



# Joint attention is intact even when visuospatial working memory is occupied

Takemasa Yokoyama<sup>a,\*</sup>, Ryosuke Kato<sup>a,b</sup>, Kazuya Inoue<sup>c</sup>, Yuji Takeda<sup>a,b</sup>

<sup>a</sup> Automotive Human Factors Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

<sup>b</sup> Graduate School of Comprehensive Human Science, University of Tsukuba, Tsukuba, Japan

<sup>c</sup> Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan



## ARTICLE INFO

### Keywords:

Gaze perception  
Joint attention  
Spatial working memory  
Dual task paradigm

## ABSTRACT

The gaze of another modulates and shifts an observer's spatial attention toward this gaze direction. This phenomenon is termed joint attention. Although previous studies found joint attention was induced during a spatial working memory task, it remains an open question whether spatial working memory interferes with joint attention. To address this question, we conducted a dual task paradigm consisting of gaze cuing and spatial working memory tasks. The gaze cuing task requires that participants respond to the location of an abrupt onset of a target while ignoring the gaze of a facial image located at the center of the display. In the spatial working memory task, participants memorized two or three locations of small dots and judged whether these locations changed in the probe screen. In the dual task condition, while participants were maintaining spatial information of the spatial working memory task, they performed the gaze cuing task. In Experiment 1, maintaining spatial information did not impair the magnitude of joint attention. In Experiment 2, we increased load of spatial working memory task, and still observed results similar to Experiment 1. In Experiment 3, we replicated the findings in Experiment 1 when an identification task was conducted. In addition, we measured electrooculographic signals to investigate eye movements of participants during the task. The study provides that joint attention is intact even if spatial working memory is occupied.

## 1. Introduction

The gaze of another individual affords rich signals for one observing this gaze. Indeed, an observer may often rely upon information within this gaze to facilitate social interactions. Through gaze perception, we can recognize, consciously as well as unconsciously, what others favor and where their interest lies (Baron-Cohen, 1995; Langton, Watt, & Bruce, 2000; Yokoyama, Ishibashi, Hongoh, & Kita, 2011; Yokoyama, Noguchi, & Kita, 2013). Thus, in many contexts people routinely pay attention to the gaze of others and use this information in attempting to anticipate a gaze giver's thoughts and desires. This information enables us to respond appropriately to the gaze giver (Baron-Cohen, 1995). Such social ability is important in conducting smooth social interactions.

Gaze perception facilitates social interactions as well as modulation of observer's attention to the direction indicated by the gaze giver. For instance, when a gaze giver looks to the left, spatial attention of the gaze observer is shifted to the gaze direction of the gaze giver (right as observer), which is termed "joint attention" (Driver et al., 1999; Friesen & Kingstone, 1998; Gregory & Jackson, 2017; Marotta, Casagrande, &

Lupianez, 2013). To examine this phenomenon, investigators have used a modified Posner paradigm (i.e. gaze cuing paradigm) (Driver et al., 1999; Friesen & Kingstone, 1998, 2003; Galfano et al., 2012; Langton et al., 2000; Yokoyama, Noguchi, & Kita, 2012). A facial stimulus looking either left or right is placed at the center of the display; following this, a target appears at the left or right peripheral location. Observers can more quickly respond to the upcoming target when it appears in a location toward which the gaze was directed (a congruent target), than when it appears elsewhere (an incongruent target). In addition, this phenomenon occurs even when gaze direction does not predict the target location (Friesen & Kingstone, 1998). Hence, because joint attention is induced despite the absence of cue validity, joint attention seems to be automatic. Furthermore, Hayward and Ristic (2013) has reported that joint attention survives even when verbal information is maintained in working memory, indicating automatic characteristics of joint attention (but see Bobak & Langton, 2015: the executive function can impair joint attention). In this sense, the gaze of others modulates spatial attention of observers even if the direction is not useful for performing a task.

Because joint attention is a primitive function in social interaction,

\* Corresponding author at: Automotive Human Factors Research Center, National Institute of Advanced Industrial Science and Technology, AIST Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan.

E-mail address: [yokoyama.takemasa@aist.go.jp](mailto:yokoyama.takemasa@aist.go.jp) (T. Yokoyama).

