



Cognitive Science (2017) 1–33

Copyright © 2017 The Authors. Cognitive Science published by Wiley Periodicals, Inc. on behalf of Cognitive Science Society.

All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12484

What Am I Looking at? Interpreting Dynamic and Static Gaze Displays

Margot van Wermeskerken,^{a,b} Damien Litchfield,^c Tamara van Gog^{a,b}

^a*Department of Education, Utrecht University*

^b*Institute of Psychology, Erasmus University Rotterdam*

^c*Department of Psychology, Edge Hill University*

Received 18 August 2015; received in revised form 28 November 2016; accepted 11 January 2017

Abstract

Displays of eye movements may convey information about cognitive processes but require interpretation. We investigated whether participants were able to interpret displays of their own or others' eye movements. In Experiments 1 and 2, participants observed an image under three different viewing instructions. Then they were shown static or dynamic gaze displays and had to judge whether it was their own or someone else's eye movements and what instruction was reflected. Participants were capable of recognizing the instruction reflected in their own and someone else's gaze display. Instruction recognition was better for dynamic displays, and only this condition yielded above chance performance in recognizing the display as one's own or another person's (Experiments 1 and 2). Experiment 3 revealed that order information in the gaze displays facilitated instruction recognition when transitions between fixated regions distinguish one viewing instruction from another. Implications of these findings are discussed.

Keywords: Eye tracking; Eye movements; Gaze interpretation; Gaze display; Gaze recognition

1. Introduction

Displays of eye movements provide a window into the mind. They show what is at the center of people's visual attention, and they are associated with their cognitive processes such as strategy use (Just & Carpenter, 1985) and insight problem solving (e.g., Grant &

Correspondence should be sent to Margot van Wermeskerken, Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: m.m.vanwermeskerken@uu.nl

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Spivey, 2003; Knoblich, Ohlsson, & Raney, 2001). As Yarbus stated in his classic study: “Eye movements reflect the human thought process; so the observer’s thought may be followed to some extent from the records of eye movements” (Yarbus, 1967, p. 190). In making this association between eye movements and cognitive processes, inferences have to be made on the basis of the display of eye movements. Essentially, this means that one has to assign meaning to a pattern of circles or dots that are overlaid on the original stimulus, either statically or dynamically.

The question of how well cognitive processes can be inferred from patterns of eye movements is important in several areas. For instance, it is important for developing pattern classifiers that are able to deduce an observer’s task based on his/her eye movement pattern (Borji & Itti, 2014; Greene, Liu, & Wolfe, 2012). Such pattern classifiers not only contribute to a richer understanding of what and how task-related cognitive processes are reflected in eye movements (e.g., cognitive relevance hypothesis, see e.g., Ballard & Hayhoe, 2009; Hayhoe & Ballard, 2005), but also are of practical importance as they may lead to technological advances enabling adaptive web search or providing adaptive information or feedback, to name just a few applications, based on a person’s eye movements (see Borji & Itti, 2014). In addition, it is important for research on cognition and learning, where replays of one’s own eye movements are being used to cue verbal reports (Hansen, 1991; Russo, 1979; Van Gog, Paas, Van Merriënboer, & Witte, 2005), or to stimulate reflection and evaluation (Kostons, Van Gog, & Paas, 2009). In addition, replays of other people’s eye movements provide insight into the conditions under which observers can take another person’s perspective and make valid inferences in terms of task or preference based on that person’s eye movements (Foulsham & Lock, 2014; Zelinsky, Peng, & Samaras, 2013). Moreover, in education and training, replays of other people’s eye movements may guide observers to new perspectives on problems and thereby improve their performance (Litchfield & Ball, 2011; Litchfield, Ball, Donovan, Manning, & Crawford, 2010) or may help them make more sense of task demonstrations, thereby improving learning (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). This study investigates whether observers are able to make inferences based on their own and another person’s gaze display in terms of (a) the task associated with the gaze display and (b) whether or not the displayed eye movements are his/her own.

1.1. Interpreting gaze displays in terms of cognitive processes

Yarbus (1967) demonstrated that cognitive processes evoked by different instructions resulted in distinct eye movement patterns when viewing a picture of a painting (i.e., *The Unexpected Visitor* by Ilya Repin). Although his findings were based on only one observer, they have been replicated and extended over the last decade (see DeAngelus & Pelz, 2009; Tatler, Wade, Kwan, Findlay, & Velichovsky, 2010). These studies suggest that eye movement patterns convey information about the task a person was performing while viewing the stimulus. However, despite Yarbus’s optimism and the findings showing that different instructions result in different viewing patterns, it is still largely an open

question to what extent people are actually able to infer what task was being performed from a display of eye movements.

Evidence that observers are able to draw inferences about the nature of the task based on a static or dynamic display of eye movements is scarce and mainly comes from studies using static gaze displays. For instance, Greene et al. (2012) presented participants with 64 grayscale photographs under four instruction conditions (i.e., “Memorize the picture,” “Determine the decade in which the picture was taken,” “Determine how well the people in the picture know each other,” and “Determine the wealth of the people in the picture”; p. 3) for 10 or 60 s. Subsequently, the displays of static eye movement patterns overlaid on those pictures were classified by human observers (Experiment 4) and pattern classifier software (Experiments 1–3). Both failed to predict a person’s task above chance level, which suggests that neither observers nor pattern classifiers were able to infer a person’s task based on a static eye movement pattern. However, some caution is warranted in drawing conclusions from this study. That is, other than concluding that observers were not able to interpret someone else’s static eye movement pattern, an alternative explanation might be that the different instruction conditions used by Greene et al. did not produce distinct gaze patterns, causing poor performance in classifying these gaze patterns (see Borji & Itti, 2014).

Evidence supporting the claim that observers *can* interpret someone else’s static gaze display comes from Zelinsky et al. (2013). In their study, participants performed a categorical search task in which they had to search for either a teddy bear or a butterfly among random category distractors (rated as high, medium, or low in similarity to the target classes). Fixation patterns of these participants were superimposed over the target-absent displays and presented to different participants who attempted to classify the search target reflected in the fixation pattern. Classification performance was well above chance (i.e., 77% and higher, with chance level being 25%). This study provides tentative evidence that observers are able to interpret another person’s *static* eye movement pattern; nevertheless, this study used rather simple static displays (i.e., four targets located at four distinct locations on a monitor) and contained relatively little eye movement data (i.e., fixation data for trials lasting 700–900 ms).

Although most studies have investigated static displays, there is also some evidence that observers are able to interpret another person’s *dynamic* gaze displays (Foulsham & Lock, 2014). In their study, participants had to first determine which fractal out of four within one display they liked most for 18 trials (“truth” block). Then, participants were presented with dynamic gaze displays from the previous participant and had to judge which of the four depicted fractals the previous participant had chosen. Finally, participants received another 18 trials with similar four-fractal-displays within which they were instructed to deceive the next participant by “hiding [their] decision or misleading the guesser” (p. 6; “lie” block; information in square brackets added). Interestingly, the intermediate block, within which the participant had to make a guess about which fractal the previous participant had chosen, included trials from both the “truth” and “lie” blocks. Results indicated that participants were well able to make inferences when the gaze displays stemmed from the “truth” block (i.e., almost 60% correct with chance level being

25%), in which the previous participant was honest about which fractal he or she preferred. Yet, on trials in which the decision was deliberately hidden, participants performed at chance level in deciding which fractal was preferred (i.e., 27%; and worse than in a control condition in which eye movements were not displayed, i.e., almost 40%). Moreover, to successfully hide their preference from subsequent observers, a common strategy was for participants to make their own fixation distributions less specific (i.e., more evenly distributed among the four fractals). This implies that without any formal instruction of how to (mis)interpret these gaze displays, participants tried to minimize any distinctive markers that would help other observers interpret such gaze displays. Different task instructions should yield distinct gaze patterns (DeAngelus & Pelz, 2009; Yarbus, 1967), and so presumably, the more distinctive the gaze pattern, the more readily it is recognized. The question still remains, however, as to whether observers are better at making inferences about static or dynamic gaze displays.

1.2. *Distinguishing your own from another's gaze displays*

In the previous studies (Foulsham & Lock, 2014; Greene et al., 2012; Zelinsky et al., 2013), participants were asked to classify static or dynamic displays of *other people's* eye movements. However, a further open question is whether participants would be able to recognize whether a gaze display is their own or someone else's and whether they would be better at classifying the instructions when a display is their own. Another open question that is explored in this study is whether seeing dynamic replays of one's own eye movements would yield higher recognition rates than static displays. The temporal aspect of dynamic replays may make them easier to follow and distinguish from other eye movement patterns. Moreover, re-viewing their own eye movements in the temporal order in which they occurred could trigger participants' memory of their prior viewing behavior, as suggested by recent findings showing that participants are able to (consciously) monitor their viewing behavior after having performed a serial search task (Marti, Bayet, & Deheane, 2015). In addition, participants who are asked to memorize images and are then presented with the image a second time, tend to refixate the regions they have also fixated during encoding (e.g., Foulsham & Kingstone, 2013a; Foulsham & Underwood, 2008; Valuch, Becker, & Ansorge, 2013). Because dynamic replays inherently provide spatiotemporal information, they can be expected to facilitate this re-enactment compared to static displays of eye movements, and thereby facilitate recognition of the scanpath as one's own or someone else's and recognition of the instruction it reflects.

One study that provided anecdotal evidence wherein participants are *not* always aware that a presented dynamic replay of eye movements was not their own comes from Hansen (1991). In his experiment, participants were trained to use a text editor program, after which they had to perform several subtasks while their eyes were being tracked. Subsequently, participants were shown a video recording of their task performance with or without a replay of eye movements and were asked to explain their problem solving (i.e., cued retrospective reports vs. retrospective reports). To explore the validity of the cued retrospective reports, two participants were presented with a replay of someone else's eye

movements displayed over a video of the performed task (on a task they also had performed themselves), while they were under the impression they would report on their own task performance. Interestingly, these participants were not told that the replay might be from someone else and were not instructed to determine whether it was their own or someone else's. Nevertheless, one of these two participants discovered the deceit and indicated that the replay was not his/hers.

