Shared Gaze in Collaborative Visual Search

Harri Siirtola, Oleg Špakov, Howell Istance & Kari-Jouko Räihä

Published online: 13 Feb 2019.

Submit your article to this journal ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

Shared Gaze in Collaborative Visual Search

Harri Siirtola ⬤, Oleg Špakov, Howell Istance, and Kari-Jouko Räihä

Visual Interaction Research Group (VIRG), TAUCHI Research Center, Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Tampere, Finland

ABSTRACT

Collaboration improves efficiency, avoids duplication of efforts, improves goal-awareness, and makes working generally more pleasurable. While collaboration is desirable, it introduces additional costs because of the required coordination. In this article, we study how visual search is affected by gaze-sharing collaboration. There is evidence that pairs of visual searchers using gaze-only sharing are more efficient than single searchers. We extend this result by investigating if groups of three searchers are more efficient, and if and how the groups of searchers develop their search strategy. We conducted an experiment to understand how the collaboration develops when groups of one to three participants perform a visual search task by collaborating with shared gaze. The task was to state if the target was present among distractors. Our results show that users are able to develop an efficient search and division-of-labor strategy when the only collaboration method is gaze-sharing.

1. Introduction

One common human task is *visual search*, or an active scan of the visual environment for a particular object (the target) among other objects (the distractors). Typical examples are detecting a gun in x-ray images of luggage, scanning tumors in mammograms, or locating your car when you can't quite remember where you parked it. Visual search has been studied extensively for the past 40 years (see Wolfe and Pashler (1998); Wolfe (1994) for reviews).

One important flavor of visual search is the act of searching together, or *collaborative search*. Good examples of this are children looking for the next piece in a puzzle, a couple in a car trying to locate a certain street, or an airborne rescue mission team trying to locate a missing person. Our focus in this article is to study what benefits shared gaze could provide in collaborative search.

Previous controlled studies on the usefulness of shared gaze have found that people are able to make use of the gaze information of their search partner. Search times have been reduced, sometimes almost to half of what individual searchers would take, without increasing the error rate. Interestingly, gaze has proved to be an even better collaboration facilitator than sound, because the collaboration overhead is smaller (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008).

However, previous studies have stopped at the case of two searchers (e.g., Niehorster, Cornelissen, Hooge, and Holmqvist (2017); Schneider and Pea (2013); Brennan et al. (2008); Malcolmson, Reynolds, and Smilek (2007)). It is interesting to explore whether the benefit continues to improve when more searchers are involved. On the one hand, one could expect an improvement in search time, but does the error rate remain reasonable? Moreover, indicating the gaze points of several searchers could be distracting as the screen gets busier. Furthermore, previous work has reported only the effect on search times and error rates. No in-depth analysis of the strategies used by the searchers has been made, nor how the strategies evolved during the experiment.

We report an experiment where groups of searchers saw each other's gaze pointer, but had no other means of communication. The questions we are addressing are as follows:

(1) Is a group of three visual searchers that share only their gaze better than a group of two searchers?
(2) Is it possible to develop a collaborative search strategy with shared gaze only?

We present results from an experiment where the task was to find a letter O (the target) among letter Qs (the distractors) on computer screen. This is similar to condition "Shared Gaze" in Brennan et al. (2008). The baseline in our study was a single searcher which is compared to groups of two and three searchers.

Our primary approach to the analysis of the results is data exploration: to look deeply into the data and explain why the data behaves the way it does (Wickham & Grolemund, 2016, advocated in Chapter 1). Collecting gaze data with detailed event logs allows us to unfold the trials in great detail, and then employ an information visualization approach to draw conclusions from the data. This makes it possible to analyze, for example, how the teams divided the search task, and what

was the impact of errors made by one searcher on continued collaboration.

2. Related work

The benefit of collaboration is a complex issue, as the classic review by Hill (1982) demonstrated. The group performance is generally better than the performance of the average individual, both qualitatively and quantitatively, but a group is often outperformed by the best individual. The result depends on the complexity of the issue at hand, and on how heterogeneous the performance levels of the individuals in the group are.

Why does not increasing the number of people carrying out a task always help, or at least not help in proportion to growth in group size? Previous research has identified several reasons for process losses, particularly concerning communication. One often discovered reason is *evaluation apprehension* (Diehl & Stroebe, 1987; Lamm & Trommsdorff, 1973): subjects working in a group may fear that their suggestions are evaluated negatively by other group members.

Of particular interest to us are studies that make use of computer-mediated communication, that is, where the members of the group are not colocated. This brings along new productivity hampering factors, or at least emphasizes them. *Social loafing* refers to some participants not doing their share of the work. Chidambaram and Tung (2005) distinguish two reasons for this: the *dilution effect*, where a person feels less motivated to participate because of more dominant or strong group members, and the *immediacy gap*, caused by an increased distance between group members. Related to social loafing is the *identifiability* of group members. Kraut (2003) suggested that anonymity of group members should not be allowed, as it works as an incentive for social loafing. However, Suleiman and Watson (2008) did not find any effect of identifiability in their experiment on social loafing. On the other hand, while computer-mediated communication may thus invite social loafing, it also has positive effects on group work. In particular, Valacich, George, Nunamaker, and Vogel (1994) found that it decreases evaluation apprehension.

Group size is a factor that affects social loafing in complex ways. Alnuaimi, Robert, and Maruping (2010) identified three cognitive elements that modulate the effect of group size on social loafing: diffusion of responsibility (accountability of transferring work outputs to others in the team), dehumanization (the denial of qualities associated with meaning, interest, and compassion), and attribution of blame (blaming the recipients of the antisocial behavior for bringing suffering upon themselves). They found that increasing group size has an increasing effect on all these three mechanisms of moral disengagement, thereby increasing social loafing. However, another study (Lowry, Roberts, Romano, Cheney, & Hightower, 2006) compared groups working through computer-mediated communication to colocated groups, and found that the negative impact of increased group size was smaller in remotely located groups.

In a fairly recent experiment, Schneider and Pea (2013) studied new ways of supporting joint attention to help students to learn more effectively, a specific area of group work. They argue that in proper joint attention the participants need to show that they are attending something in common, and they used gaze-sharing to accomplish this. In their experiment, pairs of students studied a static diagram explaining the visual information processing in the human brain. The participants were looking at the static diagram in different rooms, had audio connection, and either saw or not each other's gaze indicator. They had to collaboratively explain the model in the diagram. The results indicate that the students in the gaze-sharing condition were able to achieve a higher quality of collaboration and a higher learning gain. This was particularly the case for "followers," who were able to benefit from following the gaze of "leaders," that is, the more dominant members of the pairs. This distinction was based on the driver/passenger roles in co-operating pairs suggested by Shaer et al. (2011). For leaders, sharing the gaze did not produce any negative effects for collaboration.

Most of the previous research on collaboration has been carried out with complex tasks, such as brainstorming, idea generation, and making difficult management-related decisions. Visual search has a different flavor. In particular, the metrics of process efficiency are different: in idea generation, the number of distinct ideas produced is typically the main metric, whereas in visual search the number of correctly found items and the time taken for the task are the main metrics of interest. Moreover, while in general studies of group work "small" groups have typically consisted of three to four members and been compared to "large" groups of 6–12 members or more, collaborative visual search studies have typically stopped at comparing individual searchers to dyads (pairs of searchers). Nevertheless, it is interesting to study how much of the general findings of group work transfers to this context.