<https://doi.org/10.1016/j.visres.2018.10.011>

Received 5 January 2018; Received in revised form 4 October 2018; Accepted 23 October 2018

Available online 16 November 2018

0042-6989/ © 2018 Elsevier Ltd. All rights reserved.

the visual processing of another's gaze appears to be automatic in nature. Although the hypothesis that joint attention has an automatic characteristic seems reasonable, spatial attention also includes voluntary (top-down) processing (Bobak & Langton, 2015). In the Posner paradigm, peripheral cues capture spatial attention through exogenous (automatic) processing, whereas central cues are related to voluntary processing (Jonides, 1981). Moreover, in visual search tasks, inefficient search is associated with voluntary processing, whereas efficient search is considered to be exogenous processing (Wolfe, 1998). Previous studies have indicated that search efficiency in visual search is impaired when a different type of processing consumes resources for spatial representations (e.g., spatial working memory load) (Woodman & Luck, 2004). In short, these previous findings suggest that spatial attention shares common resources with spatial working memory, assuming that load involving spatial working memory processing degrades search efficiency in the visual search task. On the other hand, for joint attention, Law, Langton, and Logie (2010) examined the influence of spatial working memory load on the effect of gaze cuing under a dual task paradigm, and found that the gaze cuing effect can be observed even when the spatial working memory load was given. Their study provided evidence that spatial working memory did not completely eliminate the effect of joint attention, but the magnitude of this effect appears to decrease in the dual task condition relative to that associated with the single task condition, although the statistical analysis of the difference between these conditions did not reach significance. Thus, it is still unclear if joint attention operates independently of spatial working memory. Whether joint attention is resistant to the consumption of spatial working memory resources is a simple but important question, as scarce resources may be a critical cause of failure in social interactions. We investigated whether joint attention and spatial working memory functions with or without common resources for spatial representations in the current study. In addition, the present study includes the confirmation of reproducibility, the statistical analysis of a null-hypothesis significant testing using Bayesian analysis, and the examination of the effect of face and gaze perception.

The aim of this study was to investigate whether spatial working memory processing impairs joint attention. To this end, we used the dual task paradigm and compared the magnitude of joint attention in the single and dual task conditions. The magnitude of joint attention is measured with RTs of incongruent (a target appeared at the opposite gaze direction) minus congruent (target appeared at the same gaze direction) conditions. Larger RT differences imply larger joint attention effects. In our dual task paradigm, we used a working memory task similar to that of Woodman & Luck (2004), however, we used gaze cuing task alternative for their visual search task (Experiment 1). In Experiment 1, we used the dual task paradigm to ascertain whether or not the magnitude of joint attention is impaired by concurrent processing in spatial working memory. In Experiment 2, we heightened the spatial working memory load in order to examine whether joint attention persists in a more severe situation involving attentional processing. In Experiment 3, we conducted an identification task (judgment of the vertical/horizontal orientation of Gabor patches) to qualitatively replicate the findings in Experiments 1 and 2 excluding contamination from the stimulus-response compatibility effect. Also, we measured electrooculographic (EOG) signals to examine eye movements of participants during the task.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Participants

Twenty-three paid volunteers (13 female, age range: 18–28) participated in Experiment 1. They provided written informed consent, and all experimental procedures have been approved by the institutional review board of the National Institute of Advanced Industrial Science

and Technology and adhere to the principles of the Declarations of Helsinki. All participants had normal or corrected-to-normal acuity of vision, and all were naïve to the purposes of this experiment.

#### 2.1.2. Apparatus

Visual stimuli were displayed on an ASUS ROG SWIFT PG258Q LCD display of  $1920 \times 1080$  pixels (refresh rate was 120 Hz). Visual display and data collection were controlled by using the Psychophysics Toolbox of MATLAB (Brainard, 1997; Pelli, 1997) on Microsoft Windows 7. Participants were tested individually in a dark room, and the viewing distance was approximately 57 cm.

#### 2.1.3. Visual stimuli and procedure

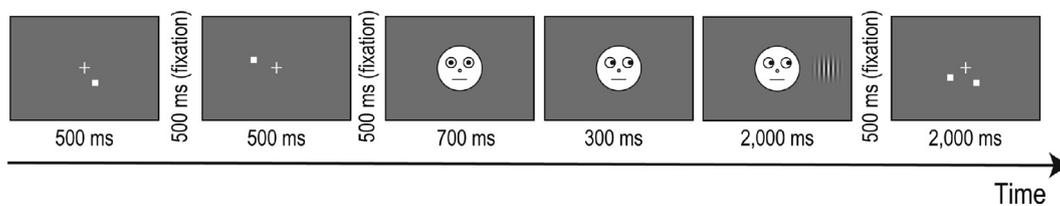
We used a dual task procedure that consisted of gaze-cuing and spatial working memory tasks. The spatial working memory task resembled that of Woodman and Luck (2004), but the gaze-cuing paradigm was used instead of their visual search paradigm. In our study, participants completed a single memory, single gaze-cuing, and dual tasks. The order of these three tasks was counterbalanced across participants. In all tasks, participants were required to repeat “a, b, c, d,” “w, x, y, z,” “1, 2, 3, 4,” or “6, 7, 8, 9” throughout each trial to prevent participants from utilizing their verbal working memory system (Besner, Davies, & Daniels, 1981). Participants were instructed to fixate their eyes to the center of the display during the tasks.