A more recent study did show that observers are able to distinguish between their own and someone else's static gaze display (Foulsham & Kingstone, 2013b). In this experiment, participants were explicitly instructed to make a choice about which of two statically displayed patterns was their own. Findings showed that observers performed above chance. Yet, although significant, performance was only slightly above chance (i.e., 54%) and the direct on-screen comparison of one's own and someone else's eye movement patterns might have facilitated recognition. Hence, the question remains as to whether observers would be able to recognize whether a single gaze display is their own or someone else's without having a direct on-screen comparison. It is also of interest to investigate whether there is a difference between static and dynamic displays in terms of recognizing whether one's own or someone else's eye movements are displayed and which instructional task the eye movements reflect. In one of the few methodological papers on gaze replays, Nalanagula, Greenstein, and Gramopadhye (2006) demonstrated that dynamic replays led to superior performance in knowledge transfer compared to static gaze displays. Although Nalanagula et al. focused on performance improvements from gaze replay modeling rather than gaze interpretation per se, in terms of this study, we can expect that dynamic replays would facilitate interpretation, as they contain information with respect to the temporal sequence of eye movements.

2. Experiment 1

This first experiment investigated whether participants, presented with either dynamic or static gaze displays overlaid on the original stimulus, were (a) able to meaningfully interpret these gaze displays and (b) able to recognize whether a gaze display was their own or someone else's. Participants were presented with Ilya Repin's painting *The Unexpected Visitor* under three different instruction conditions stemming from the original work of Yarbus (1967), two of which ("estimate the ages of the people in the painting" and "estimate how long the unexpected visitor had been away from the family") involved much focus on the faces of the depicted individuals, while a third required hardly any focus on people ("remember the positions of the objects in the room"; see also DeAngelus & Pelz, 2009). Participants were given 15 s to inspect the image under each of the instructions, as pilot tests indicated that this duration resulted in displays comparable to those used in the original work of Yarbus (1967; note that Yarbus viewing times were 10× longer) and that this was long enough to complete the task. Subsequently, they were presented with either static or dynamic gaze displays overlaid on the original image that were or were not their own, and they had to indicate whose eye movements were shown (own or other) and which instruction they reflected.

First, we explored whether performance on these two tasks was above chance or not. Second, it was hypothesized that dynamic eye movement replays would lead to better performance on both tasks than static displays, as replays contain more information with respect to the temporal sequence of eye movements. Showing not only which areas were fixated, but also in what order, might make it more obvious whether the pattern resembles one's own viewing behavior, as well as which instruction it represents, especially because two of the instructions would result in looking at faces a lot, but only one ("estimate how long the unexpected visitor had been away from the family") would involve a lot of transitions between the various people depicted in the painting.

2.1. Method

2.1.1. Participants

Participants were 44 Dutch undergraduate students (26 males; $M_{\text{age}} = 20.3$, $SD = 1.4$), who took part in the study for course credit or a monetary reward of 5 Euro. Participants were randomly assigned to either the dynamic ($n = 22$) or static ($n = 22$) gaze display condition. Seventeen additional participants (9 males; $M_{\text{age}} = 20.6$, $SD = 1.8$) enrolled in the study but had to be excluded due to technical problems ($n = 7$) or bad calibration measures (i.e., deviations $\geq 0.7^\circ$, $n = 10$). All participants had normal or corrected-to-normal vision.

2.1.2. Materials

2.1.2.1. Stimulus and instructions: The stimulus consisted of a picture of Ilya Repin's painting entitled *The Unexpected Visitor* (cf. Yarbus, 1967). The stimulus image and instruction conditions were presented, using SMI Experiment Center software (version 3.3; SensoMotoric Instruments GmbH, Teltow, Germany). The image covered $1,087 \times 1,050$ pixels and was presented centrally on a monitor with a resolution of $1,680 \times 1,050$ pixels (22" monitor; 48×30 cm; angular subtense of $\pm 43.6^\circ$). Participants were first given 15 s to freely examine the painting. Subsequently, they received the following instructions in random order and were given 15 s of viewing time per instruction: (a) estimate the ages of the people in the painting ("ages"), (b) remember the positions of the objects in the room ("objects"), and (c) estimate how long the unexpected visitor had been away from the family ("away"). In between the instructions, the painting disappeared and a fixation cross was presented in the middle of the screen for an accumulated dwell time of 1 s in order to assure that participants always started looking from the same location for each instruction condition.

2.1.2.2. Gaze displays: Participants' eye movements were recorded while they viewed the painting under the different instructions, binocularly at 250 Hz, using an SMI RED250 eye tracker (SensoMotoric Instruments GmbH, Teltow, Germany). From these records, both static and dynamic gaze displays were generated, using SMI BeGaze software (Version 3.2; SensoMotoric Instruments) with gaze displays being overlaid on the original stimulus. The static displays of eye movements were generated using the "Scan Path" utility in BeGaze with fixations defined as lasting at least 100 ms with a maximal

dispersion of 50 pixels (cf., Litchfield et al., 2010; Litchfield & Ball, 2011; a low-speed event detection algorithm was used to enhance generalizability to other eye-tracking systems with lower sampling frequencies). Fixations were displayed by yellow circles with a diameter of 48 pixels and a line width of 4 pixels with consecutive fixations being connected by a yellow solid line (4 pixels). Dynamic replays of eye movements were generated using the “Bee Swarm” option in BeGaze, with raw eye movement data being displayed by a yellow circle with a diameter of 48 pixels and a line width of 6 pixels. This type of display was chosen as the dynamic version of the “Scan Path” resulted in staccato eye movements that were more difficult to follow with the eyes as compared to the “Bee Swarm.” Frame rate of the output for the dynamic displays was set to 25 Hz.

For the “other” eye movement records, one of the authors acted as a model prior to the experiment, observing the painting under the same three instruction conditions and two additional instructions (again, the same starting location was assured for each instruction condition so that differences between *own* vs. *other* gaze displays could not be derived from the starting location). Static and dynamic gaze displays were generated in a similar way as the participants’ gaze displays (see Fig. 1 for a sample of static displays). The two additional instructions were also based on Yarbus’s study (i.e., “surmise what the family had been doing before arrival of the unexpected visitor” and “estimate the material circumstances of the family in the picture”). These displays were included in the recognition test as fillers in order to decrease the likelihood of guessing correctly (see the next section, “Recognition Test”).

2.1.2.3. Recognition test: The recognition test consisted of 11 gaze displays in total. Participants saw their own eye movements under each instruction condition twice, plus the “other” gaze displays: three displays corresponding to the three instruction conditions participants had seen and two filler displays that did not match the instruction conditions participants had seen. The displays were presented in random order on the computer screen, using E-Prime (Version 2.0; Psychology Software Tools, Inc., Sharpsburg, PA, USA). Each gaze display was presented for a maximum of 15 s while the participants determined (a) whether the current gaze display reflected their own eye movements or those of another person (i.e., own/other recognition) and (b) which instruction was reflected in the gaze display: (1) estimate the ages, (2) remember the positions of the objects, (3) estimate how long the visitor had been away, or (4) none of these instructions (i.e., instruction recognition). After 15 s, the gaze display area turned black, but the questions remained visible if they had not been answered yet. Questions were presented in sequential order underneath the stimulus, with the instruction recognition task always being presented after the own/other recognition question had been answered. Participants answered the questions by typing the response corresponding to the answering options on the monitor on the numeric keyboard (1 or 2 for the own/other recognition task, 1, 2, 3, or 4 for the instruction recognition task).

2.1.3. Procedure

The experiment was run in individual sessions of approximately 30 min duration. Participants were seated in front of the eye tracker with their head positioned in a chin and

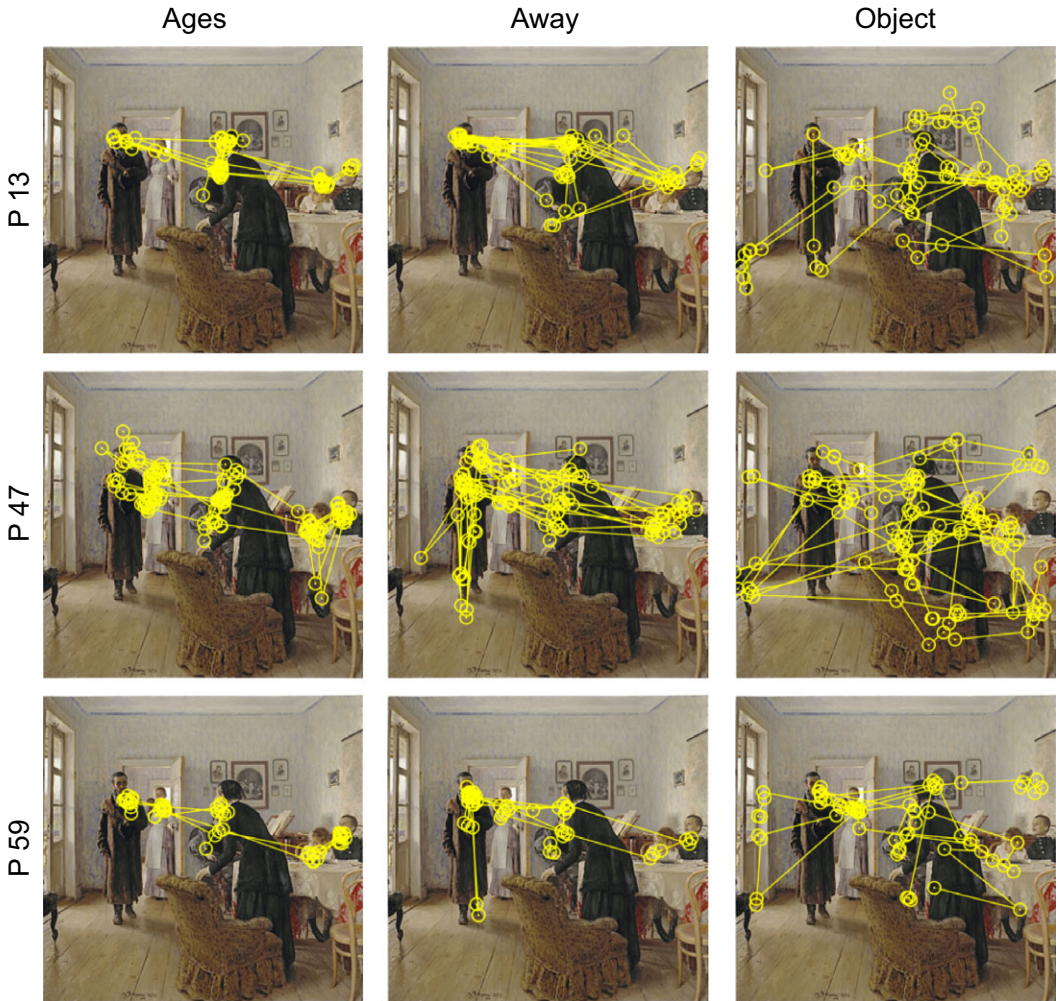


Fig. 1. Representation of static gaze displays of three participants of Experiment 1 in response to each instruction condition. From left to right: “Estimate the ages of the people in the room,” “Estimate how long the unexpected visitor had been away from his family,” and “Remember the positions of the objects in the room.”

forehead rest in order to stabilize the head. Distance to the monitor’s center was approximately 60 cm. First, the system was calibrated using a 5-point calibration plus a 4-point validation (after removal of participants who deviated $\geq 0.7^\circ$: dynamic condition: accuracy $M = 0.38^\circ$, $SD = 0.14^\circ$; static condition: accuracy $M = 0.39^\circ$, $SD = 0.13^\circ$). Then, participants were presented with the stimulus and were allowed to engage in free viewing. Next, they were presented with the same stimulus again three times, accompanied by the three instructions, presented in random order. Subsequently, participants performed a short filler task for 2–3 min (i.e., a puzzle) while the experimenter generated the gaze displays for the recognition test. Prior to performing the recognition test, participants were

presented with a display of eye movements (static or dynamic, depending of condition) overlaid on a different image (i.e., image of Penguins) in order to familiarize them with such a display of eye movements. Successively, participants performed the recognition test.