Probably the first study on collaborative visual search was carried out by Malcolmson et al. (2007). They report two experiments of how collaboration influences visual search performance. Individuals working together ("collaborative pairs") were compared to the pooled responses of the individuals working alone ("nominal pairs"). The task was to find a "backward C" amongst distractor items that had gaps on both sides from a display that was shown for a limited time (300 ms, 400 ms, or 500 ms). The collaborative pairs were encouraged to discuss how they might best perform the task together, and for every 40 trials the participants were instructed to evaluate their strategy, and discuss whether they want to change it. Participants were sitting in front of the same computer and indicated the presence of target by pressing the space bar (and apparently had a chance to communicate verbally during the experiment, although the article does not report this). The target was present in half of the trials. They found that collaborative pairs had fewer hits than the nominal pairs, but also had much lower proportion of false alarms. Together, this means that the collaborative pairs showed greater sensitivity for detecting the target, and had also more conservative response bias, that is, they needed more evidence to make a decision.

In general studies of group work, remotely located groups presented both disadvantages and advantages compared to colocated groups. What is the case in visual search? In an influential study, Brennan et al. (2008) studied visual search of remotely located pairs. In particular, they studied the effect of different forms of remote communication. Their experimental task was the same we use in this study (find "O" in "Qs"), but they had different collaboration conditions: searching alone, as pairs (shared gaze by seeing the collaborator's gaze-cursor),

shared-voice (by speaking to each other), and shared-gaze-plus-voice. They found that the shared gaze condition was the best of all, twice as fast and efficient as single search. Their main conclusion was that people can successfully communicate and coordinate their search using shared gaze alone.

This finding was corroborated by the experiment of Niehorster et al. (2017), where the objects were images of Gabor filters, and the task was to find the one with a different orientation from the others. Thus, determining the correct choice required comparison of the images, contrary to the setup of Brennan et al. (2008) where the shape of the searched item was known in advance. In the experiment of Niehorster et al. (2017), the images appeared in fixed locations in a grid of hexagons, and the gaze point of a searcher's partner was indicated by highlighting the border of the corresponding hexagon. Niehorster et al. (2017) found that collaborating searchers were able to find the target about twice as fast as individual searchers, that is, they formed an efficient collaboration strategy. When the searchers competed and saw each other's gaze markers, they found the targets even faster.

To the best of our knowledge, the studies by Brennan et al. (2008) and Niehorster et al. (2017) are the only ones that share two important properties: (1) collaboration was facilitated by showing the gaze-pointer of the search partner on the screen, and (2) no consensus was required, that is, as soon as one searcher found the target, he or she could indicate it and the next search task was presented. Both studies found that pairs could work about twice as fast as single searchers, but there was no detailed analysis of what kind of search strategy emerged. Our goal is to investigate the search patterns in more detail.

A follow-up study to the one by Brennan et al. (2008) was carried out by Neider, Chen, Dickinson, Brennan, and Zelinsky (2010). Their experiment was designed to find out how a location in space is referenced under time pressure. They compared speech alone (shared voice), gaze cursors alone (shared gaze), or both. The task was to locate and reach consensus on a sniper target in a pseudo-realistic city scene. In the case where gaze alone was used, consensus was indicated by both searchers fixating on the same target before pressing a physical button. The results were twofold. First, contrary to the finding of Brennan et al. (2008), the time for finding the target did not differ between the conditions. This indicates that the search space was not divided between the participants in this task. Second, they observed that less time was needed for one partner to locate the target after it had been located by the other partner. Thus, gaze-sharing was more efficient than speaking when collaborating on tasks that require rapid communication of spatial information. Similarly in the experiment of Yamani, Neider, Kramer, and McCarley (2017) where consensus was also required, gaze-sharing failed to reduce reaction time and produced a cost to it, contrasting the finding of Brennan et al. (2008) where there was no consensus requirement.

The benefits of gaze in sharing information on an already found search result have been repeatedly observed in later studies that used more abstract tasks, in the spirit of Brennan et al. (2008). In one of the tasks used by Zhang et al. (2017), the searchers had to find an oval object among 364 colored circles. Yamani et al. (2017) performed a two-person visual search-and-consensus experiment

where the task was to locate a randomly set letter T among 87 letter Ls. Both studies showed the benefits of gaze for communication and making deictic references. However, both allowed the use of speech for communication in addition to gaze, and did not attempt to study the effect of gaze alone.

In the many studies that transfer information on the point of gaze to a search partner, the indicator of the gaze point has taken many different forms. The variations include a yellow circle (Brennan et al., 2008; Neider et al., 2010), highlighting a border area (Niehorster et al., 2017), light blue dot (Schneider & Pea, 2013), circled eye icon (D'Angelo & Gergle, 2016), pink cross mark (Yamani et al., 2017), and an eye icon trail (Li, Manavalan, D'Angelo, & Gergle, 2016). The form of the gaze indicator was used as a controlled variable in the first experiment of Zhang et al. (2017). The four conditions were a gaze cursor (rendered as a circle), a gaze trajectory, highlighting a set of objects in the proximity of the gaze point, and a spotlight (highlighting an area around the gaze point, both the background and the objects). They found a trade-off between the prominence of gaze indicators and user satisfaction. Users preferred subtle feedback and found detailed gaze information distracting and less agreeable. Especially the trajectory, even though it provided the most detailed information of the gaze path, was evaluated negatively.

Similar to Zhang et al. (2017), Li et al. (2016) used a condition with shared gaze and voice. They found that gaze information helped the participants in making more deictic references. However, in the beginning they were overwhelmed by the amount of visual information available when gaze was rendered as a trail of the last few fixations. In another condition, a dynamically shrinking zoom focus was used, but then the visualization was considered cognitively demanding. In a related study (D'Angelo & Gergle, 2016), an additional problem was caused by the inaccuracy of the gaze data produced by a low-cost eye tracker. This resulted in the deictic references being sometimes interpreted incorrectly.

In summary, previous research on collaboration has highlighted both advantages in process efficiency and disadvantages in communication, especially evaluation apprehension and social loafing. Results on identifiability of group members have been conflicting. Most studies on group size have compared large groups to smallish groups of four to six members. Remotely located groups have showed some benefits in avoiding evaluation apprehension, but made it easier to engage in social loafing.

In collaborative visual search, not much is known beyond comparing individual searchers to dyads. An interesting finding pertinent to this domain is that sharing the gaze of a collaborating searcher has been found to be the fastest way of communication, faster than using speech. This indicates that searchers have been able to divide the search space efficiently, but past research has not studied how they did it. Exploring this in more detail, moving from dyads to larger groups, and examining whether the general findings on group work carry over to this context are of our interest in this article.

3. Method

3.1. Participants

Twenty-four students from Tampere University participated in the experiment for a reward. They were 16 men and 7 women

(one participant declined to state the gender). Their age ranged from 19 to 46 years, with median of 22 years. Fifteen of the participants had normal vision and nine had corrected-to-normal vision.

The participants were randomly allocated into groups having one, two, or three members, four groups of each size. The mean of age and its standard deviation in each group was ($M = 26.0, SD = 6.78$), ($M = 23.5$, $SD = 5.50$), and ($M = 24.6, SD = 7.57$), respectively. They were asked to give consent to collect their gaze data. It was explained that they would remain completely anonymous, and that they could opt out at any point of the experiment or afterward if they so wished.

3.2. Apparatus

Three comparable eye trackers (two *Tobii T60s*[1] and one *myGaze n* by *Visual Interaction GmbH/SensoMotoric Instruments*[2]) were used to collect participants' gaze data. Both tracker models can provide at least 30 gaze points per second with reported $0.5°$ gaze position accuracy and $0.1°$ spatial resolution. In addition, both tracker models work with most glasses and lenses. Tobii T60 has a 17" TFT display with a 1280×1024 resolution, a Dell display with same specifications was used with the myGaze tracker.