**2.1.3.1. Single memory task.** Following the presentation of a cross fixation in the center of a uniform gray background for  $500 \pm 100$  ms, two white dots ( $0.2^\circ \times 0.2^\circ$ ) were sequentially presented. Duration of each white dot on the screen was 500 ms, and these white dot screens were separated by a fixation screen for 500 ms. The white dots were presented at randomly chosen locations within  $1.7^\circ \times 1.7^\circ$  centered in the middle of the display, and 2000 ms to make the decision. Two blocks of 36 trials were employed, preceded by 12 practice trials.

Each dot was situated at least  $0.50^\circ$  from the center of the other dot's location and from the central cross fixation. After a 4000 ms retention interval, two dots were shown at the same time as a test display. In half of trials, the two dots were presented at the identical locations as the initial presentation. In the remaining trials, one of the two dots appeared at a different location that was located at least  $0.50^\circ$  from its original location, the other dot, and the central cross fixation. Participants made a two-alternative forced choice based on whether or not the location of one of the dots had changed. They used their left hand and pressed z button to indicate same and x button to indicate different, and were given 2000 ms to make the decision. Two blocks of 36 trials were employed, preceded by 12 practice trials.

**2.1.3.2. Single gaze-cuing task.** In the gaze cuing task, we used a schematic face ( $2.3^\circ \times 2.3^\circ$ ). An example of the stimulus is depicted in Fig. 1. A fixation display ( $500 \pm 100$  ms) was presented, followed by a schematic face with direct gaze for 700 ms in the center of the display. Subsequently, the gaze of the face shifted to the left or right for 300 ms, and then a target Gabor patch (size:  $1.4^\circ \times 1.4^\circ$ , spatial frequency: 3.2 c/d, Michelson contrast: 0.625, orientation: vertical) was presented  $5.6^\circ$  to the left or right of the face stimulus for 2000 ms. Gaze direction was non-predictive of the target location, and participants were informed of the manipulation prior to the onset of the experiment. Participants were required to press the 1 (left) or 2 (right) key with their right hand as quickly and accurately as possible to judge the Gabor target location (i.e., a target localization task) while ignoring gaze direction of the face image. Congruent (gaze direction and target location were the same) and incongruent (gaze direction and target location were different) conditions were intermixed in a block. Four blocks of 36 trials were employed, preceded by 12 practice trials.

**2.1.3.3. Dual task.** Participants completed both a memory and a gaze-



**Fig. 1.** An example of the sequence of events in a typical trial in the dual task condition. Two white dots were presented sequentially. While participants were maintaining visual information regarding the locations of the dots, they responded to whether the target Gabor was presented left or right. Thereafter, they answered whether the locations of the two dots were the same as or different to the initial presentation.

cuing task on each trial in the dual task. Fig. 1 shows an example of a trial sequence of the dual task. While participants were maintaining visual information in the memory task, they conducted the gaze cuing task. After the central gaze cue with the target Gabor display were shown, a fixation display was presented for 500 ms, and then the memory test display was presented for 2000 ms. Participants performed the gaze cuing task as quickly and accurately as possible, and they answered the memory task as accurately as possible within 2000 ms. Participants were told to perform the two tasks with equal effort. Four blocks of 36 trials were employed, preceded by 12 practice trials.

2.2. Results

2.2.1. Spatial memory task

Mean accuracy was  $89.8 \pm 1.2$  (SE) % in the single task condition and  $82.6 \pm 1.7$  (SE) % in the dual task condition. A paired *t*-test showed mean accuracy in the single task was significantly higher than that in the dual task ( $t_{22} = 5.53, p < .001, d = 1.01$ ).