2.1.4. Data analysis

For the own/other recognition task, the number of correct (and incorrect) judgments was calculated as a function of whether it was a participant's own or someone else's gaze pattern. These numbers were then used to determine d' , which is a signal detection theory measure and reflects the overall recognition sensitivity while taking into account the false alarm rate (i.e., a large d' indicates a relatively high hit rate and a relatively low rate of false positives; see MacMillan, 2002, for formulas).¹ For the instruction recognition task, the total number of correct judgments was calculated as well as the number of correct judgments as a function of instruction type (i.e., age, away, and object).

One-sample t -tests were used to assess whether performance was above chance (i.e., for own/other recognition task: $d' > 0^2$; for instruction recognition task: number correct judgments $>$ chance [i.e., 25%]). Independent t -tests were used to analyze differences in performance between the dynamic and static condition. Note that descriptive statistics for the filler trials are provided (Table 1), but that these trials were excluded from the analyses.

2.2. Results and discussion

For all analyses, we used a significance level of 0.05. Cohen's d and partial eta-squared are reported as a measure of effect size, with $d = 0.20$ and $\eta_p^2 = 0.02$, $d = 0.50$

Table 1

Mean percentage (SD) of correct responses for both recognition tasks per condition and as function of whether the gaze display was a participant's own or someone else's, and performance on each instruction condition in Experiment 1

	Dynamic Condition	Static Condition
Own/other recognition	57.6 (14.8)	55.6 (17.1)
Own gaze	54.5 (22.5)	52.3 (18.0)
Other gaze	63.6 (25.0)	62.1 (23.7)
Fillers	75.0 (33.6)	65.9 (32.3)
d'	0.42 (0.67)	0.32 (0.88)
Instruction recognition	69.7 (16.8)	57.1 (23.3)
Own gaze	72.7 (22.1)	56.1 (27.5)
Other gaze	63.6 (25.0)	59.1 (25.1)
Fillers	34.1 (32.3)	29.5 (36.7)
Age	63.6 (25.0)	53.0 (30.3)
Object	89.4 (18.9)	80.3 (26.5)
Away	56.1 (26.0)	37.9 (31.4)

Note. Chance levels for own/other recognition and instruction recognition was 50% and 25%, resp.; total number of trials for overall performance: $n = 9$; Own gaze: $n = 6$; Other gaze: $n = 3$; Fillers: $n = 2$.

and $\eta_p^2 = 0.13$, and $d = 0.80$ and $\eta_p^2 = 0.26$ corresponding to small, medium, and large effects, respectively. For nonparametric tests, r is reported as an effect size with $r = .10$, $r = .30$, and $r = .50$ denoting small, medium, and large effects, respectively (Cohen, 1988).

2.2.1. Main analysis of performance

An overview of the data is provided in Table 1 (see also Fig. 2). One-sample t -tests revealed that the dynamic condition performed above chance for both recognition tasks (own/other recognition: $t(21) = 2.90$, $p = .009$, $d = 0.63$; instruction recognition: $t(21) = 12.43$, $p < .001$, $d = 2.66$). The static condition only performed above chance for the instruction recognition task, $t(21) = 6.45$, $p < .001$, $d = 1.38$, but not for the own/other recognition task, $t(21) = 1.73$, $p = .099$, $d = .36$ (the Bonferroni adjusted significance level given 2 comparisons per judgment task: $\alpha = 0.025$).

Comparison of the static and dynamic conditions revealed that performance on the own/other recognition task did not differ between the two conditions, $t(42) = 0.39$, $p = .696$; $d = 0.12$, whereas the dynamic condition yielded better performance than the static condition on the instruction recognition task, $t(42) = 2.06$, $p = .046$, $d = 0.63$. In

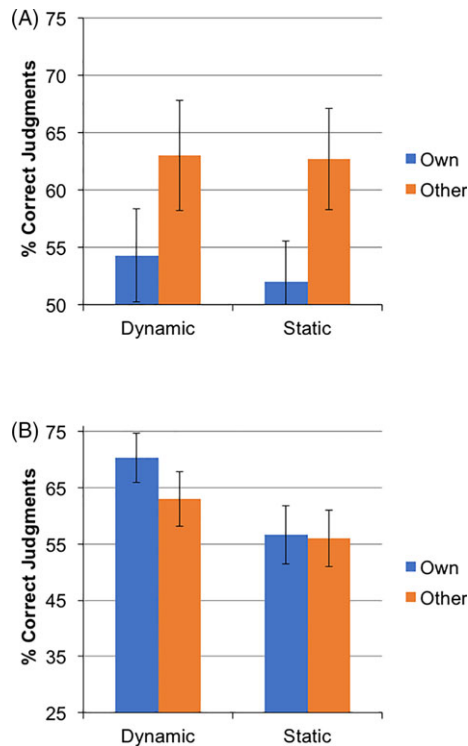


Fig. 2. Percentage of correct judgments in Experiment 1 for (a) *own/other recognition task*, and (b) *instruction recognition task* presented for both conditions and for own ($n = 6$) and other ($n = 3$) gaze trials. The y-intercept represents chance level; error bars represent standard error of the mean.

other words, dynamic replays seem to yield above chance performance on both recognition tasks and seem to enhance recognition of the instruction reflected in the gaze display compared to static displays.

As shown in Table 1, with respect to recognizing a gaze display as someone else's or one's own, participants in both conditions tended to be better in recognizing someone else's as someone else's than at recognizing one's own as one's own (i.e., difference of 9.1%–9.8%). This difference was borderline significant when tested with a 2 (display condition: static or dynamic) \times 2 (gaze: own or other) mixed factor ANOVA with percentage correct on the own/other recognition task as dependent variable: $F(1, 42) = 4.01$, $p = .052$, $\eta_p^2 = 0.087$. However, there was no main effect of condition nor an interaction between display condition \times gaze (both F s < 1). When recognizing instructions from own gaze trials and other gaze trials, it appears that in the dynamic condition, participants were somewhat better in interpreting their own gaze displays than other gaze displays (difference of 9.1%), although this pattern was reversed for the static condition, with other gaze displays yielding better performance than own gaze displays (i.e., difference of 3.0%). A 2 (display condition: static or dynamic) \times 2 (gaze: own or other) mixed ANOVA with percentage correct on the instruction recognition task as the dependent variable did not reveal any effects (all p s $\geq .082$). Note, however, that the number of observations in these analyses was unbalanced in terms of own ($n = 6$) and other ($n = 3$), and that the number of "other" observations was very low. As such, caution is warranted when interpreting these results.

2.2.2. Performance differences among instructions

As mentioned in the Introduction, two of the instructions could be expected to result in many eye movements toward the faces of the depicted people (i.e., "ages" and "away"), whereas one would lead to very little if focus is on the people (i.e., "objects"). One would expect that this would make the "objects" gaze patterns more distinctive than the others, which would make this instruction easier to recognize. To formally assess this, we explored participants' performance for each instruction condition separately. These data show that, indeed, recognition performance on the "objects" instruction was much higher than on the other two instructions (see also Table 1). This was confirmed by a 2 (display condition: static or dynamic) \times 3 (instruction: away, ages or object) mixed-factor ANOVA with the correct percentage as the dependent variable, which revealed a main effect of instruction: $F(2, 84) = 37.40$, $p < .001$, $\eta_p^2 = 0.471$ (note, however, the low number of observations per participant for each instruction condition, $n = 3$). The Bonferroni correction follow-up analyses confirmed that performance was best for the "object" instruction compared to the other two instructions (both p s $< .001$) and that the "ages" instruction yielded a reliable better performance than the "away" instructions ($p = .029$). In addition, a main effect of the condition replicated the finding that the dynamic group outperformed the static group: $F(1, 42) = 4.23$, $p = .046$, $\eta_p^2 = 0.092$. There was no interaction effect ($F < 1$).

2.2.3. Response times

In order to explore whether observers made a decision at first sight or took some time, response times were analyzed.³ As can be seen in Table 2, observers in the dynamic

Table 2

Mean response times (*SD*) per condition and occurrences of response times lasting shorter/longer than 15 s presentation time in Experiment 1

	Dynamic Condition	Static Condition
Mean (<i>SD</i>) response time for own/other task (s)	12.9 (3.7)	8.0 (3.6)
Mean (<i>SD</i>) total response time (s)	16.5 (4.0)	13.2 (3.5)
Occurrences of < 15 s and > 15 s response time for own/other task (% RT > 15 s)	105/93 (47.0)	175/23 (11.6)
Occurrences of < 15 s and > 15 s response time in total (% RT > 15 s)	73/125 (63.1)	134/64 (32.3)

Note. Fillers were excluded.

condition tended to watch the whole dynamic replay of eye movements before deciding on the recognition tasks. Indeed, mean response times in the static condition were shorter for both judgments ($t[42] = 4.45$, $p < .001$, $d = 1.37$ and $t[42] = 2.82$, $p = .007$, $d = 0.87$, respectively), but they still suggest that observers did not respond at first sight, but rather inspected the gaze display at hand.

2.3. Conclusion

The aim of Experiment 1 was to explore whether participants who have just inspected a painting under different instruction conditions would be able to recognize whether or not a display of eye movements was their own and would be able to recognize which instruction was reflected in the display. In line with our hypothesis, the data showed that performance in the dynamic condition was above chance on both recognition test tasks, while the static condition scored only above chance on the instruction recognition task but not on the own/other recognition task. Note that the dynamic display condition did not yield a better performance than the static condition on the own/other recognition task. On the instruction recognition task, however, the dynamic condition did yield a better performance than the static condition. This latter finding was true for all instruction conditions, which underlines the importance of spatial-temporal information, which was inherently provided in the dynamic replays, but not in the static displays. Finally, we showed that the more distinctive gaze patterns, such as those derived from the “object” instruction, were much easier to recognize.

