does not

contradict our conclusion, as we claimed that only a proportion of individuals' perspectives were computed and not all of them. It is possible that the number of spontaneously tracked perspectives has a bottleneck, which needs to be addressed in the future. Regardless of the mechanism, our findings support the idea of the spontaneous computation of what (some) individuals in a high entitative group see.

This study extends the target of perspective taking from one individual to multiple individuals in a crowd (Apperly & Butterfill, 2009; Kovács et al., 2010; Ward et al., 2019; Wu, Street, Richardson, Street, & Kirkham, 2010). However, the spontaneity of this process has its boundary condition when extended to multiple perspectives and depends upon the entitativity of the displayed crowd of individuals. To some extent, this conclusion implies that the processing of others' perspectives is partially automatic, in accordance with the research of Capozzi et al. (2014). In their study, when the avatars had converging lines of sight, their perspectives interfered with the participant's perspective, whereas when they had conflicting perspectives (e.g., one avatar looked at discs but the other looked at an empty wall), the altercentric intrusion effect disappeared. In our study, at a global level, the avatars' perspectives were competing, although a proportion of them converged to the same discs, but we still found that altercentric intrusion was apparent with the high-entitative group. It is possible that in the study by Capozzi et al. (2014), avatars with diverging perspectives may be considered a low-entitative group, showing the same null effect that the current study revealed. If this is the case, the factor that determines the boundary of spontaneously processing others' perspectives is not the convergence of multiple perspectives but the entitativity of a group that contains multiple individuals.

Whether the spontaneous perspective-taking effect (e.g., the altercentric intrusion effect) truly reflects the social mentalizing nature of tracking perspectives remains controversial (Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Marshall, Gollwitzer, & Santos, 2018; Ramsey et al., 2013; Santiesteban et al., 2014; Todd et al., 2019; Ward et al., 2019). Some scholars believe that this effect simply reflects a domain-general process as the effect was unaltered when the avatar was replaced with an inanimate arrow stimulus (Conway et al. 2017; Santiesteban et al., 2014). Although the current study did not aim to address this debate, it documents that the social nature of processing perspectives can be manifested in this effect as the social information of group entitativity alters the altercentric intrusion effect. This observation may indicate how the individuals embedded in a context would change the portent of perspective taking, and the previous findings should be further examined while controlling the contexts. Furthermore, the ability to spontaneously track individuals' perspectives provides a new avenue for understanding "groupthink," which may serve as a cognitive

process to rapidly trace the group's mind when engaging with a cohesive group. When individuals' perspectives are available, it would be easy to identify what individuals in the group agree with and what information the group has obtained. The finding of spontaneously computing the perspective of individual displayed in a high-entitative group has its functional adaptation. It has been suggested that the ability of perspective taking is adapted to form shared reality with others (Hodges, Denning, & Lieber, 2018). High-entitative groups are usually created to maintain high cohesion and pursue a common goal, and accordingly, the shared reality across members is necessary. This driving force may lead participants to efficiently calculate what the individuals in a high entitative group see. In this case, the individuals' perspectives can be immediately available, and the commonness across members could easily be extracted to create shared reality.

Additionally, future studies could investigate whether the number of persons in a group influences the computation of the group's perspective to examine the generalization of the current conclusion. Due to space limitations in the visual scene, the current study used only four avatars, which is within the capacity of visual working memory (Luck & Vogel, 2013). Future research could place more avatars in the visual scene and give each avatar different viewing contents. Additionally, we presented conditions in which a proportion of individuals see discs but the other individuals see nothing to simulate what people usually face in daily life and to be able to match the experimental displays. However, individuals in a crowd often see their own targets but with different contents. For example, in a crowd two individuals see two objects, and two individuals see one object. Future studies could adopt this setting to test the generalizability of the current conclusion. As a null effect was found under the condition of low entitativity, it is important to replicate this finding using other designs and methods – for instance, manipulating the number of individuals in the crowd, and find whether there are boundary conditions of the effect reported.

In conclusion, our results suggest that spontaneously calculating the perspectives of individuals displayed in a crowd has a boundary condition and that people can spontaneously compute what individuals in high entitative groups see. This ability might be a core cognitive process to understanding crowd behaviors and guide social activities in humans.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-020-01857-x>.