2.2.2. Gaze-cuing task

Error trials were removed from further analysis (Total: 1.25%, Single task: 0.90%, Dual task: 1.60%). As in previous studies, we used median for reaction time data to examine gaze-cuing effects as computed for every participant in each condition from the remaining data. (Bobak & Langton, 2015; Burton, Bindemann, Langton, Schweinberger, & Jenkins, 2009; Driver et al., 1999; Ivanoff & Saoud, 2009; Marotta et al., 2013; Ristic, Friesen, & Kingstone, 2002), and Table 1 shows RTs in the gaze cuing task. Fig. 2 shows gaze cuing effects (RTs in the incongruent – congruent conditions) on RTs. We performed 2 *Task type* (single, dual)  $\times$  2 *Congruency* (congruent, incongruent) repeated-measure analysis of variance (ANOVA). We found the main effect of *Task type* [ $F_{(1, 22)} = 290.722, p < .001, \eta_p^2 = 0.93$ ]. We also observed the main effect for *Congruency* [ $F_{(1, 22)} = 10.577, p < .005, \eta_p^2 = 0.325$ ]. However, we did not find an interaction between *Task type* and *Congruency* [ $F_{(1, 22)} = 1.582, p = .221, \eta_p^2 = 0.067$ ]. We compared the gaze cuing effects in the single and dual task conditions. The RT differences between gaze cuing effects in the single and dual conditions did not reach significance [ $t_{22} = 1.25, p = .22, d = -0.33$ ]. However, because this result might stem from a lack of power of the *t*-test, we also conducted a Bayesian analysis of a null-hypothesis significant testing. When the posterior probability for the null hypothesis acquired by this

analysis exceeds 0.75, the evidence is interpreted as positive (in favor of the null hypothesis) (Masson, 2011). The value of the posterior probability ( $p_{BIC}(H_0|D) = 0.78$ ) constitutes positive evidence in favor of the null hypothesis.

2.3. Discussion

We observed decreased accuracy of the spatial working memory task and delayed overall RTs in the gaze-cuing task in the dual task condition, compared with the single task condition. These results indicated that participants divided their resources for spatial representations into two tasks (spatial memory and gaze cuing task) in the dual task condition. Our results also showed that the congruent RTs were significantly faster than the incongruent RTs (i.e., the occurrence of the gaze cuing effect), even though the gaze cue was uninformative because it did not reliably predict the upcoming target location. More importantly, the magnitude of the gaze cuing effects in the dual task condition was statistically comparable with that in the single task condition, which suggests that attentional shifts induced by another’s gaze were intact even if resources of spatial representations were consumed by the different task.

In the next experiment, we added a greater load for spatial working memory (memorization of two locations in Experiment 1; memorization of three locations in Experiment 2). Experiment 2 was designed to confirm the results in Experiment 1 and to examine if joint attention persists even when more spatial processing resources are allocated to the memory task.

3. Experiment 2

3.1. Method

The method in Experiment 2 was identical to Experiment 1, with the exception of the following details.

3.1.1. Participants, stimuli, and procedure

Twenty-three paid volunteers participated (9 female, age range: 20–34) in Experiment 2. In the memory task, one more dot was added after the second dot (500 ms each for the 3rd dot screen and interval fixation screen), and participants were required to memorize three locations simultaneously.

**Table 1**  
Means of (individual) median reaction time (ms) and errors (%) for the gaze cuing task. The parentheses represent standard error of mean.

	Experiment 1		Experiment 2		Experiment 3	
	(2 spat WM load)		(3 spat WM load)		(Identification)	
	Single task	Dual task	Single task	Dual task	Single task	Dual task
Congruent RT	332 (9.4)	570(17.8)	335(10.2)	585(17)	511(14.2)	783(19.1)
% error	0.85(0.24)	1.81(1.07)	0.72(0.19)	0.85(0.38)	7.58(1.34)	2.97(1.15)
Incongruent RT	339(9.7)	584(19)	349(10.1)	596(16.2)	526(14.1)	799(21.1)
% error	0.97(0.20)	1.39(1.06)	1.45(0.63)	0.66(0.35)	9.11(1.45)	4.1(1.26)

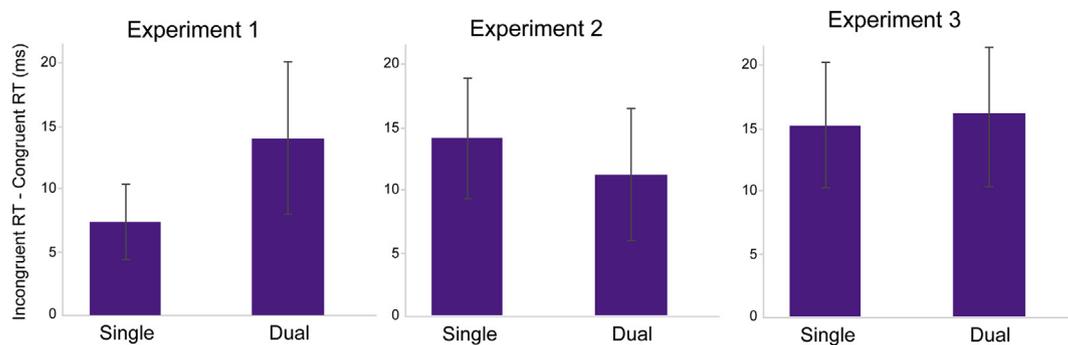


Fig. 2. The results of the magnitude of joint attention (Incongruent RTs – Congruent RTs). We observed joint attention in Experiments 1–3.