3. Experiment 2

Although the results from Experiment 1 were promising in showing that participants are able to recognize whether or not a gaze replay was their own or not and to meaningfully interpret dynamic and static displays of their own and other person’s eye movements, performance on the own/other recognition task was only slightly above chance and only so for the dynamic condition. One of the questions addressed in Experiment 2,

therefore, was whether recognition performance would improve when participants are informed about the judgment task prior to recording their eye movements. This might make participants monitor their viewing behavior more consciously, which would be expected to result in a stronger memory trace of their own scan paths, which, in turn, would facilitate recognition.

This assumption was based on previous research that showed that participants were able to consciously monitor their own viewing behavior and to recollect this information (Marti et al., 2015). In their study, participants performed serial search tasks, after which they had to introspect on where they had looked on the display. Results suggest that although introspection was neither perfect (i.e., reported fixations were more on targets of the display, whereas real fixations were closer to the center of the screen) nor complete (i.e., about half of the real fixations could be matched to the reported fixations), participants were able to provide valid reports of their eye movements that were closely related, spatially and temporally, to the real eye movements. In addition, participants tend to produce similar scanpaths during recognition as during earlier encoding of the same stimulus (Foulsham & Kingstone, 2013a; Foulsham & Underwood, 2008; Valuch et al., 2013). That is, when instructed to memorize multiple images for later recognition purposes, participants tend to refixate earlier fixated regions during recognition, even if the original image had been mirrored or shifted during recognition (Valuch et al., 2013). Importantly, these similar-looking patterns are also observed when participants are instructed to “attentively examine” the images without the purpose of later recognition (Valuch et al., 2013, p. 4), but to a lesser extent. This might indicate that a stronger memory trace of the scanpath is formed when participants know that recognition is required.

Although this study is very different, as we present only one image under different instruction conditions rather than a vast number of images with one general instruction, and our presentation duration is 3–5 times longer, it can be hypothesized based on these findings that being informed about the subsequent recognition task would result in stronger memory traces of one’s own scanpath. This should improve recognition performance on the own/other recognition task, as well as the recognition of the instruction of one’s own scanpaths. To examine this assumption, participants in Experiment 2 were assigned to either an *informed* condition or a *not informed* condition. We hypothesized that participants in the *informed* condition would outperform participants in the *not informed* condition on the own/other recognition task and—at least for gaze displays that were their own—on the instruction recognition task.

Moreover, Experiment 1 fell short in providing insight into *how* participants performed the judgments with respect to their temporal unfolding, as well as the order in which these judgments were made. After all, in Experiment 1 the questions were asked sequentially with the own/other question always preceding the instruction question. As such, it is unclear whether participants really make separate decisions and, if so, whether they make them in this order, or whether they first recognize the instruction before recognizing whether or not their own eye movements are displayed. In order to get an insight into this matter, a think-aloud protocol was introduced in Experiment 2 (e.g., Ericsson & Simon, 1993; Van Gog et al., 2005). As such, participants were instructed to answer the

recognition questions aloud, as soon as they knew the answer, without imposing an order. This allows for determining *how* participants performed the judgments and provides some information on the timing and order of their judgments.

Lastly, Experiment 1 was unbalanced in terms of “own” (each instruction presented twice = 6 items) vs. “other” (each instruction presented once = 3 items) recognition items. We therefore increased the number of “other” items in Experiment 2 by including recordings from each instruction condition by a second other person.

3.1. Method

3.1.1. Participants

Participants were 106 Dutch undergraduate students (32 males, $M_{\text{age}} = 20.6$, $SD = 3.0$), who had not participated in Experiment 1 and took part for course credit or a monetary reward of 5 Euro. Twenty-nine additional participants (5 males; $M_{\text{age}} = 21.2$, $SD = 2.9$) enrolled in the study but had to be excluded due to bad calibrations (i.e., deviations $\geq 0.7^\circ$; $n = 16$), indicating during debriefing that they had forgotten about one or more of the answering options on one or more of the trials (despite repetitive presentation of the instructions in between trials, see below; $n = 12$) or technical problems ($n = 1$). All participants had normal or corrected-to-normal vision.

3.1.2. Design

A 2 (display condition: static or dynamic) \times 2 (information condition: informed or not informed) between-subjects design was used, resulting in four conditions to which participants were randomly assigned. The order in which the recognition tasks (i.e., own/other recognition and instruction recognition) were presented to participants during recognition task instruction was counterbalanced within each condition, both during verbal instruction prior to the experiment and on screen in-between trials (this was done to ensure that the order of presenting the instructions would not affect the think aloud recognition data). As such, 27 participants in the dynamic condition were assigned to the informed condition (of whom 13 received the order instruction recognition—own/other recognition), leaving 26 in the not-informed condition (of whom 13 received the order instruction recognition—own/other recognition). In the static condition, 26 participants were assigned to the informed condition (of whom 13 received order instruction recognition—own/other recognition) and 27 to the not-informed condition (of whom 13 received order instruction recognition—own/other recognition).

3.1.3. Materials

3.1.3.1. Stimulus and instructions: The same stimulus and instruction conditions (i.e., “ages”, “objects,” and “away”) were used as in Experiment 1. An additional image (i.e., Dutch government) was used in two practice trials and was presented to participants once under one instruction condition (i.e., “Count the number of people and those who wear a blue tie.”), followed by a practice recognition test to familiarize participants with the think aloud method (see “practice task”).

3.1.3.2. Gaze displays: The eye tracker and recording set-up was the same as in Experiment 1 and the gaze displays were generated, using the same settings as in Experiment 1. The “other” eye movement records covered each of the three experimental instructions plus the two filler instructions and were generated by two different people: the same as in Experiment 1, plus those from one other person, recorded prior to the experiment.

3.1.3.3. Recognition test: The recognition test consisted of 16 dynamic or static gaze displays, depending on the condition (i.e., the same as in Experiment 1, plus the five additional ones in the “other” condition). As in Experiment 1, gaze displays were presented in random order on the computer screen, using E-Prime (Version 2.0; Psychology Software Tools, Inc.), and each gaze display was presented for a maximum of 15 s.

Participants had to perform the same judgment tasks as in Experiment 1 (i.e., own/other recognition and instruction recognition), but now participants were instructed to provide their answers orally; that is, participants were asked to think aloud while watching the dynamic or static gaze display and to answer both questions: “Whose eye movements were these? Were these your own or someone else’s? And which instruction is reflected by the eye movements? Is that the instruction of judging the ages of the people in the painting, remembering the positions of the objects in the room, estimating how long the unexpected visitor had been away, or none of these instructions?”. As stated above, the order within which the questions were presented was counterbalanced within conditions. Participants were instructed to provide their judgments as fast and as accurately as possible once they knew the answer to one or to both questions, and they were allowed to provide their judgments after the display had turned black (i.e., after 15 s) if they not yet had done so. Participants’ answers were recorded by a microphone attached to the stimulus PC. In between the trials, a slide appeared on screen (for 10 s) that reminded participants of both questions and answering options (in the same order as during verbal instruction).

3.1.3.4. Practice task: A practice task was used to familiarize participants with thinking aloud during the recognition test. Overlaid on the practice image, two “other” gaze displays were presented, one recorded under the same instruction that participants had carried out themselves and a different one. The gaze display was either static or dynamic depending on the condition. Participants were instructed to judge aloud whether their own or someone else’s eye movements were displayed and whether the display reflected the same instruction they had been presented with or not.

3.1.4. Procedure

Participants were seated in front of the monitor and were shown Ilya Repin’s *The Unexpected Visitor* for free inspection (i.e., 15 s). Then, participants were instructed that they would be presented with the image three more times under different instruction conditions, while their eye movements were recorded. The *informed* group was then told that they had to carefully follow each instruction while inspecting the image, because in the second part of the experiment they would be shown their gaze displays and would be

asked to judge which instruction a display reflected. The *not informed* group was told that they had to carefully follow each instruction while inspecting the image, because their eyes were being tracked (which was identical to the instruction in Experiment 1). Hereafter, participants were positioned in the head- and chinrest and the eye tracker was calibrated, using the same procedure as in Experiment 1 (after removal of participants who deviated $\geq 0.7^\circ$: dynamic condition: accuracy $M = 0.36^\circ$, $SD = 0.12^\circ$; static condition: accuracy $M = 0.40^\circ$, $SD = 0.13^\circ$), and the instruction conditions were presented in random order. Finally, the additional practice image was presented under one instruction condition. Subsequently, the experimenter generated the participant's dynamic or static gaze displays (i.e., $n = 3$), while participants performed a filler task (i.e., a puzzle) for approximately 5 min. Participants then performed the practice task to familiarize them with thinking aloud and subsequently performed the recognition test. Afterwards, participants were debriefed and asked whether they experienced any difficulties in performing the task or remembering all answering options (and, if so, they were excluded from the sample; see "Participants" section).

3.1.5. Data analysis

Analysis of performance data was similar as in Experiment 1, except that the experimenter kept track of the provided oral judgments during the course of the experiment.

3.1.5.1. Order of response: In order to assess whether participants had a preference for first providing one judgment or the other in both the static and dynamic condition, we computed the relative number of occasions on which they first provided their judgment for the own/other recognition task. This was done by dividing the total number of instances on which they first provided the judgment of the instruction task by the total number of trials (excluding the filler trials; $n = 1,272$). A value > 0.50 shows that the instruction recognition was mentioned first, < 0.50 the own/other recognition was mentioned first, and at 0.50 both are mentioned first equally often. Subsequently, separate one-sample *t*-tests were conducted for each condition to assess whether the proportion of answering the own/other recognition task first differed significantly from chance (i.e., 0.50).

3.1.5.2. Response times: An independent rater scored the response times with respect to both judgment tasks (i.e., onset and end of first judgment and onset of second judgment). A second independent rater scored a subset of the data (i.e., 10%); Intra Class Correlation was high ($ICC's \geq 0.990$). Responses for the filler trials were not included in the analyses. Additionally, a total of 20 responses (i.e., 1.6%) had to be excluded due to participants taking too much time for providing their judgments (> 30 s with the display turning black after 15 s, $n = 7$), correcting their judgments or making comments (because this rendered it unclear which onset to use in the analyses; $n = 5$), or forgetting to provide a judgment (i.e., the experimenter reminded him/her to do so; $n = 8$), leaving a total of 1,252 responses.