Acknowledgements This work was supported by the Special Foundation of Philosophy and Social Science in Zhejiang Province for Young Talents (浙江省哲学社会科学领军人才培养专项课题 (青年英才培育); Grant no. 21QNYC12ZD).

Data availability All materials and data can be accessed at https://osf.io/fp49e/?view_only=474d334ecd11433e95cab94d16f64c7c. None of the experiments was preregistered.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970.
- Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*(1), 14–25.
- Capozzi, F., Bayliss, A. P., & Ristic, J. (2018). Gaze following in multiagent contexts: Evidence for a quorum-like principle. *Psychonomic Bulletin & Review*, *25*(6), 2260–2266.
- Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric Intrusions from Multiple Perspectives: Beyond Dyads. *PLOS ONE*, *9*(12), e114210.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Routledge.
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454–465.
- Cracco, E., & Cooper, R. P. (2019). Automatic imitation of multiple agents: A computational model. *Cognitive Psychology*, *113*, 101224.
- Cracco, E., de Coster, L., Andres, M., & Brass, M. (2016). Mirroring multiple agents: motor resonance during action observation is modulated by the number of agents. *Social Cognitive and Affective Neuroscience*, *11*(9), 1422–1427.
- Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of personality and social psychology*, *83*(5), 1076–1094.
- 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.
- 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.
- Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2015). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 158–163.
- Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: categorization, reciprocation, or fear? *Journal of Personality and Social Psychology*, *79*(1), 77–94.
- Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Gamier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences*, *109*(19), 7245–7250.
- Hamilton, A. F. de C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37–44.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336–355.

- Hodges, S. D., Denning, K. R., & Lieber, S. (2018). Perspective taking: motivation and impeding to shared reality. *Current Opinion in Psychology*, 23, 104–108.
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830–1834.
- Lakens, D. (2010). Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology*, 46(5), 701–708.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400.
- Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two level-1 perspective-taking tasks. *PLOS ONE*, 13(3), e0194101.
- Mattan, B., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective-taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4), 1100–1117.
- Morewedge, C. K., Chandler, J. J., Smith, R., Schwarz, N., & Schooler, J. (2013). Lost in the crowd: Entitative group membership reduces mind attribution. *Consciousness and Cognition*, 22(4), 1195–1205.
- Moussaïd, M., Kapadia, M., Thrash, T., Sumner, R. W., Gross, M., Helbing, D., & Hölscher, C. (2016). Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface*, 13(122), 20160414.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Petersen, L.-E., & Blank, H. (2003). Ingroup bias in the minimal group paradigm shown by three-person groups with high or low state self-esteem. *European Journal of Social Psychology*, 33(2), 149–162.
- Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, 110(41), E3965–E3972.
- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, 117(2), 230–236.
- Ramsey, R., Hansen, P., Apperly, I., & Samson, D. (2013). Seeing it my way or your way: Frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *Journal of Cognitive Neuroscience*, 25(5), 670–684.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266.
- Santesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929–937.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76.
- Sun, Z., Yu, W., Zhou, J., & Shen, M. (2017). Perceiving crowd attention: Gaze following in human crowds with conflicting cues. *Attention Perception & Psychophysics*, 79(4), 1039–1049.
- Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, 129(2), 426–438.
- Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences*, 116(13), 6025–6034.
- Todd, A. R., & Simpson, A. J. (2016). Anxiety impairs spontaneous perspective calculation: Evidence from a level-1 visual perspective-taking task. *Cognition*, 156, 88–94.
- Todd, A. R., Simpson, A. J., & Cameron, C. D. (2019). Time pressure disrupts level-2, but not level-1, visual perspective calculation: A process-dissociation analysis. *Cognition*, 189, 41–54.
- 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.
- Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, 29(5), 874–880.
- Wu, R., Street, M., Richardson, D. C., Street, G., & Kirkham, N. Z. (2010). Social cues support learning about objects from statistics in infancy. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).
- Yin, J., Ding, X., Zhou, J., Shui, R., Li, X., & Shen, M. (2013). Social grouping: perceptual grouping of objects by cooperative but not competitive relationships in dynamic chase. *Cognition*, 129(1), 194–204.
- Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-based attention on social units: Visual selection of hands performing a social interaction. *Psychological Science*, 29(7), 1040–1048.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.