### 3.2. Results and discussion

#### 3.2.1. Spatial memory task

Mean accuracy was lower in the dual task ( $82.6 \pm 1.8(\text{SE})\%$ ) compared with the single task ( $86.9 \pm 1.6(\text{SE})\%$ ). We conducted a paired *t*-test and found that accuracy in the single task was significantly higher than that in the dual task ( $t_{22} = 3.41, p < .001, d = 0.51$ ).

#### 3.2.2. Gaze-cuing task

We analyzed median RT after removing error trials (Total: 0.92%, Single task: 1.08%, Dual task: 0.75%). Median RT data were evaluated according to a 2 *Task type* (single, dual)  $\times$  2 *Congruency* (congruent, incongruent) repeated-measures ANOVA. The main effects of *Task type* [ $F_{(1, 22)} = 509.815, p < .001, \eta_p^2 = 0.959$ ] and *Congruency* [ $F_{(1, 22)} = 11.778, p < .005, \eta_p^2 = 0.349$ ] were both robust. Nonetheless, there was no significant interaction between *Task type* and *Congruency* [ $F_{(1, 22)} = 0.422, p = .522, \eta_p^2 = 0.019$ ]. In addition, we analyzed gaze-cuing effects between the single and dual task conditions, and found no significant difference [ $t_{22} = 0.649, p = .522, d = 0.15$ ]. The value of the posterior probability for the null hypothesis ( $p_{\text{BIC}}(\text{H}_0|\text{D})$ ) was 0.85. As in Experiment 1, this constitutes positive evidence in favor of the null hypothesis.

### 3.3. Discussion

Experiment 2 qualitatively replicated the results in Experiment 1. Note that gaze cuing effects persist in a situation where more resources for spatial representations are required. Thus, we can conclude that the gaze cuing effects are robust, and attentional processing is active even when available resources are scarce.

In Experiments 1 and 2, we used the localization task. In the localization task, response (left/right for the Gabor locations) and cue directions (leftward/rightward gaze) were consistent in the congruent condition, so stimulus–response compatibility might have influenced the findings in Experiments 1 and 2. To reject the possibility, we used an identification task in Experiment 3. In addition, we measured EOG signals to examine eye movements of participants during the task.

## 4. Experiment 3

### 4.1. Method

The method in Experiment 3 was identical to Experiment 1, with the exception of the following details.

#### 4.1.1. Participants and procedure

Twenty-four paid volunteers participated (8 female, age range: 20–35) in Experiment 3. We excluded data of two participants from the analysis due to poor performance in the gaze cuing task (below 50% chance performance in the dual task condition). In the gaze-cuing task, we conducted an identification task, and participants judged an

orientation (horizontal/vertical) of the Gabor patch.

#### 4.1.2. EOG recordings and analysis

The EOG signals were recorded with a digital amplifier (BrainProducts, BrainAmp Standard system). A horizontal EOG was recorded using electrodes placed at the left and right outer canthi with AFz as the ground electrode. The time-constant was set at 10 s. EOG signals were digitized at a sampling rate of 1000 Hz, and they were band-pass filtered at 0.1–30 Hz with a second-order Butterworth filter. We measured mean EOG amplitudes in two time periods. One is mean EOG amplitudes ranging from the onset of the gaze cue to the onset of the target presentation (i.e., a cue presentation period). Another is mean EOG amplitudes ranging from the onset of the target presentation to when a participant pressed a key in the gaze-cuing task (i.e., a target presentation period).

### 4.2. Results and discussion

#### 4.2.1. Spatial memory task

Mean accuracy was lower in the dual task ( $82.7 \pm 1.3(\text{SE})\%$ ) compared with the single task ( $86.6 \pm 1.5(\text{SE})\%$ ). We conducted a paired *t*-test and found that accuracy in the single task was significantly higher than that in the dual task ( $t_{21} = 3.60, p < .001, d = 0.58$ ).