Response times were used to determine when both judgments were provided and the time interval between the judgments. Subsequently, the interval between judgments was either categorized as short (i.e., interval between judgments < 500 ms), moderate (500–

1,000 ms), or long (> 1,000 ms; see Boomer & Dittmann, 1962; Rochester, 1973). We then explored whether performance (i.e., both judgments correct, one judgment correct, or both judgments incorrect) was associated with short, moderate, or long intervals, to get an indication of whether the two recognition decisions are integrated (short intervals) or separate (moderate and long intervals) decisions.

3.2. Results and discussion

3.2.1. Main analysis of performance

An overview of the data is provided in Table 3 (see also Fig. 3). One-sample *t*-tests (Bonferroni adjusted $\alpha = 0.0125$) revealed that the dynamic conditions performed above chance for the own/other recognition task (informed: $t[25] = 4.23$, $p < .001$, $d = .83$; not informed: $t[25] = 3.45$, $p = .002$, $d = 0.67$), but that the static conditions did not (informed: $t[24] = 2.53$, $p = .018$, $d = 0.51$; not informed: $t[25] = 2.27$, $p = .032$, $d = 0.44$; note that this is not significant after correction of the alpha).⁴ A univariate ANOVA with display condition (dynamic or static) and information condition (informed or not informed) as between-subjects variables and performance as dependent variable revealed neither main effects (display condition: $F[1, 99] = 3.056$, $p = .084$, $\eta_p^2 = 0.030$, information condition: $F < 1$), nor an interaction effect ($F < 1$). Overall, our findings replicated those from Experiment 1 in that participants only performed better than chance at making own/other recognition judgments with dynamic gaze displays, but not with static gaze displays.

For the instruction recognition task, one-sample *t*-tests (Bonferroni adjusted $\alpha = 0.0125$) revealed that performance was above chance irrespective of display condition or information condition (dynamic/informed: $t[26] = 15.40$, $p < .001$, $d = 2.96$; dynamic/not informed: $t[25] = 13.32$, $p < .001$, $d = 2.61$; static/informed: $t[25] = 12.56$, $p < .001$, $d = 2.47$; static/not informed: $t[26] = 10.60$, $p < .001$, $d = 2.04$). A univariate ANOVA with display condition (dynamic or static) and information condition (informed or not informed) as between-subjects variables and performance as dependent variable only revealed a significant main effect of the display condition, $F(1, 102) = 8.51$, $p = .004$, $\eta_p^2 = 0.077$, but no main effect of information condition ($F < 1$) and no interaction effect ($F < 1$). The main effect of the display condition replicated the finding from Experiment 1, in that the dynamic group outperformed the static group on the instruction recognition task. However, being informed about the subsequent judgment task did not result in any improvements (or decrements) in the performance.

In sum, no evidence was found that informing participants about the subsequent judgment task led to higher recognition performances. When comparing performance on own versus other gaze trials (see Table 3), no such advantage was observed either: A 2 (display condition: static or dynamic) \times 2 (information condition: informed or not informed) \times 2 (gaze: own or other) mixed-factor ANOVA on percentage correct on the own/other recognition task did not reveal any main or interaction effects (gaze: $F[1, 102] = 3.58$, $p = .061$, $\eta_p^2 = 0.34$; display condition: $F[1, 102] = 3.62$, $p = .060$, $\eta_p^2 = 0.34$; information condition: $F < 1$; all interactions: $F_s[1, 102] \leq 1.27$, $p_s \geq 0.262$, $\eta_p^2_s = 0.012$). The same applied to the instruction recognition task: A mixed-factor ANOVA on percentage

Table 3

Mean percentage (*SD*) of correct responses for both recognition tasks per condition and as function of whether the gaze replay was a participant's own or someone else's, and performance on each instruction condition in Experiment 2

	Dynamic Condition		Static Condition	
	Informed	Not Informed	Informed	Not Informed
Own/other recognition	60.8 (13.6)	62.2 (18.1)	56.1 (12.8)	55.6 (16.2)
Own gaze	62.4 (23.4)	58.3 (23.7)	51.9 (22.8)	50.6 (25.5)
Other gaze	59.3 (20.3)	66.0 (22.8)	60.3 (17.7)	60.5 (16.8)
Fillers	84.3 (24.2)	88.5 (14.5)	85.6 (16.1)	83.3 (19.6)
<i>d'</i>	0.63 (0.76)	0.68 (1.01)	0.36 (0.71)	0.37 (0.84)
Instruction recognition	66.7 (14.1)	69.9 (17.2)	60.6 (14.4)	58.3 (16.3)
Own gaze	64.2 (24.8)	71.8 (24.4)	63.5 (25.4)	60.5 (25.8)
Other gaze	69.1 (17.7)	67.9 (18.8)	57.7 (20.1)	56.2 (16.8)
Fillers	38.0 (32.1)	38.5 (29.4)	32.7 (24.3)	38.0 (22.3)
Age	55.6 (18.8)	56.7 (25.1)	51.0 (21.8)	52.8 (26.3)
Object	86.1 (20.0)	89.4 (18.9)	81.7 (20.7)	78.7 (21.6)
Away	58.3 (26.9)	63.5 (28.5)	49.0 (22.9)	43.5 (25.6)

Note. Chance levels for own/other recognition and instruction recognition was 50% and 25%, resp.; total number of trials for overall performance: $n = 12$; Own gaze: $n = 6$; Other gaze: $n = 6$; Fillers: $n = 4$.

correct on the instruction recognition task only revealed the main effect of display condition ($F[1, 102] = 8.51, p = .004, \eta_p^2 = 0.077$), with the dynamic condition yielding a better performance than the static condition, but no other effects (all F s < 1). Since we did not find any effects of being informed on performance, either overall or when distinguishing between own and other trials, the data of information conditions will collapse within each display condition in subsequent analyses.

3.2.2. Performance differences among instructions

In order to replicate the findings of Experiment 1, we assessed performance on each instruction task separately. Consistent with Experiment 1, the data revealed that performance on the “objects” instruction was much higher than on the other two instructions (see Table 3). This was confirmed by a 2 (display condition: dynamic or static) \times 3 (instruction: ages, away or objects) mixed-factor ANOVA with percentage correct as the dependent variable, which showed, next to the main effect of condition ($F[1, 104] = 8.59, p < .001, \eta_p^2 = 0.076$; dynamic $>$ static), a main effect of instruction ($F[2, 208] = 72.48; p < .001, \eta_p^2 = 0.411$), but no interaction effect ($F[2, 208] = 1.68, p = .190, \eta_p^2 = 0.016$). Bonferroni-corrected follow-up comparisons on instruction revealed that the “objects” instruction yielded better performance (both p s $< .001$) than the other two instructions, for which performance did not differ ($p = 1.000$).

3.2.3. Analysis of response sequence

To assess whether or not participants had a preference for first making one judgment or the other, we conducted one-sample t -tests (Bonferroni adjusted $\alpha = 0.025$)

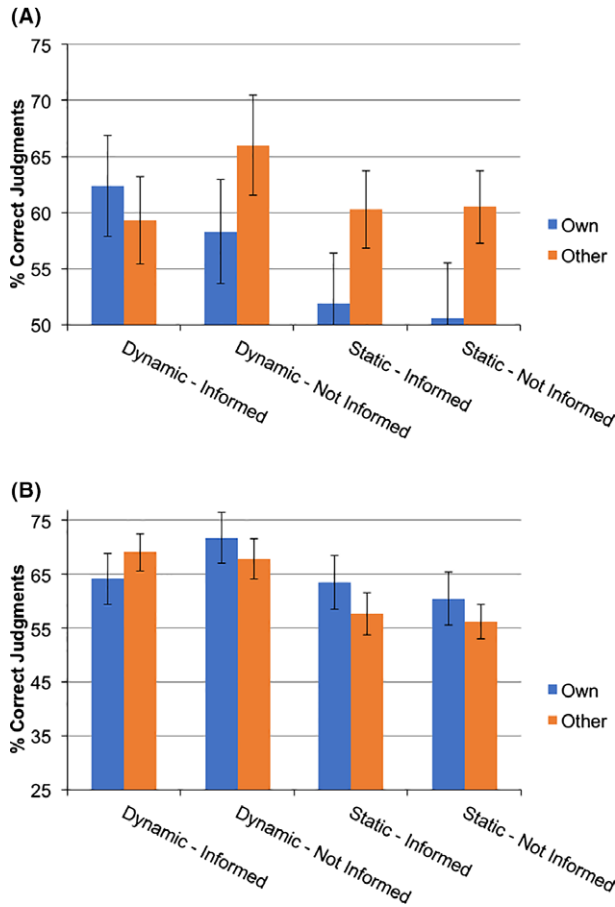


Fig. 3. Percentage of correct judgments in Experiment 2 for (a) *own/other recognition task*, and (b) *instruction recognition task* presented for each display condition and information condition and for own ($n = 6$) and other ($n = 6$) gaze trials. The y-intercept represents chance level; error bars represent standard error of the mean.

for both display conditions (dynamic or static) with the relative number of occasions in which they first provided their judgment of the instruction recognition task as the dependent variable. These analyses revealed that participants in the dynamic condition were more inclined to first provide the instruction judgment and then the own/other judgment ($M_{\text{instruction}} = 0.64$, $SD = .41$; $t(52) = 2.59$, $p = .013$, $d = 0.36$). In the static condition, this preference was not statistically significant ($M_{\text{instruction}} = 0.61$, $SD = 0.42$, $t(52) = 1.87$, $p = .067$, $d = 0.26$). This preference for first providing the instruction judgment, at least in the dynamic condition, resonates well with the finding that it was harder to recognize whether or not a gaze display was one's own than to recognize the instruction reflected in a gaze display, as evidenced by performance on the recognition tasks.

3.2.4. Response times and timing of judgments

An overview of the response times is provided in Table 4. The pattern of the response times is similar to the pattern observed in Experiment 1; that is, participants in the dynamic condition took more time to provide both judgments than participants in the static condition (1st judgment: $U = 317.00$, $z = -6.872$, $p < .001$, $r = .67$; 2nd judgment: $U = 443.00$, $z = -6.075$, $p < .001$, $r = .59$). When comparing the response times to the ones in Experiment 1, it appears that participants were considerably faster in the current experiment (i.e., approximately 3–5 s on average), which might be due to the changes in the procedure (i.e., think-aloud response, resolving the need to move the eyes away from the display toward the numeric keyboard, and the opportunity to provide the judgments in any order). In line with this observation, the occurrences in which participants took more time to make their decision than the presentation time of the gaze display of 15 s was considerably lower (i.e., *dynamic*: 28.0%; *static*: 5.1%) than in Experiment 1 (i.e., *dynamic*: 67.1%; *static*: 30.7%).