Software for the experiment was developed in-house by utilizing *ETU-Driver*.[3] Each client machine sent its gaze data to a server machine which distributed it to all client machines. TCP/IP network protocol was employed, and visual side-by-side comparison of displays did not show perceivable latency. The inevitable jitter resulting from low-cost eye trackers was smoothed with a two-state low-pass filter (Špakov, 2012). The gaze cursors were simply relocated after each gaze sample, there was no animation between locations.

No chin rest was used as the experiment does not require to indicate the gaze position with high accuracy. However, it was checked after calibration that the gaze cursor was within its own size from the control point (approximately 2º).

3.3. Procedure

The participants were assigned randomly into groups, did not know other group members, and were not allowed to discuss with other group members before the experiment, to prevent the possible development of prior collaboration strategy.

The task was to search for a letter O among letter Qs, all oriented at 0º, 90º, 180º, or 270º angle set in random locations on a computer screen (Figure 1, the screen resolution was 1280×1024). Each letter had a small invisible surrounding area to make sure that the letters did not overlap or touch each other or the borders of the screen, and each letter occupied approximately 0.6º on the screen. Two object set sizes, 21 characters and 35 characters, were used, as in Brennan et al. (2008).

A trial begun when one of the participants in a group pressed a key. Searchers were instructed to indicate the presence or absence of target as quickly and accurately as possible by pressing one of two keys (*enter* = target present, *space bar* = target missing). There was no time limit for the search. A trial ended as soon as any of the participants pressed a key. The screen then showed briefly whether a correct decision had been made by displaying the text 'Correct' or 'Error'. The screen then went blank, and a new trial could be started by any participant.

Before the experiment the members in the group were informed of the rules of the competition. Each participant was awarded a movie ticket for participation. In addition, the participants in the best one-member, two-member, and three-member groups each received a prize of 10 euros. Performance was computed as the total time used in the experiment, with 3 s subtracted for each correct decision, and 6 s added for each incorrect decision (smaller is better). This was as in the Brennan et al. (2008) experiment, and was to motivate the participants to be quick, yet careful.

The group members had no other means of communication than the near-real-time pointers of the other group members' gaze on the screen (Figure 1). The participants did not see their own gaze pointer.

Participants filled afterward a paper questionnaire concerning their opinions about the test setup and their collaboration strategies.

3.4. Design

The experiment was a $3 \times 2 \times 2$ mixed factorial design. The number of group members was between-groups factor, and the within-groups factors were the number of characters (target plus distractors), and the target presence/absence. The dependent variables were the reaction time and outcome. The amount of collected data was 12 groups $\times$ 2 object sets $\times$ 4 orientations $\times$ target absence/presence $\times$ 12 task repetitions = 2304 trials. Each group did 192 trials and encountered the same set of tasks and orientations in randomized order.

We used Statistical System R (R Core Team, 2018) and the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *afex* (Singmann, Bolker, Westfall, & Aust, 2017) to fit a linear mixed effects model and perform an analysis of variance. The fixed effects were the number of group members and the presence of target. As random effects, we had a separate intercept and slope for the number of characters and the subject.

4. Results

4.1. Time

The number of group members ($F_{2,20.01} = 21.86, p < .001$) and the presence of target ($F_{1,2296.50} = 1313.95, p < .001$) both had a significant effect on reaction time. In addition, there was
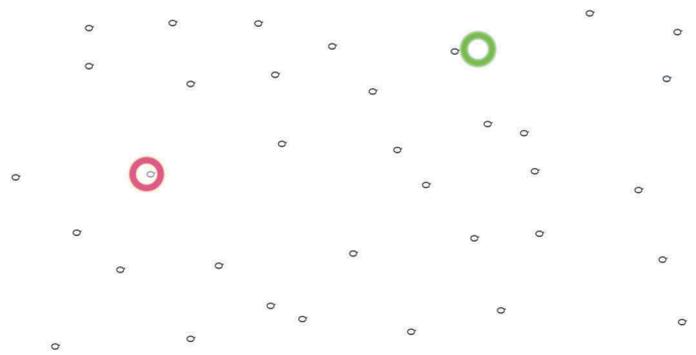

Figure 1. Stimulus with two gaze indicators, that is, a group of three participants is searching for a target "O" among letters "Q" (target not present).

significant non-crossing interaction ($F_{2,2292.91} = 7.73, p < .001$) between them (Figure 2), that is, adding a group member lowered the reaction time more if the target was absent.

Post hoc tests (with Bonferroni correction and Satterthwaite's degrees-of-freedom approximation) showed that the differences between group sizes with the target absent were statistically significant (One–Two: $t(22.63) = -3.173, p < .001; d = 0.57$, Two–Three: $t(27.51) = -3.462, p < .001; d = 0.51$). With the target present, between subsequent group sizes, the difference between group sizes one and two was statistically significant (One–Two: $t(22.12) = -4.893, p < .01; d = 0.89$, Two–Three: $t(26.15) = -1.577, ns$).

The overall reaction times are shown in Table 1, and Figure 8 shows the breakdown into groups and participants, and shows how the decision making was distributed. Figure 3 shows a chart of reaction times in the four combinations of target existence ("Absent," "Present") and group response ("Not found," "Found").

When the target "O" was absent, the response was correct in 99.3% of the trials, and false alarm was reported in 0.7% of cases. When the target was present, the response was correct in 88.7% of the trials, and the "O" was missed in 11.3% of the cases. Figure 3 shows the reaction times of trials side-by-side, with medians and distributions.

Another aspect of the reaction time is how it evolves in a repeated-measures experiment. The reaction time improved monotonically until about trial 75 in all group sizes. After this point, the groups of one and two searchers show an increment (perhaps fatigue), which would eventually fade away. The groups of three searchers improved their performance throughout the experiment.

4.2. Errors

A generalized linear model with logistic regression was used to model the number of errors. The group size ($F_{2,15.56} = 4.04, p < .05$) had a significant effect on the correctness of task outcome, likewise the presence of target ($F_{1,2052.72} = 126.23, p < .001$), and there was also a significant interaction ($F_{2,2008.53} = 11.84, p < .001$). The noncrossing
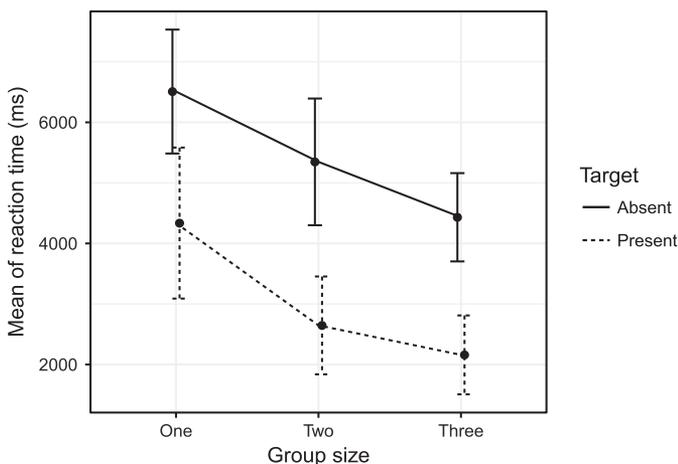
Table 1. Mean reaction times (ms) per the number of group members and outcome. Variation is visualized in Figure 3.

Group size	Target absence confirmed	False alarm	Miss	Target absence confirmed	Mean
One	6508	–	6272	3927	5422
Two	5364	3091	4480	2490	4003
Three	4446	3349	3922	1994	3295
Mean	5444	3252	5262	276	4240
Count	1144	8	130	1020	2304

interaction shows that adding group members improved more when the target was not present.