#### 4.2.2. Gaze-cuing task

We analyzed median RT after removing error trials (Total: 5.27%, Single task: 6.44%, Dual task: 4.10%). Median RT data were evaluated according to a 2 *Task type* (single, dual)  $\times$  2 *Congruency* (congruent, incongruent) repeated-measures ANOVA. We found a clear main effect of *Task type* [ $F_{(1, 21)} = 177.811, p < .001, \eta_p^2 = 0.894$ ] and *Congruency* [ $F_{(1, 21)} = 20.306, p < .001, \eta_p^2 = 0.492$ ]. There were no significant interaction between *Task type* and *Congruency* [ $F_{(1, 21)} = 0.016, p = .9014, \eta_p^2 = 0.001$ ]. In the direct comparison, we did not find any differences of gaze-cuing effects between the single and dual task conditions [ $t_{21} = 0.125, p = .901, d = -0.04$ ], and the value of the posterior probability for the null hypothesis ( $p_{\text{BIC}}(\text{H}_0|\text{D}) = 0.86$ ) shows the result constitutes positive evidence in favor of the null hypothesis.

#### 4.2.3. EOG

Table 2 shows EOG data in the cue and target presentation periods. For the cue presentation period, mean amplitudes were evaluated according to a 2 *Task type* (single, dual)  $\times$  2 *Cue direction* (left, right) repeated-measures ANOVA. We did not find main effect of *Task type* [ $F_{(1, 21)} = 1.569, p = .2241, \eta_p^2 = 0.067$ ] and *Cue direction* [ $F_{(1, 21)} = 0.946, p = .3419, \eta_p^2 = 0.041$ ]. Moreover, there were no significant interaction between *Task type* and *Cue direction* [ $F_{(1, 21)} = 0.450, p = .5098, \eta_p^2 = 0.020$ ].

For the target presentation period, mean amplitudes were evaluated according to a 2 *Task type* (single, dual)  $\times$  2 *Congruency* (congruent, incongruent)  $\times$  2 *Target location* (left, right) repeated-measures ANOVA. We did not find main effect of *Task type* [ $F_{(1, 21)} = 0.036,$

**Table 2**

Mean EOG amplitudes ( $\mu\text{V}/\text{ms}$ ) of (A) Cue presentation period and (B) Target presentation period. The positive and negative voltages indicated participants' eye movements toward the left and right directions, respectively. The parentheses represent standard error of mean.

A. Cue presentation period(from the onset of the cue to onset of the target)				
	Single task	Dual task		
Cue Left	1.28(1.05)	-1.03(0.96)		
Cue Right	1.53(1.27)	0.13(0.55)		
B. Target presentation period(from the onset of the target to a response)				
	Single task		Single task	
	Congruent	Incongruent	Congruent	Incongruent
Target Left	20.54(4.92)	18.61(4.74)	13.88(3.91)	14.93(4.25)
Target Right	-19.74(4.92)	-19.08(5.77)	-12.91(4.88)	-16.69(4.23)

$p = .8511$ ,  $\eta_p^2 = 0.002$ ] and *Congruency* [ $F_{(1, 21)} = 1.380$ ,  $p = .2533$ ,  $\eta_p^2 = 0.062$ ], but found the main effect of *Target location* [ $F_{(1, 21)} = 16.541$ ,  $p < .001$ ,  $\eta_p^2 = 0.441$ ]. There were no significant interactions between *Task type* and *Congruency* [ $F_{(1, 21)} = 0.086$ ,  $p = .7716$ ,  $\eta_p^2 = 0.004$ ], *Task type* and *Target location* [ $F_{(1, 21)} = 2.124$ ,  $p = .1598$ ,  $\eta_p^2 = 0.092$ ], *Congruency* and *Target location* [ $F_{(1, 21)} = 0.165$ ,  $p = .6889$ ,  $\eta_p^2 = 0.008$ ], and *Task type*, *Congruency*, and *Target location* [ $F_{(1, 21)} = 3.498$ ,  $p = .0754$ ,  $\eta_p^2 = 0.143$ ].