With regard to time intervals between judgments, which can be used to explore whether the two judgments were likely made in one integrated manner or separately, there was a similar pattern for the dynamic and static conditions: 33.2–34.4% judgment intervals were short (< 500 ms), 15.8–16.9% were moderate (500–1,000 ms), and 49.8–49.9% were long (> 1,000 ms). In addition, we explored whether short intervals were more often associated with correct answers than long intervals by computing the frequencies for both correct, only one judgment correct, or both judgments incorrect for each interval. This did not reveal a systematic trend, but rather it showed that for the dynamic condition only 34.1% of the cases in which both judgments were correct had a short interval, 18.1% a moderate interval, and 47.8% a long interval. The same was true for the static condition (i.e., 30.4%, 18.5%, and 51.1%, resp.). Hence, making one integrated decision (as reflected in a short time interval) was not unambiguously associated with the judgments being correct, nor was making separate decisions (as reflected in long intervals) unequivocally associated with having both judgments incorrect (i.e., only 14.5% [dynamic condition] to 17.9% [static condition] of the cases in which both judgments

Table 4

Mean response times (*SD*) per condition, occurrences of response times lasting shorter/longer than 15 s presentation time, and mean inter-response interval (*SD*) for each condition in Experiment 2

	Dynamic Condition		Static Condition	
	Informed	Not Informed	Informed	Not Informed
Mean (<i>SD</i>) response time for 1st judgment (s)	9.3 (4.1)	8.9 (3.3)	5.1 (2.9)	4.7 (2.8)
Mean (<i>SD</i>) total response time (s)	12.4 (4.4)	11.9 (3.3)	8.0 (3.1)	7.9 (2.9)
Occurrences of < 15 s and > 15 s response time for 1st judgment (% RT > 15 s)	257/61 (19.2)	261/46 (15.0)	293/12 (3.9)	310/12 (3.7)
Occurrences of < 15 s and > 15 s response time in total (% RT > 15 s)	224/94 (29.6)	224/83 (27.0)	289/16 (5.2)	306/16 (5.0)
Inter-response Interval (s)	1.61 (0.82)		1.46 (0.77)	

Note. Fillers were excluded.

were incorrect had a long interval between the judgments). In sum, participants seemed to make an integrated decision on approximately one-third of the trials, but a separate decision on at least half of the trials, and this was not associated with correctness. The time interval between judgments might, however, be associated with participants' *confidence* in the correctness of their decisions (i.e., when an answer comes to the mind fluently, people tend to be more confident in its correctness; e.g., Ackerman & Zalmanov, 2012). This we did not assess, but it might be interesting to take into account in future research.

3.3. Conclusion

In line with the findings from Experiment 1, recognition of whether a gaze display was or was not one's own was above chance only for the dynamic condition, but not for the static condition. However, as in Experiment 1, the dynamic condition did not yield better performance than the static condition on the own/other recognition task. Moreover, the dynamic group outperformed the static group in interpreting which instruction was reflected in the gaze displays, with both groups performing well above chance in both conditions. The findings of Experiment 2 extend those of Experiment 1 by providing insight into *how* participants performed the task. Using a think-aloud method, it was revealed that participants had a preference for first providing the instruction judgment, although this preference was only significantly different from chance in the dynamic condition. Note that the think-aloud method only resulted in participants providing their judgment (e.g., "This is objects, and . . . mine"); it did not yield any information on how they came to that judgment. This might be due to the limited time of the stimulus presentation, which was clearly long enough to perform the tasks, but might have been too short to provide an elaborate justification of why they chose a particular instruction condition or of why they thought it was or was not their own gaze display.

Next to replicating and extending the findings from Experiment 1, a second aim of Experiment 2 was to uncover whether being informed about the subsequent judgment task would improve recognition performance. Contrary to our predictions, however, no benefits of being informed were found. This suggests that participants did not strategically change their viewing behavior to facilitate later judgments when they knew in advance about the judgment tasks. This is in line with findings by Marti et al. (2015), who found that participants were quite good at introspection, but knowing that they would have to introspect, did not result in fundamentally different eye movement patterns compared to participants who were not instructed to introspect.

4. Experiment 3

Experiments 1 and 2 showed that the dynamic condition resulted in better instruction recognition performance than the static condition, presumably because the dynamic condition inherently provided spatial-temporal information, which was only limitedly available

in the static displays. That is, whereas the dynamic condition inherently provided information of what was inspected and in what order, this information was present only to some degree in the static condition by the connecting lines between two consecutive fixations. To investigate the role of order information for instruction recognition performance, we conducted a third experiment in which we compared three different display types: (a) dynamic displays (i.e., exact order information); (b) static with lines between consecutive fixations (i.e., some but limited order information); and (c) static displays without lines between consecutive fixations (i.e., no order information). In all displays, temporal information was provided by displaying relative fixation duration (i.e., smaller circles for shorter fixations, larger circles for longer fixations) to eliminate any positive effects of being provided with temporal information in the dynamic condition and not in the static conditions.

In line with the findings from Experiments 1 and 2, we hypothesized that the dynamic condition would yield higher performance than the two static conditions. In addition, we expected that being provided with some information about the order of fixations in the static display would lead to better performance than not having any order information available, especially for the “away” instruction for which the (limited) order information is important to distinguish it from the “ages” instruction (i.e., both yield many fixations on the faces but differ in the amount of switching between faces; see also DeAngelus & Pelz, 2009). Hence, it is expected that performance on these instructions is worst for the static without order information compared to the static with limited order information and dynamic display conditions. As we now were only interested in interpreting another person’s eye movements in terms of the task instruction, this third experiment was conducted online via Amazon’s Mechanical Turk (Paolacci, Chandler, & Ipeirotis, 2010).

4.1. Method

4.1.1. Participants and design

Participants ($n = 60$) were recruited via Amazon’s Mechanical Turk (see Paolacci et al., 2010) and were paid a small incentive for their participation (\$2). Three participants had to be excluded due to incomplete data ($n = 1$) or due to technical difficulties ($n = 2$). The remaining 57 participants ($M_{\text{age}} = 38.8$ years, $SD = 12.3$ years, range 19–71 years; 29 males) were U.S. residents, performed the study on a laptop or computer in an environment without distractions (i.e., self-reported noise rating lower than seven on a scale of one to nine), and were not color blind.

A within-subjects design was employed, so all participants performed the instruction recognition task under each display condition. Order of display conditions was counter-balanced across participants.

4.1.2. Materials

4.1.2.1. *Gaze displays*: Gaze displays were presented in Qualtrics software (Qualtrics, Provo, UT, USA). They were generated using the data of the five participants with the highest calibration accuracy in Experiment 1. SMI BeGaze software (Version 3.4;

SensoMotoric Instruments) was used to generate the displays for three instruction conditions (i.e., “ages,” “away,” and “objects”). For each of these five participants and each instruction condition, three gaze displays were generated (one for each display condition), yielding 45 different gaze displays. The static and dynamic displays of eye movements were generated using the “Scan Path” utility in BeGaze, with fixations defined as lasting at least 100 ms with a maximal dispersion of 50 pixels. Fixations were displayed as yellow circles with a line width of 4 pixels. Size varied depending on duration of the fixation (e.g., the diameter of a fixation of 500 ms was 80 pixels). In the static displays with order information, consecutive fixations were connected by a yellow solid line (4 pixels). In the dynamic displays and the static displays without order information, no lines between fixations were visible (i.e., no gaze trail). See Fig. 4 for screenshots of the different gaze displays.

4.1.2.2. Recognition test: The recognition test consisted of three blocks of 15 gaze displays for each display condition, consisting of five times each instruction condition. Within each display condition, gaze displays were presented randomly and each gaze display was presented for a maximum of 15 s. Participants were instructed to judge which instruction is reflected in the gaze displays shown to them, by inspecting the image and then proceeding to the next page (either manually by clicking the next page button or automatically after 15 s) and providing the answer by selecting one of the three instruction conditions. Prior to each gaze display, participants were reminded of the question and answering options, which remained on screen for a maximum of 10 s, after which the system automatically proceeded to the next page or shorter, when participants proceeded to the next page earlier by clicking on the next-page button.

4.1.3. Procedure

After logging on to the system, participants were asked to fill out some demographic questions. As in Experiments 1 and 2, they then were provided with Ilya Repin’s *The Unexpected Visitor* for free inspection (i.e., 15 s). Then, participants were instructed that they would be presented with the image three more times under different instruction conditions and that they had to carefully follow each instruction while inspecting the image. Participants then performed the recognition test in three blocks of 15 displays. The entire experiment lasted approximately 30 min.

4.1.4. Data analysis

Analysis of performance data was similar as in Experiment 1: For each display condition, a one-sample *t*-test was used to assess whether performance was above chance (i.e., number correct judgments > chance [i.e., 33.33%]). A repeated measures analysis of variance (RM-ANOVA) was used to analyze differences in performance with the three display conditions and instruction conditions as within-subjects factors. Helmert contrasts were applied to test our first hypothesis that the dynamic condition would yield higher recognition performance than the two static conditions and to test our second hypothesis that the

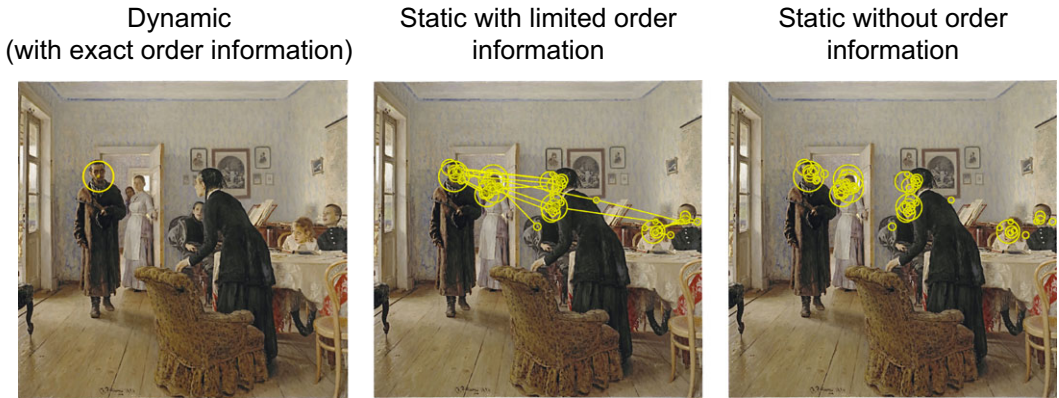


Fig. 4. Screenshots of the different displays in response to the “age” instruction condition presented in Experiment 3. From left to right: dynamic gaze display, static gaze display with limited order information, and static gaze display without order information.

static condition with limited order information would yield higher performance than the static condition without order information.