Post hoc comparisons (with Bonferroni correction and Satterthwaite's degrees-of-freedom approximation) show that only the difference between single and a pair of searchers when the target is present is statistically significant ($z = 3.070, p < .001; d = 0.29$). However, Cohen's d indicates that the effect size is small.

Figure 4 shows the overview of the error counts. Overall, 138 (6.0%) out of 2304 tasks failed, and the outcome was a miss (130, 5.6%) or a false alarm (8, 0.4%). The single searchers did not make errors in the absence of target, and they made approximately twice as many errors compared to collaborative searchers (combined) when the target was present.

4.3. Cooperation between searchers

We collected qualitative feedback with a paper questionnaire that participants filled after experiment, and asked about policies and strategies affecting collaboration formation. This information was then compared with recorded gaze paths and other log data. We concentrate mainly on the group strategies: groups of two or three participants.

The visualizations about search strategies are based on the trials without target (Figures 5–7). If the target is present, the subjects respond as soon as they locate it, but when the target is absent they have to do an exhaustive search before the decision. This allows to better detect if there are any patterns in this behavior. Participants' gaze paths were stored for offline analysis, allowing study of the patterns of collaboration and how they evolved.
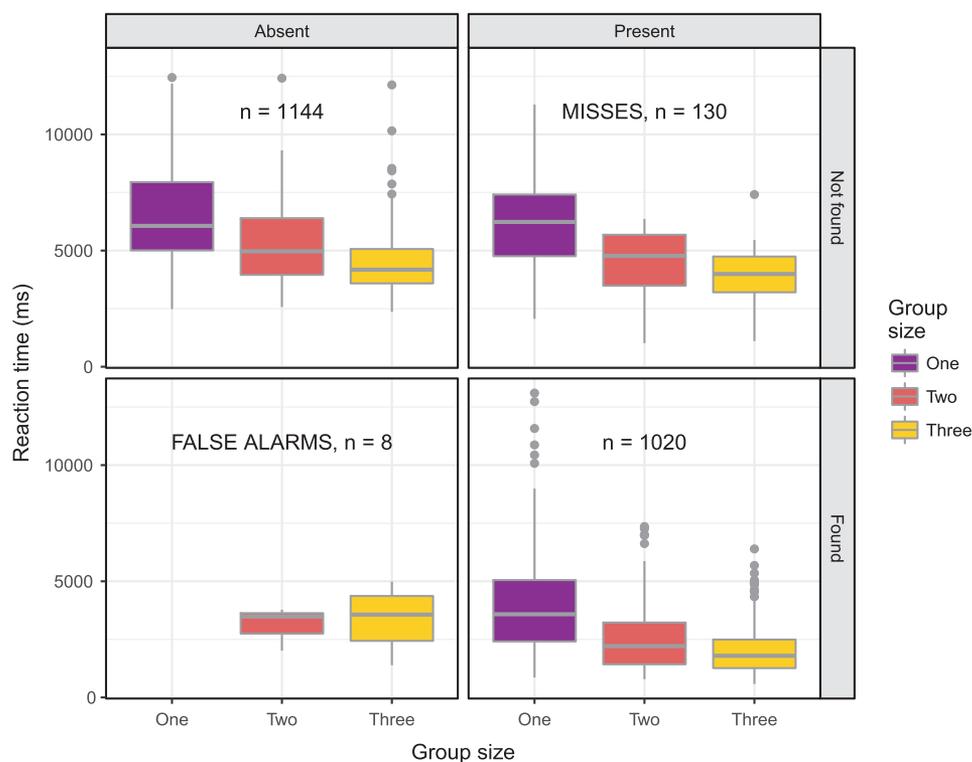
4.3.1. Policy for starting a trial
Eight participants, coming both from the two and three member groups, felt that the decision to start a trial was truly random (Table 2). Eight participants never started a trial, and four participants always started a trial. One group formed a policy that the one who made the previous choice would start the next trial.

4.3.2. Policy for deciding target not present
There were several strategies how to end an exhaustive search (Table 3).

Seven participants replied that they always inspected the whole display before making the decision although they had separate search areas, and one participant replied that covering about 75% of the display was enough for him. Thirteen participants inspected only their part of the display, and waited for a certain circumstance before deciding ("if eyes of others are wandering around," threshold of 1–2 s, or letting the others check the whole display).

Figure 2. Interaction of the number of group members and the presence of target, with within-group standard error bars.

Figure 3. Reaction time distributions in the four combinations of target presence and response. Two observations with reaction time over 18 s were excluded from all charts for readability, but not from the computation.
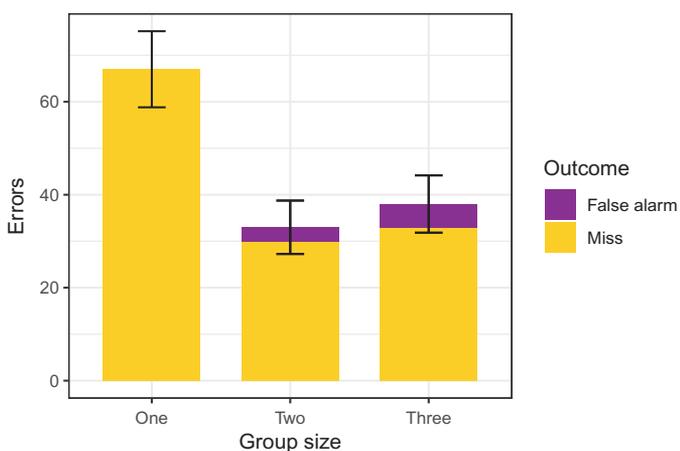
Figure 4. Number of false alarms and misses by group size.

4.3.3. How teams divided the search task

Figure 5 shows how participants verbalized their search strategy, and displays a heat map of each participant's gaze samples. The heat map fuses together all trials from trial number 49, that is, after the first three repetitions of 16 trials. It was observed that, if a search strategy emerged, it was during the four first repetitions, if ever. The heat map was constructed by dividing the screen area into 40×30 pixel rectangles which were colored according to gaze sample numbers, and the alpha channel was adjusted to let colors blend. The heat maps have (hand-drawn) overlays of observed search patterns (or approximations of them). The most popular search pattern was clockwise (11), and counterclockwise (2) and S-shape (2) were also present.

The groups with three members also tried to communicate their strategy to others with gaze. This was done, for example, by always starting from the same location, or by indicating the search area by keeping the gaze position stationary for a while (also in the blank screen, between trials).

All participants in the groups responded that they observed other participants' search order, and over half commented that it did affect their strategy. They consciously tried to avoid the area that their collaborators seemed to cover.

Figure 6 shows a heat map of gaze samples that was created by overlaying trials 49–192. The alpha channel is adjusted in a way that allows the prevailing color to come through. For single searchers the time cutoff is set at 6500 ms, which is the average time to confirm the absence of target (from Table 1), and for the groups of two and three searchers the cutoff is set at the proportion of searchers (single searchers at t, groups of two searchers at $t/2$, and groups of three searchers at $t/3$).

One possible metric for a successful strategy is how long the searchers stay away from the area their fellow searchers are covering, that is, how long do they work in parallel, without overlap. We computed iteratively the convex hulls of the gaze samples of each participant, and an overlap of search areas was observed when the intersection of the convex hulls was nonempty. Figure 7 shows how this parallel working develops throughout the trials.