#### 4.3. Discussion

In Experiment 3, we obtained qualitatively similar results in Experiments 1 and 2. Thus, we rejected the possibility that stimulus–response compatibility was confounded in the findings in the previous experiments. In addition, in the cue presentation period, we did not observe differences in EOG signals between gaze cues (leftward/rightward) of the facial image. On the other hand, in the target presentation period, we observed differences in EOG signals between target locations (left/right) while we did not observe the effects of congruency (congruent/incongruent). These EOG results suggest that the gaze-cuing task used in Experiment 3 required overt attentional shifts toward a target in order to identify its orientation (vertical/horizontal), although there was no sign of overt attentional shifts during the cue presentation period. It should be noted that we cannot completely rule out the possibility that participants moved their eyes during the cue presentation period (i.e., involving overt attention) because EOG does not have high accuracy to measure the absolute eye positions. More specifically, in the experiment, EOG signals were amplified by an alternating current amplifier, which distorts direct congruent signals (corresponding to the absolute eye positions). Therefore, it is hard to make a strong claim that attentional shifts induced by the gaze cue occurred without overt eye movements. Furthermore, because we did not measure EOG in Experiments 1 and 2, it is still possible that significant eye movements occurred during the task in those experiments. Investigating the effect of covert and overt processes on joint attention can be an important research topic, although it is beyond the scope of the current study. To clarify the issue, further studies are needed.

#### 5. General discussion

In the present study, we investigated the effects of spatial working memory on joint attention. Across the three experiments, maintaining spatial information in visual working memory did not impair attentional shifts induced by gaze perception.

Joint attention is induced even when an additional spatial working memory task is performed simultaneously, a result that dovetails with previous studies that investigated joint attention during the dual task paradigm (Law et al., 2010). Although previous studies have demonstrated that the effect of joint attention could be observed even when

resources for spatial representations were consumed (Law et al., 2010), it was still unclear whether such resource consumption reduced the effect. Beyond the previous findings, the current study indicates that the magnitude of joint attention remains *intact* even when spatial working memory is occupied.

Since overall RTs in the gaze cuing task were approximately 250 ms slower in the dual task condition than the single task condition, spatial working memory obviously interfered with the gaze cuing task performance. Interestingly, it did not impair joint attention (i.e., facilitation effects induced by gaze cue), implying that the spatial working memory task levied an impact on different processing during the gaze cuing task. The processing in the gaze cuing task can be divided into two phases. The first phase is related to gaze perception including an attentional shift in direction; the second phase is a response to an abrupt onset target appearing at a peripheral location. Because the magnitude of joint attention did not differ between single and dual task conditions, a possible explanation is that spatial working memory impaired the second phase, but not the first phase.

Why is joint attention not impaired by other spatial representations? Two interpretations can be considered. The first interpretation is that joint attention and spatial working memory draw on independent mechanisms, although it has been proposed that spatial working memory shares a common processing resource with some types of an attentional task, such as the visual search task (Woodman & Chun, 2006; Woodman & Luck, 2004). Another interpretation is that joint attention may be controlled in a resource-free or resource-light manner. Joint attention is induced even when gaze direction does not predict target location (Friesen & Kingstone, 1998). Hence, because, without participants' intent, the spatial attention of gaze observers automatically shifts to the direction of a giver's gaze, joint attention transpires automatically. Such an automatic process might be consistent with the interpretation that joint attention is controlled in a resource-free or resource-light manner. That is, joint attention can survive even when resources for spatial representations were consumed by retention of three items in spatial working memory. Because the current study did not directly investigate the relationship between the mechanisms of joint attention and spatial working memory, it is still unclear whether these mechanisms shared the common processes or not. Further studies are needed to clarify this issue.

We previously showed that focused attention is necessary to perceive leftward and rightward gaze (i.e., averted gaze) (Yokoyama, Sakai, Noguchi, & Kita, 2014). In addition, flanker effects caused by averted gaze perception did not occur without focus of attention (Burton et al., 2009). On the other hand, direct gaze is preferentially processed in the brain when it is presented unconsciously (Yokoyama et al., 2013), and perception of direct gaze does not require focused attention (Yokoyama et al., 2014). Hence, our visual systems focused judge that visual inputs of gaze direction are “direct or not” before attentional selection and conscious perception occur. Given that

applying focused attention to averted gaze is necessary for encoding its direction, once averted gaze is processed through focused attention, the triggering of an attentional shift might be obligatory.

Hayward and Ristic (2013) found retention of digit sequence did not decrease gaze cuing effects, while Bobak and Langton (2015) observed generating random numbers reduced the magnitude of the gaze cuing effects. Those two studies manipulated verbal working memory of numbers, but the processes associated with these studies are different. That is, Hayward and Ristic (2013) manipulated the retention, and Bobak and Langton (2015) manipulated the executive function of working memory. Our study manipulated the retention of “spatial” working memory, and revealed that retention of spatial information did not influence joint attention. However, it is still possible that executive processing of spatial working memory influences joint attention as demonstrated in verbal working memory.