4.2. Results and discussion

An overview of the data is provided in Table 5. One-sample *t*-tests (Bonferroni adjusted $\alpha = 0.017$) revealed that all conditions performed above chance in recognizing what instruction was reflected in the gaze displays (dynamic: $t[56] = 19.11$, $p < .001$, $d = 2.53$; static with limited order information: $t[56] = 18.79$, $p < .001$, $d = 2.49$; static without order information: $t[56] = 17.23$, $p < .001$, $d = 2.28$).

To assess whether there were any differences in performance as a result of display condition and instruction, a 3 (display condition: dynamic, static with limited order information, static without order information) by 3 (instruction: ages, objects, away) repeated measures analysis of variance was conducted with planned Helmert contrasts to test our hypotheses. This analysis revealed a main effect of display condition: $F(2, 112) = 4.23$, $p = .017$, $\eta_p^2 = 0.070$. The planned Helmert contrasts confirmed our first hypothesis that the dynamic condition would yield higher performance than the two static conditions: $F(1, 56) = 6.47$, $p = .014$, $\eta_p^2 = 0.104$, but not our second hypothesis that the static condition with limited order information would yield higher performance than the static condition without order information: $F(1, 56) = 1.43$, $p = .237$, $\eta_p^2 = 0.025$. In addition, a main effect of instruction was revealed: $F(2, 112) = 103.72$, $p < .001$, $\eta_p^2 = 0.649$. Bonferroni post hoc comparisons revealed that participants performed significantly better on the “objects” instruction than on the “ages” and “away” instructions (both p 's $< .001$). Performance on the “ages” and “away” instructions did not differ significantly ($p = .110$). These main effects were qualified by a significant interaction between display condition and instruction: $F(4, 224) = 5.15$, $p = .001$, $\eta_p^2 = 0.084$. We further analyzed this interaction by testing whether performance in the display conditions differed for each

Table 5

Mean percentage (*SD*) of performance on each instruction condition in Experiment 3

	Dynamic Condition With Exact Order Information	Static Condition With Limited Order Information	Static Condition Without Order Information
Overall performance	70.9 (14.8)	67.6 (13.8)	65.5 (14.1)
Age	63.5 (24.8)	57.2 (23.4)	58.9 (24.3)
Object	88.1 (18.5)	91.2 (15.1)	91.9 (16.0)
Away	61.1 (25.5)	54.4 (24.4)	45.6 (21.0)

instruction. The planned Helmert contrasts revealed that performance differed between display conditions only for the “away” instruction, with the dynamic condition (i.e., exact order information) yielding higher performance than the static conditions, $F(1, 56) = 11.77$, $p = .001$, $\eta_p^2 = 0.174$, and the static condition with limited order information yielding better performance than the static condition without order information, $F(4, 224) = 7.49$, $p = .008$, $\eta_p^2 = 0.118$. On the “ages” and “objects” instructions, there were no significant performance differences among conditions (all p 's $\geq .102$).

4.3. Conclusion

The aim of this third experiment was to investigate the role of order information in instruction recognition by comparing display conditions that provide exact, limited, or no order information. Overall, the main effect of display condition seemed to indicate that — in line with our hypothesis — the dynamic condition (exact order information) had higher instruction recognition performance than the static conditions, but in contrast to our hypothesis, the limited order information present in the static condition with connecting lines did not yield higher overall performance than the static condition without order information. However, these findings were qualified by an interaction with instruction: As expected, benefits of order information were especially apparent for the “away” instruction. Only on this instruction did the dynamic condition significantly outperform the static conditions and did the static condition with limited order information outperform the static condition without order information.

5. General discussion

The aim of Experiments 1 and 2 was to investigate observers' ability to recognize whether a static or dynamic gaze display was their own or someone else's and to interpret which viewing instruction was reflected in these gaze displays, and whether being informed about these tasks beforehand would lead to increased performance on these tasks (Experiment 2). Both experiments showed that observers were only able to recognize whether a gaze display was their own or someone else's above chance level when they were shown dynamic replays, not when they were shown static displays, though

there were no significant differences in performance between the static and dynamic conditions. Moreover, both experiments showed that observers could interpret (i.e., match the corresponding instruction to) both static and dynamic gaze displays above chance level, with dynamic displays yielding better performance than static displays. Being informed beforehand about the instruction recognition task (Experiment 2) did—rather surprisingly—not affect performance on either the own/other or the instruction recognition task.

In Experiment 3, we investigated the role of order information in instruction recognition, by comparing exact (i.e., dynamic displays), limited (i.e., static displays with connecting lines between fixations), and no order information (i.e. static displays without connecting lines). Order information only helped recognition for one type of instruction: Recognition performance on the “away” instruction was highest in the dynamic condition and was higher in the static condition with connecting lines (limited order information) than in the static condition without order information.

5.1. *Distinguishing your own from another’s gaze displays*

The first finding that for static displays recognition of whether a gaze display was one’s own or another person’s was not above chance is consistent with previous studies (Foulsham & Kingstone, 2013b). Although performance was consistently above chance when using dynamic displays, the dynamic condition did not yield higher performance than the static condition on this own/other recognition task. Indeed, it is rather surprising that the temporal unfolding inherent in dynamic replays did not seem to foster recognition of the one’s own eye movements as one’s own, even when the participants were informed about the judgment task prior to eye movement recordings. Possibly, the instructions used in this study resulted in highly similar scanpaths of self and “others.” In that case, it would be hard for participants to recognize the difference, as the other displays would also provide good memory cues. The finding that being informed did not result in any performance benefits is in line with recent observations suggesting that people’s awareness of where they have looked is poor (Clarke, Mahon, Irvine, & Hunt, 2016; Vö, Aizenman, & Wolfe, 2016) but seems to be in contrast with the findings of Marti et al. (2015), who reported that participants were able to consciously monitor their viewing behavior and recollect and report on this information (note that the eye movement patterns did not differ fundamentally when introspection was and was not required). However, our study differs from the Marti et al. study, in which participants provided reports about their viewing behavior directly after completing a serial search, in two respects. First, our participants had to *recognize* whether an eye movement pattern was theirs, which might ask for a different recollection of information as compared to actively reconstructing one’s own scan path. Second, in this study, participants had to retain information on where they had looked for a longer period of time, as the recognition test was presented only after they had inspected the image under each instruction condition. Thus, taken together, it seems that memory of one’s own viewing behavior is short-lived, making recollection of this information at a later stage difficult (which would be in line with previous accounts on the validity of introspection; e.g., Ericsson & Simon, 1980).

5.2. Interpreting gaze displays in terms of cognitive processes

The finding that participants performed above chance in both the static and dynamic conditions at recognizing which instruction condition was reflected in the displays, contrasts with the findings by Greene et al. (2012), who suggested that observers were unable to infer a person's task based on a static gaze display. Possibly, this is due to the fact that our instructions were a selection of three of Yarbus's (1967) original instructions, one of which we could reasonably expect to stand out from the other two (i.e., "objects"), whereas we could expect the other two to partially overlap in terms of the areas fixated, but not in the order of fixations (i.e., "ages" and "away"; DeAngelus & Pelz, 2009). In the Greene et al. study, it was unclear whether the various instructions would yield a distinct gaze pattern on the various images that they used (see Borji & Itti, 2014).

Tentative evidence that the distinctiveness of gaze patterns in response to the different instructions plays a role in recognition performance comes from the difference in performance among instructions. The "objects" instruction, which distinguishes itself from the other two in that it requires little if any focus on people, resulted in much higher recognition performance than the other two instructions, which bore more resemblance to each other and often were confused with one another (see Table A1 in the Appendix). However, resemblance among instructions seemed to have more deleterious effects on instruction recognition performance in the static than in the dynamic condition in all experiments. Overall, we found an advantage of dynamic over static displays with regard to instruction recognition. This advantage seemed most apparent in the "away" instruction (see Tables 1, 3, and 5), which differs from the "ages" instruction mainly in terms of *transitions* between the faces of people depicted in the painting (see DeAngelus & Pelz, 2009). This is not surprising because the dynamic gaze displays inherently provide temporal and spatial information with regard to transitions. The importance of order information was underlined by the findings from Experiment 3, which showed that the exact order information (dynamic displays) and limited order information (static displays with connecting lines) was only helpful (compared to static displays with no order information) for recognition performance on the "away" instruction, presumably because the exact and limited order information highlighted the transitions between the faces of the people in the picture.

These differences in people's ability to reliably recognize the task instruction from a given scanpath, but struggle to identify one's own scanpath from someone else's, can be reconciled when considering the scale of the variability in eye movement patterns across tasks compared to the variability across individuals (Litchfield et al., 2010). Litchfield et al. argued that the variability in task-specific eye movements dwarfs the variability in the eye movements across individuals performing the same task and therefore the task-specificity of eye movements would be a major contributing factor as to how well gaze displays could be interpreted (and exploited). In their study, observers could benefit from viewing another person's eye movement patterns but gained little advantage in watching an expert over a novice performing the same task. Moreover, the beneficial effects in performance were eliminated if the eye movements shown were not task-specific. It is well

established that expert and novice eye movement behaviors are statistically different from each other (for a meta analysis, see Gegenfurtner, Lehtinen, & Säljö, 2011), but this is not apparent when the eye movement patterns are actually observed as gaze displays. Instead, eye movements need to be substantially different from each other (i.e., coming from a different task) to lead to differential effects on performance.

Unlike Litchfield et al. (2010), this study explicitly tested how well people can recognize these recorded eye movement patterns. As mentioned earlier, the task instructions may have resulted in observers producing highly similar scanpaths for each instruction and so it was always going to be difficult to differentiate one scanpath from another when the task is held constant. In contrast, different task instructions yield distinct gaze patterns (DeAngelus & Pelz, 2009; Land, 2006; Yarus, 1967), and so it follows that observers would find it easier to recognize task instructions from scanpaths.

5.3. *Limitations and future research*

While the restriction to a single picture and just three carefully chosen instructions can be regarded as a limitation of this study, we felt that this was a good starting point for addressing our research questions, considering the lack of evidence for people's ability to meaningfully interpret gaze displays thus far (Greene et al., 2012). Yet future research should address these questions, using other stimulus materials.

Another limitation is the rather limited number of observations in each cell (i.e., nine observations in Experiments 1 and 16 in Experiment 2), which results in a rather large variability in our signal detection measure d' . Hence, future research including more observations should replicate the finding that recognition of whether a gaze display is one's own or another person's is above chance for dynamic but not for static displays.