4.3.4. When did strategies emerge and how these develop

Participants were asked in the questionnaire about their opinion on when did the strategy emerge (Table 4). Six participants
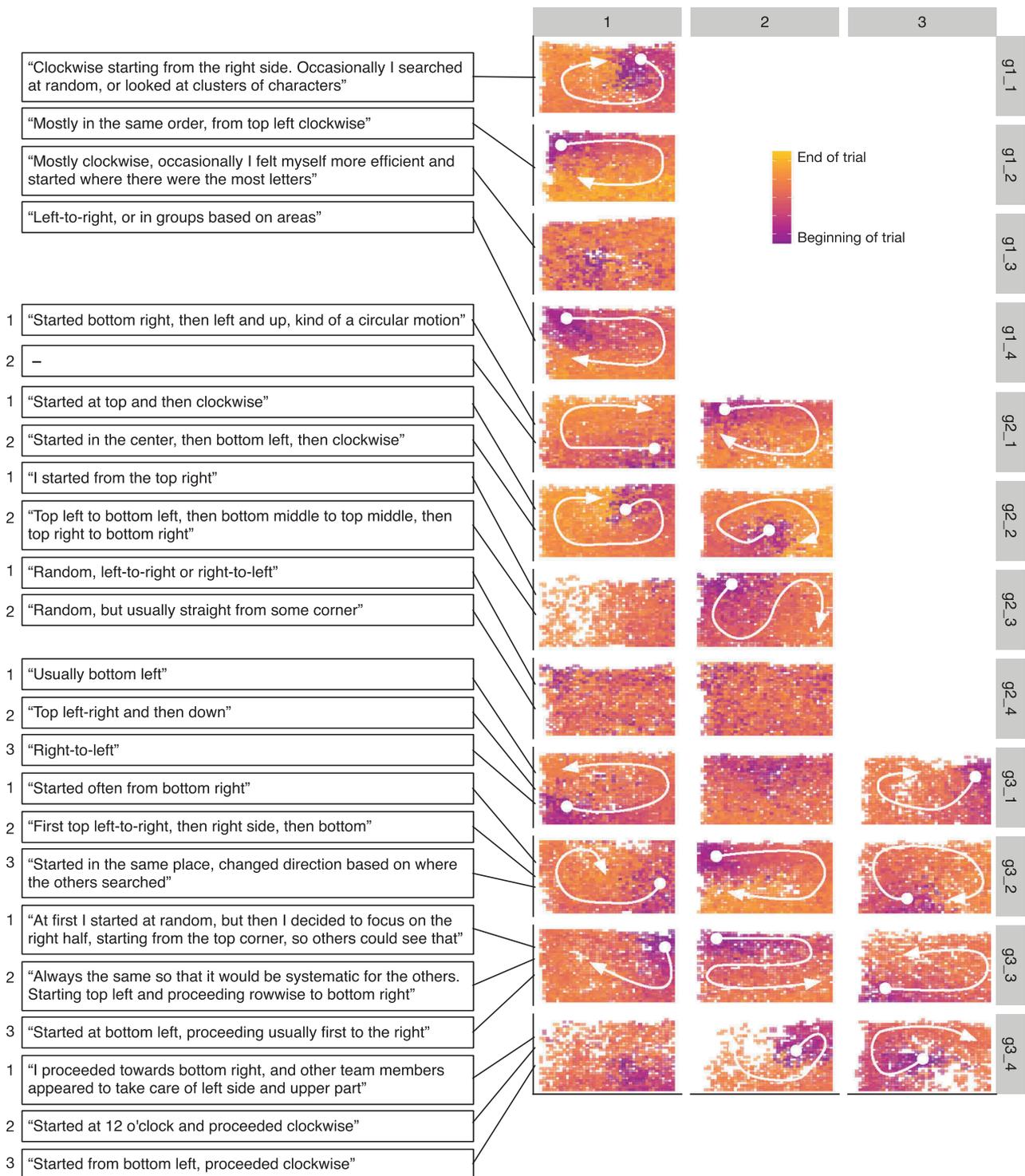
Figure 5. Participants' descriptions of their search patterns and how they appeared in the gaze data. The gaze rectangles (40 × 30 pixels) are color-coded chronologically (from dark blue to yellow), and trials 49–192 are overlaid, revealing possible consistent patterns in the gaze paths. It was assumed that if a strategy would emerge, it would be in place after 48 trials.

replied that the strategy was in place right in the beginning or after a couple of trials, and one participant answered that it took about 10 trials. Another six participants thought that the strategy formation took about 20 trials, and one thought it was about 50 trials. Finally, two participants estimated that the strategy emerged half-way through the experiment (96 trials).

4.3.5. Usefulness of gaze pointers

All participants stated that the gaze pointers were useful, although two participants in both group conditions commented that the pointers were distracting initially and got to be get used to. They attracted your visual attention when it would be more beneficial to look elsewhere.
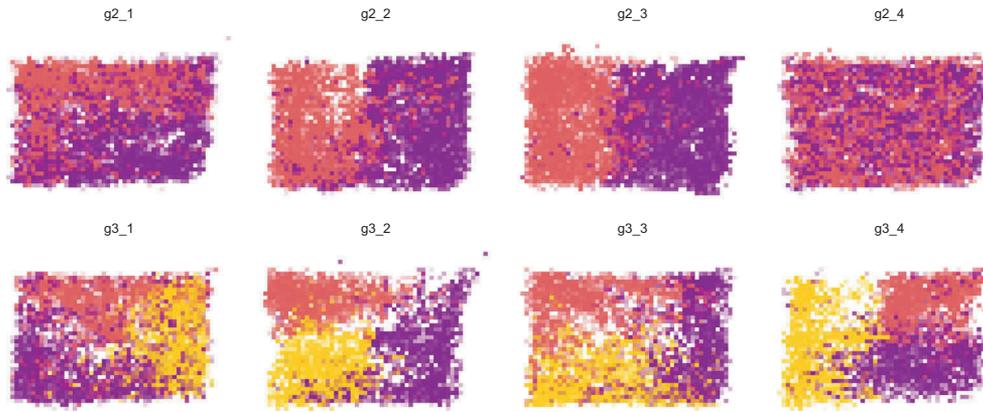
Figure 6. Overlaid heat maps of fixations for groups without target in trials 49–192. The participants in groups estimated that the search strategy emerged after about three rounds (3 × 16), if ever. The color encodes a user, and the number of gaze samples is shown as the intensity of color by adjusting the transparency, that is, the alpha channel. Each heat map is allotted the same amount of search time, by adjusting according to the number of participants: If a single searcher has spent $t = 6500$ms (the mean time from Table 1) on the task, the groups of two searchers have spent $t/2$ per participant, and the groups of three searchers have spent $t/3$ per participant.
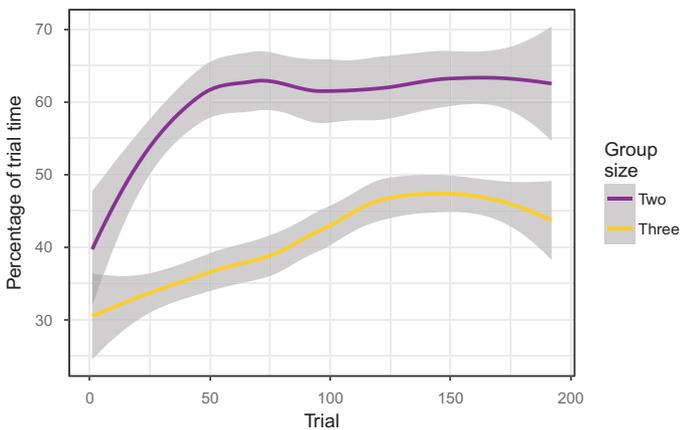
Figure 7. Point in time when the area covered by group of searchers starts to overlap, as percentage of total trial time with loess smoothing and 95% confidence interval. After about 50 trials, the groups of two searchers spend about 60% of time looking at the area not covered by their fellow searcher, while after about 125 trials, the groups of three searchers spend about 45% of time on their own area.