In conclusion, we investigated whether spatial working memory interferes with the effects of joint attention. We found that joint attention was not impaired by maintenance of additional spatial representations in working memory. Furthermore, the results are replicated even with an increased load in a spatial working memory task. Here, we report that attentional shifts by gaze perception are automatically induced even during maintaining of spatial information and that additional spatial representations in working memory do not impair joint attention.

#### Declarations of interest

None.

#### Acknowledgement

This study was supported by a grant from the JSPS Grant-in-Aid for Young Scientists [18K18350] to the first author (T.Y.) and from the National Institute of Advanced Industrial Science and Technology to the first author (T.Y.) [AAZ202761C3] and the fourth author (Y.T.) [AAZ202761C1]

#### References

- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, Mass: MIT Press.
- Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning – The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, 33(Nov), 415–437.
- Bobak, A. K., & Langton, S. R. (2015). Working memory load disrupts gaze-cued orienting of attention. *Frontiers in Psychology*, 6, 1258. <https://doi.org/10.3389/fpsyg.2015.01258>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Burton, A. M., Bindemann, M., Langton, S. R., Schweinberger, S. R., & Jenkins, R. (2009). Gaze perception requires focused attention: Evidence from an interference task. *Journal of experimental psychology. Human perception and performance*, 35(1), 108–118. <https://doi.org/10.1037/0096-1523.35.1.108>.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, 6(5), 509–540.
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, 5(3), 490–495.
- Friesen, C. K., & Kingstone, A. (2003). Abrupt onsets and gaze direction cues trigger independent reflexive attentional effects. *Cognition*, 87(1), B1–B10. <https://doi.org/10.1016/S0010027702001816>.
- Galfano, G., Dalmazo, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *Quarterly Journal of Experimental Psychology*, 65(10), 1895–1910. <https://doi.org/10.1080/17470218.2012.663765>.
- Gregory, S. E., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 43(2), 237–249. <https://doi.org/10.1037/xlm0000294>.
- Hayward, D. A., & Ristic, J. (2013). The uniqueness of social attention revisited: Working memory load interferes with endogenous but not social orienting. *Experimental Brain Research*, 231(4), 405–414. <https://doi.org/10.1007/s00221-013-3705-z>.
- Ivanoff, J., & Saoud, W. (2009). Nonattentional effects of nonpredictive central cues. *Attention, Perception and Psychophysics*, 71(4), 872–880. <https://doi.org/10.3758/APP.71.4.872>.
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long, & A. D. Baddeley (Eds.). *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.
- Langton, S. R., Watt, R. J., & Bruce, I. I. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2), 50–59. <https://doi.org/10.1016/S1364661399014369>.
- Law, A. S., Langton, S. R., & Logie, R. H. (2010). Assessing the impact of verbal and visuospatial working memory load on eye-gaze cuing. *Visual Cognition*, 18(10), 1420–1438.
- Marotta, A., Casagrande, M., & Lupianez, J. (2013). Object-based attentional effects in response to eye-gaze and arrow cues. *Acta Psychol (Amst)*, 143(3), 317–321. <https://doi.org/10.1016/j.actpsy.2013.04.006>.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43(3), 679–690. <https://doi.org/10.3758/s13428-010-0049-5>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin and Review*, 9(3), 507–513.
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.). *Attention*. London UK: University College London Press.
- Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, 14(4–8), 808–830. <https://doi.org/10.1080/13506280500197397>.
- Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin and Review*, 11(2), 269–274.
- Yokoyama, T., Ishibashi, K., Hongoh, Y., & Kita, S. (2011). Attentional capture by change in direct gaze. *Perception*, 40(7), 785–797. <https://doi.org/10.1068/P7003>.
- Yokoyama, T., Noguchi, Y., & Kita, S. (2012). Attentional shifts by gaze direction in voluntary orienting: Evidence from a microsaccade study. *Experimental Brain Research*, 223(2), 291–300. <https://doi.org/10.1007/s00221-012-3260-z>.
- Yokoyama, T., Noguchi, Y., & Kita, S. (2013). Unconscious processing of direct gaze: Evidence from an ERP study. *Neuropsychologia*, 51(7), 1161–1168. <https://doi.org/10.1016/j.neuropsychologia.2013.04.002>.
- Yokoyama, T., Sakai, H., Noguchi, Y., & Kita, S. (2014). Perception of direct gaze does not require focus of attention. *Scientific Reports*, 4, 3858. <https://doi.org/10.1038/srep03858>.