In addition, the current findings leave the question unanswered of how a display of eye movements is encoded and whether participants attempt to compare it with the episodic memory of where they looked under each instruction condition. Although we used a think-aloud method in Experiment 2, it did not provide much information on *how* people interpreted or processed the displays, presumably because of the limited presentation time (15 s). Findings from standardized memory recognition tasks may shed light on this complicated issue. For example, research by Gallo (2004) and Gallo, Bell, Beier, and Schacter (2006) suggests that the *presence* of specific information can be used as a disqualifying criterion as to whether an event is believed to have actually occurred. In our case, participants may identify a specific feature or features of the display that helps them distinguish the task instruction associated with it and/or whether it was their own. For instance, if the "other" eye movement recording showed fixations on an area that the participant consciously remembers never having looked at, this would be a criterion for rejecting it as being their own eye movement recording and as reflecting one of the three viewing instructions. In addition to, and separate from disqualifying monitoring processes, Gallo et al. suggest that the *absence* of specific information can be exploited by diagnostic monitoring processes, which compare recollected information with that which is expected to be recollected (i.e., recalling that one has looked at a certain area under a certain

instruction condition, but the eye movement pattern not showing a fixation at that particular area). As the present study did not provide information on this matter, future research in gaze interpretation could address this, for instance, by interviewing participants about *why* they thought a display was their own or not, or reflected one instruction and not another. In addition, in order to get insight into *what* information participants exploit in making their judgments, future research might manipulate the duration of the gaze displays or number of visible fixations to see whether or not this affects performance.

Another way of making further progress in this regard would be to systematically investigate the inferences that participants can make from recorded eye movement patterns, given what we already know about how eye gaze is used in face-to-face situations (Argyle & Cook, 1976; Kleinke, 1986; Shepherd, 2010), and the extent to which people can infer what other people are thinking (e.g., Apperly & Butterfill, 2009; Baron-Cohen, 1995; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). For example, it is well established that we have a predisposition to process and follow eye gaze from a face; however, understanding why another person was looking at a particular item can involve relatively simple intentionality processes (e.g., Baron-Cohen's intentionality detector) or more advanced "theory-of-mind" processes that help the observer take the perspective of others in more complex situations. The question is still open whether these same processes are also being recruited to help recognize and interpret recorded gaze patterns (cf. Litchfield et al., 2010). The fact that observers can regulate their own gaze displays to help or deceive observers depending on the sociocommunicative context (Foulsham & Lock, 2014; see also Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008) suggests that there may be some overlap in processes and strategies employed in normal face-to-face situations, but clearly further research is required that makes specific comparisons to face-to-face gaze interpretation abilities. Previous gaze display interpretation studies (e.g., Foulsham & Lock, 2014; Greene et al., 2012; Zelinsky et al., 2013) have required observers to make either relatively simple inferences or more complex inferences about these gaze displays, with varying degrees of success. This study adds to this growing research area by showing that observers can recognize task instructions from these gaze displays but struggle to recognize which gaze display is their own, and that these respective abilities are affected by the type of display used (dynamic, static).

To conclude, we found that observers have difficulty recognizing whether a gaze display (static or dynamic) is their own or someone else's. On dynamic gaze replays own/other recognition performance was above chance, but performance differences between static and dynamic displays were not significant. In contrast, we found that observers are quite good at correctly identifying which task the viewer was performing from a gaze pattern, at least from the small selection of three instructions used in this study. Instruction recognition was facilitated by dynamic gaze displays compared to static gaze displays in the first two experiments, and the third experiment showed this is likely due to the order information that is inherently present in the dynamic displays. Order information was especially relevant for distinguishing instructions that result in similar fixation locations, but differ in the transitions among locations. In order to more fully understand under what conditions observers are able to make inferences about eye movement

displays, what kind of inferences they can make, and how this is achieved, future research should further explore the role of task complexity and distinctiveness of eye movement displays on inference making.

Acknowledgments

This research was supported by a Vidi grant (#452-11-006) from the Netherlands Organization for Scientific Research (NWO) awarded to Tamara van Gog. The authors thank Jacqueline de Nooijer, Kirsten Versijde, and Susan Ravensbergen for their help in conducting the experiment and scoring the data.

Notes

1. In order to calculate d' for the gaze recognition task, we applied the following matrix:

Response Category	Correct Answer	Provided Answer
Hit	Own	Own
False negative	Own	Other
False positive	Other	Own
Correct rejection	Other	Other

2. Note that when $d' > 0$, participants managed to correctly recognize whether a gaze display was their own or someone else's; that is, they had high hit/correct rejection rates and low rates of false negatives/positives. On the contrary, $d' < 0$ indicates that participants consistently recognized their own gaze display as someone else's and vice versa; that is, they had low hit/correct rejection rates and high rates of false negatives/positives. A $d' = 0$ indicates performance at chance level with hit/correct rejection rates being equal to rates of false negatives/positives (MacMillan & Creelman, 2005).
3. Note that all responses were included; however, the results are the same when excluding incorrect responses.
4. Note that we could not compute a d' for three participants as the hit rate was zero for two participants and for one participant the rate of false positives was = 1.

References

- Ackerman, R., & Zalmanov, H. (2012). The persistence of fluency-confidence association in problem solving. *Psychonomic Bulletin & Review*, 19, 1187–1192. doi:10.3758/s13423-012-0305-z
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970. doi:10.1037/a0016923

- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, UK: Cambridge University Press.
- Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*, 1185–1204. doi:10.1080/13506280902978477
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Boomer, D. S., & Dittmann, A. T. (1962). Hesitation pauses and juncture pauses in speech. *Language & Speech*, *5*, 215–220. doi:10.1177/002383096200500404
- Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, *14* (3), 1–22. doi:10.1167/14.3.29
- 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*, 1465–1477. doi:10.1016/j.cognition.2007.05.012
- Clarke, A. D. F., Mahon, A., Irvine, A., & Hunt, A. R. (2016). People are unable to recognize or report on their own eye movements. *The Quarterly Journal of Experimental Psychology*. doi: 10.1080/17470218.2016.1231208
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.
- DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*, 790–811. doi:10.1080/13506280902793843
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251. doi:10.1037/0033-295X.87.3.215
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev ed). Cambridge, MA: MIT Press.
- Foulsham, T., & Kingstone, A. (2013a). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*, 41–56. doi:10.1037/a0028227
- Foulsham, T., & Kingstone, A. (2013b). Where have eye been? Observers can recognise their own fixations. *Perception*, *42*, 1085–1089. doi:10.1068/p7562
- Foulsham, T., & Lock, M. (2014). How the eyes tell lies: Social gaze during a preference task. *Cognitive Science*, *39*, 1704–1726. doi:10.1111/cogs.12211
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6, 1–17. doi:10.1167.8.2.6
- Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120–128. doi:10.1037/0278-7393.30.1.120
- Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*, 730–741. doi:10.1080/09658210600648506
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*, 523–552. doi:10.1007/s10648-011-9174-7
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*, 62–466. doi:10.1111/1467-9280.02454
- Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*, 1–8. doi:10.1016/j.visres.2012.03.019
- Hansen, J. P. (1991). The use of eye mark recordings to support retrospection in software testing. *Acta Psychologica*, *76*, 31–49.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Science*, *9*, 188–194. doi:10.1016/j.tics.2005.02.009

- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, K. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction, 25*, 62–70. doi:10.1016/j.learninstruc.2012.11.004
- Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137–172. doi:10.1037/0033-295X.92.2.137
- Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 78–100. doi:10.1037/0033-2909.100.1.78
- Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*, 1000–1009. doi:10.3758/BF03195762
- Kostons, D., Van Gog, T., & Paas, F. (2009). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*, 121–132. doi:10.1016/j.learninstruc.2011.08.004
- Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research, 25*, 296–324.
- Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *The Quarterly Journal of Experimental Psychology, 64*, 649–656. doi:10.1080/17470218.2011.558628
- Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*, 251–262. doi:10.1037/a0020082
- MacMillan, N. A. (2002). Signal detection theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed). Vol. 4 (pp. 43–90). *Methodology in experimental psychology*. New York: John Wiley & Sons.
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marti, S., Bayet, L., & Deheane, S. (2015). Subjective report of eye fixations during serial search. *Consciousness and Cognition, 33*, 1–15. doi:10.1016/j.concog.2014.11.007
- Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*, 289–300. doi:10.1016/j.ergon.2005.11.008
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411–419. doi:10.2139/ssrn.1626226
- Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research, 2*, 51–81. doi:10.1007/BF01067111
- Russo, J. E. (1979). A software system for the collection of retrospective protocols prompted by eye fixations. *Behavior Research Methods & Instrumentation, 11*, 177–179. doi:10.3758/BF03205643
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266. doi:10.1037/a0018729
- Shepherd, S. V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience, 4*, 1–13. doi:10.3389/fnint.2010.00005
- Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichovsky, B. M. (2010). Yabus, eye movements and vision. *i-Perception, 1*, 7–27. doi:10.1068/i0382
- Valuch, C., Becker, S. I., & Ansorge, U. (2013). Priming of fixations during recognition of natural scenes. *Journal of Vision, 13*(3):3, 1–22. doi:10.1167/13.3.3
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*, 785–791. doi:10.1016/j.chb.2009.02.007
- Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*, 237–244. doi:10.1037/1076-898X.11.4.237

Võ, M. L. H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1477–1481. doi:10.1037/xhp0000264

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Zelinsky, G. J., Peng, Y., & Samaras, D. (2013). Eye can read your mind: Decoding gaze fixations to reveal categorical search targets. *Journal of Vision*, 13(14), 10, 1–13. doi:10.1167/13.14.10

Appendix

Table A1. Response matrix for all trials

		Instruction Condition	Given Response			
			Ages	Away	Objects	None
Experiment 1	Static	Ages	35	27	0	4
		Objects	3	2	53	8
		Away	31	25	0	10
	Dynamic	Filler	4	13	14	13
		Ages	42	17	1	6
		Objects	0	3	59	4
		Away	19	37	3	7
Experiment 2	Static—Informed	Filler	7	6	16	15
		Ages	53	42	0	9
		Objects	0	0	85	44
	Static—Not informed	Away	42	51	6	17
		Filler	9	11	13	34
		Ages	57	39	3	8
	Dynamic—Informed	Objects	0	3	85	45
		Away	41	47	0	14
		Filler	10	19	20	41
		Ages	60	36	0	12
Objects		6	2	93	7	
Away		33	63	1	11	
Dynamic—Not informed	Filler	14	14	39	41	
	Ages	59	40	0	5	
	Objects	2	1	93	8	
	Away	31	66	0	7	
Experiment 3	Static without order information	Filler	12	13	39	40
		Ages	168	111	6	
		Objects	11	12	262	
	Static with limited order information	Away	139	130	16	
		Ages	163	112	10	
		Objects	8	17	260	
	Dynamic (exact order information)	Away	117	155	13	
		Ages	181	99	5	
		Objects	17	17	251	
		Away	100	174	11	