Table 2. Policy for starting a trial.

Strategy	Group size	
	Two	Three
It was random	5	3
I started	1	3
Someone else started	2	6

Table 3. Policy for deciding target not present.

Strategy	Group size	
	Two	Three
Check 75% of the display		1
Check the whole display	4	3
Check own area and decide	2	8
Others' gaze is wandering	2	1

Table 4. When did the strategy emerge.

Strategy emerged	Group size	
	Two	Three
No comment		1
It didn't		1
Right away, kept changing	4	1
"Early"	2	2
Between trials 10–25	2	4
About half-way through (ca. 96)		3

4.3.6. Benefits of a search team

All participants in two- and three-person groups answered that – in their opinion – multiple searchers were better than one. The assumed benefits were increased search speed and the reassuring feeling of being a member of a team. However, one participant in a three-participant group suspected that a team could make more errors than a single searcher.

4.4. What was the impact of making an error on collaboration

Half of the participants replied that they made an unintended choice one or more times during the experiment. The effect of these slips was that participants were more cautious in the following trials, at least for a while ('after an error I waited a little to calm down the decision making'). One of the participants commented that all the errors he made were actually slips.

Single searchers had four incidents of consecutive errors, none longer ones, and the rest were isolated. On the average, every 11th trial resulted in an error, and the longest stretch of successful trials was 44. Three of the four single searchers made slightly more errors toward the end of the experiment.

All groups of two searchers made three consecutive errors and the rest were isolated ones. Three times a false alarm was made. The average number of consecutive successful trials was 24, and the longest stretch without errors was 69 trials. In addition, the performance of the groups remained consistent throughout the experiment.

All groups of three searchers made three consecutive errors and one sequence of four errors, and the rest were isolated

incidents. The average number of consecutive successful trials was 22, and the longest sequence without errors was 63 trials.

4.5. What was the impact of competition

As described in section *Procedure*, the experiment was gamified by subtracting 3 s from successful trial's time, and by adding 6 s to a failed trial's time. The fastest team in each group sizes won movie tickets.

The winning groups were g3_4, g2_3, and g1_1. The bonus and penalty did not have an effect on the winners, but they did affect the order of the other groups of the same size (Table 5).

5. Discussion

Theoretically, if a group of searchers could divide the task optimally, two searchers would do an exhaustive search in half of the time, and three searchers would do it in third of the time (as Brennan et al. (2008) suggested). Table 6 shows a comparison, based on Table 1, where the average time (100%) taken by a single searcher is compared to average times of groups with two and three searchers (single searchers did not make any false alarms). So in this experiment, the overhead in two member groups was $72\% - 50\% = 22\%$, on average, and $61\% - 33\% = 28\%$ in the groups of three members.

This overhead – or coordination cost – is probably due to a superficial check of the other group members' search area. Only one of the two-member groups divided the search area perfectly (group g2_3, Figures 6 and 9), and even they did not do it all the time. For groups of three, the clearest indication of how separate search areas evolved over time is shown in Figure 10.

It can be seen from Figure 6 that the groups of searchers were able to develop a search strategy with shared gaze only. There was only one group (g2_4) which did not

divide the search area, and they even mentioned this in the questionnaire (Figure 5). All the other groups had some kind of spatial division-of-labor strategy which is clearly visible in Figure 6.

Previous studies (D'Angelo & Gergle, 2016; Li et al., 2016) found the rendering of the gaze indicator problematic, either being overwhelming or incorrect. In our experiment, it was sufficient that a searcher stayed aware where the partners were searching. Therefore, the simple fixation-based gaze pointer was sufficient, and accuracy was not critical.

Although there was no consensus requirement for the result, the group members had to develop a *trust* in their fellow searchers – trust that they were covering their own patch. It is known from consensus-requiring experiments that two heads are better than one only if the individuals can accurately communicate their level of confidence to each other (Bahrami et al. 2010; Koriat, 2012). It is not simple to communicate your confidence with gaze only. Only one pair of searchers had a particularly strong trust in each other which shows in their gaze plots (Figure 9). Interestingly, in the post-test questionnaire the members of that group stated that they developed their strategy very quickly, after just a couple of trials, but Figure 9 indicates that developing confidence in the strategy took considerably longer. However, the most common strategy to end an exhaustive search was to check your own patch carefully and then cross-check the other areas superficially until someone made the decision (Figure 5).

Perhaps the most interesting observation in the error patterns of groups is what happened after an error occurred. In the groups of two searchers, the probability for the same participant to make the decision after an error was slightly lowered (by 2%), but after correct decision it was clearly elevated (by 12%). In the groups of three searchers the phenomenon was opposite: the probability for the same participant to answer after an error was slightly elevated (by 4%), but there was practically no difference if the answer was correct. The explanation might lie in peer pressure. In the case of two searchers, the other searcher knows that the partner failed, but in the case of three searchers, the failed searcher remains "anonymous" – there are two possibilities. Although we did not explicitly control the identifiability of group members, this different behavior of dyads and three-member groups lends support to the effect of identifiability (Kraut, 2003).

In many cases, the leader/follower pattern observed by Schneider and Pea (2013) emerged. From Figure 8, especially the case where the target was absent (upper part of Figure 8), there is a trend for one of the searchers in groups g2_1, g2_3, g2_4, g3_1, and g3_2 to be dominant. In all but one of these groups (g2_3) the dominant searcher also started the next trial. As might be expected, the winning groups in Table 5 did not exhibit this behavior, and were able to split the work more evenly. In two of the groups of three searchers (g3_2, g3_4) there was also a clearly submissive participant present, making only 7 out of 96 target-not-present decisions. When the target was present (lower part of Figure 8), the distribution of decision-making was more even.

Table 5. The grand sum of experiment time for each group, and the grand sum with bonus/penalty adjustment. Each correct outcome reduced the time with 3 s and each error increased it with 6 s.

Group	Sum of time	Sum of time with bonus/penalty	Error count
g3_4	541,714	46,714	9
g2_3	638,277	116,277	6
g3_3	629,192	125,192	8
g3_1	599,075	149,075	14
g3_2	761,094	248,094	7
g2_1	753,531	249,531	8
g2_4	738,549	270,549	12
g2_2	944,355	431,355	7
g1_1	954,160	486,160	12
g1_4	957,730	516,730	15
g1_3	995,913	662,913	27
g1_2	1,256,470	797,470	13

Table 6. Reaction times compared to a single searcher.

Group size	Absence confirmed	False alarm	Miss	Presence confirmed	Mean
One	100%	–	100%	100%	100%
Two	82%	NA	71%	63%	72%
Three	68%	NA	63%	51%	61%
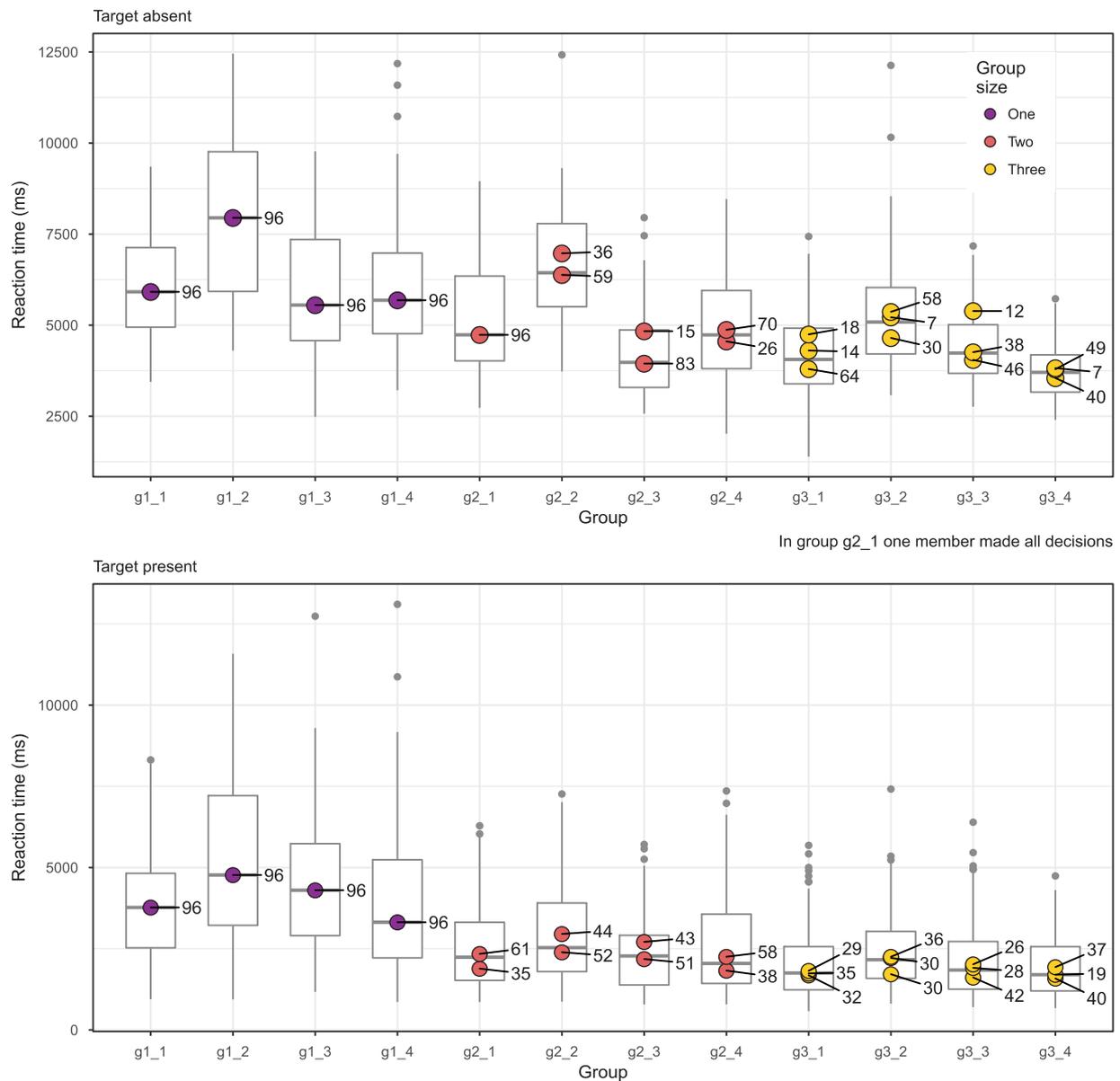
Figure 8. The distribution of reaction time per group, and the number of decisions made by each participant. The box-plot shows each group's distribution, with median, and the dots shows each participant's median of reaction time. Two observations exceeding 18,000 ms were excluded for readability from the figures, but not from the computation of box-plots.

In our experiment, the target was present in half of the trials. The results do not apply to situations like searching for a gun from x-ray image of luggage, as the number of errors is known to increase rapidly if the target appears rarely (Wolfe, Horowitz, & Kenner, 2005).

Figure 7 reveals how the search strategy develops in groups. In the beginning, the searching pairs start to look at each other's areas well before the midpoint of the trial, but after about 50 trials the beginning of gaze overlap starts to stabilize at well over 60% of the trial time. The groups of three participants increase their non-crossing search until trial number 150, after which they spend almost half of their trial time on their own areas.

6. Limitations of the study

We acknowledge that our study has certain limitations concerning our sample and experimental setup that affect the generalizability of the results. We had 24 participants in our experiment (four solitary searchers, four groups of two searchers, and four groups of three searchers), and each group repeated the task 192 times, but still the breadth of the sample is relatively small.

We did not use a control condition with no gaze sharing in our experiment. We took the results of Brennan et al. (2008) as the starting point and focused on shared gaze. However, even without gaze sharing, knowing that another searcher is simulta-

Figure 9. Areas where the fastest group of two searchers (g2_3) looked at as convex hulls enclosing their gaze samples. Trials are in chronological order and target is not present. "Error" indicates the only trial where the group made the wrong decision.

neously carrying out the same task might affect the results, as the image viewing experiment by Richardson et al. (2012) indicates.

Another issue is the stimulus size – the set of 21 characters is fairly low for groups of three searchers, and the benefits would probably be more compelling with sets larger than 35 characters. However, we chose to perform our study with these set sizes to be comparable with earlier results. Finally, our participants were recruited among the students of the university, so our result is limited to this population.

7. Conclusions

To summarize, the results showed that three searchers are better than two, and that it is possible to develop an efficient visual search strategy by shared gaze only. The groups of three searchers outperform the reaction times of groups of two searchers in exhaustive search while the number of errors is about the same. All but one of the eight groups were able to develop a collaborative search strategy by shared gaze only.

The imminent follow-up question is whether there are benefits in growing the group size beyond three. The rectangular search area lends itself to dividing the work in a more obvious way between four searchers than three searchers. We believe that there is a point where the benefit of adding more searchers is consumed by increased coordination costs, and a point where the additional gaze indicators become too distracting. In addition, the optimal group size probably depends on the number of distractors in the search space as well. Answering these questions requires further studies.

Our findings are applicable, for example, to situations where humans need to perform collaborative visual search that is time-critical, and allows only very limited exchange of information between collaborators. A good example of such a circumstance is an emergency room or a situation room type of scenario where a group of people try to quickly solve a problem together. Technologically, gaze-sharing can be enabled by optical head-mounted displays
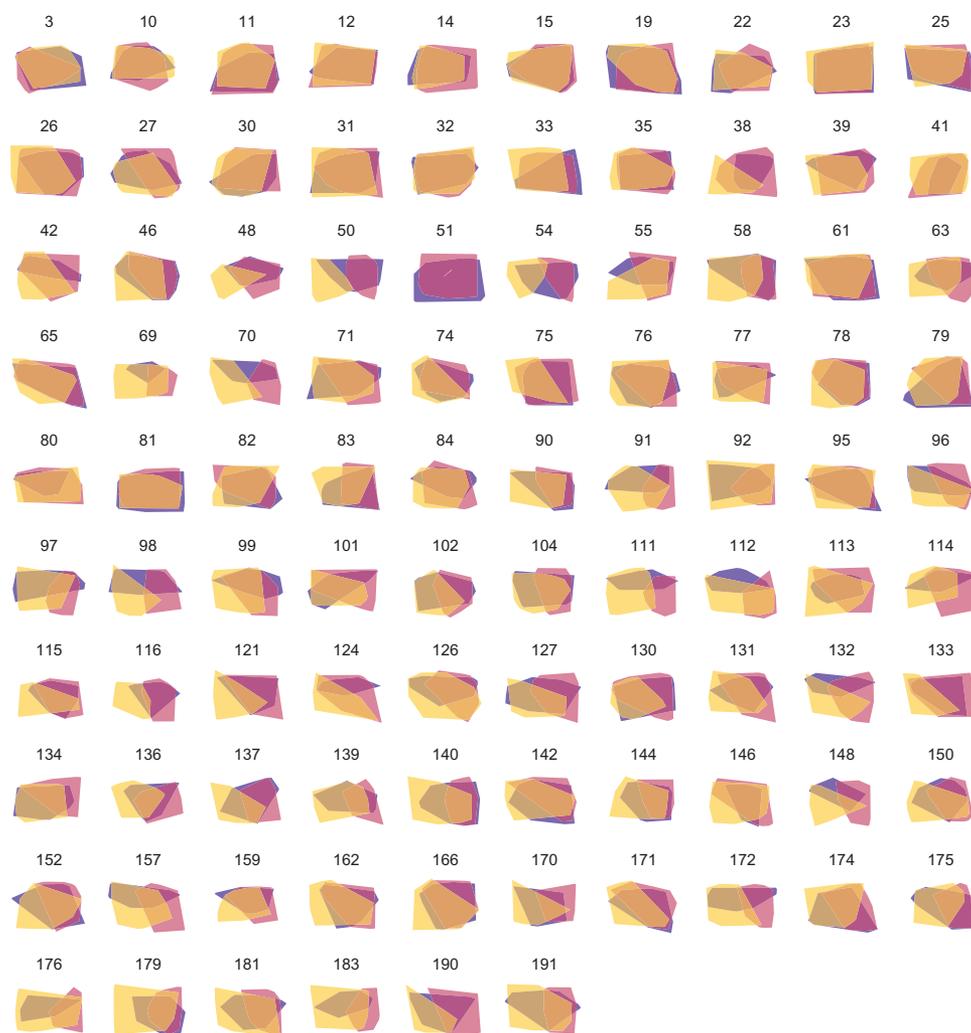
Figure 10. Areas where the fastest group of three searchers (g3_4) looked at as convex hulls enclosing their gaze samples. Trials are in chronological order and target is not present.

with eye trackers. Such devices are already available as augmented reality headsets.

Notes

Acknowledgments

ORCID

Harri Siirtola http://orcid.org/0000-0002-4720-1702

References

Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems*, *27*(1), 203–230. doi:10.2753/MIS0742-1222270109

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329* (5995), 1081–1085. doi:10.1126/science.1185718

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477. doi:10.1016/j.cognition.2007.05.012

Chidambaram, L., & Tung, L. L. (2005). Is out of sight, out of mind? An empirical study of social loafing in technology-supported groups. *Information Systems Research*, *16*(2), 149–168. doi:10.1287/isre.1050.0051

D'Angelo, S., & Gergle, D. (2016). *Gazed and confused: Understanding and designing shared gaze for remote collaboration*. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, (pp. 2492–2496). San Jose, CA.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3), 497. doi:10.1037/0022-3514.53.3.497

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, 91(3), 517. doi:10.1037/0033-2909.91.3.517

Koriat, A. (2012). When are two heads better than one and why? *Science*, 336 (6079), 360–362. doi:10.1126/science.1216549

Kraut, R. E. (2003). Applying social psychological theory to the problems of group work. *HCI models, theories and frameworks*. New York, NY: Morgan Kaufman.

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European Journal of Social Psychology*, 3(4), 361–388. doi:10.1002/(ISSN)1099-0992

Li, J., Manavalan, M., D'Angelo, S., & Gergle, D. (2016). *Designing shared gaze awareness for remote collaboration*. In Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, CSCW '16 Companion, (pp. 325–328). San Francisco, CA. doi:10.1145/2818052.2869097.

Lowry, P. B., Roberts, T. L., Romano, N. C., Jr, Cheney, P. D., & Hightower, R. T. (2006). The impact of group size and social presence on small-group communication: Does computer-mediated communication make a difference? *Small Group Research*, 37(6), 631–661. doi:10.1177/1046496406294322

Malcolmson, K. A., Reynolds, M. G., & Smilek, D. (2007). Collaboration during visual search. *Psychonomic Bulletin & Review*, 14(4), 704–709. doi:10.3758/BF03196825

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5), 718–724. doi:10.3758/PBR.17.5.718

Niehorster, D., Cornelissen, T., Hooge, I., & Holmqvist, K. (2017). Searching with and against each other. *Journal of Vision*, 17(11), 222. doi:10.1167/17.10.222

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org

Richardson, D. C., Street, C. N. H., Tan, J. Y. M., Kirkham, N. Z., Hoover, M. A., & Cavanaugh, A. G. (2012). Joint perception: Gaze and social context. *Frontiers in Human Neuroscience*, 6, 194. doi:10.3389/fnhum.2012.00194

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397. doi:10.1007/s11412-013-9181-4

Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). *Enhancing genomic learning through tabletop interaction*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (pp. 2817–2826). Vancouver, BC: ACM.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). *afex: Analysis of factorial experiments*. R package version 0.18-0, Retrieved from https://CRAN.R-project.org/package=afex

Špakov, O. (2012). *Comparison of eye movement filters used in HCI*. In Proceedings of the Symposium on Eye Tracking Research and Applications, (pp. 281–284). Santa Barbara, CA: ACM.

Suleiman, J., & Watson, R. T. (2008). Social loafing in technology-supported teams. *Computer Supported Cooperative Work (CSCW)*, 17(4), 291–309. doi:10.1007/s10606-008-9075-6

Valacich, J. S., George, J. F., Nunamaker, J. F., Jr, & Vogel, D. R. (1994). Physical proximity effects on computer-mediated group idea generation. *Small Group Research*, 25(1), 83–104. doi:10.1177/1046496494251006

Wickham, H. (2010). *ggplot2: Elegant graphics for data analysis (Use R!)*. New York, NY: Springer.

Wickham, H. (2017). *tidyverse: Easily install and load 'tidyverse' packages*. Retrieved from https://blog.rstudio.org/2016/09/15/tidyverse-1-0-0/

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. New York, NY: O'Reilly Media.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238. doi:10.3758/BF03200774

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: Rare items often missed in visual searches. *Nature*, 435 (7041), 439–440. doi:10.1038/435439a

Wolfe, J. M., & Pashler, H. (1998). Visual search. *Attention*, 1, 13–73.

Yamani, Y., Neider, M. B., Kramer, A. F., & McCarley, J. S. (2017). Characterizing the efficiency of collaborative visual search with systems factorial technology. *Archives of Scientific Psychology*, 5(1), 1. doi:10.1037/arc0000030

Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., & Gellersen, H. (2017). Look together: Using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing*, 21(1), 173–186. doi:10.1007/s00779-016-0969-x

About the Authors

Harri Siirtola earned his PhD in Interactive Technology from University of Tampere in 2007, and he is a Docent of Interactive Technology. He is a member of Visual Interaction Research Group (VIRG) in TAUCHI Research Center, and an active member of the European information visualization community.

Oleg Špakov received his PhD in Interactive Technology from University of Tampere in 2008. His research area is gaze tracking, with a focus on user interaction and attention analysis. He is a co-author of over 40 research articles and number of applications that utilize gaze input.

Howell Istance is a researcher in human-computer interaction specializing in the application of eye tracking for interacting with computers. He was awarded his PhD by Loughborough University in 2017 for work in the area of gaze-assisted interaction techniques for motor-impaired people, particularly concerning their access to games.

Kari-Jouko Räihä is a professor emeritus at Tampere University in Finland. His research area is human-computer interaction. He has mostly worked on eye tracking, both for analyzing interaction and as an interaction technology. He received his PhD in Computer Science at the University of Helsinki in 1982.