Running Head: Recollection benefits from second eye fixations

Manipulating the depth of processing reveals the relevance of second eye fixations for recollection but not familiarity

Charlotte Schwedes[1], Demian Scherer[2] and Dirk Wentura[1]

[1]Saarland University, Saarbrücken, Germany

[2]University of Münster, Münster, Germany

Correspondence address:

Charlotte Schwedes
University of Saarland
Department of Psychology
Campus A2 4
66123 Saarbrücken

Phone: 0049 681 302 4641
Fax: 0049 681 302 4049
E-mail: c.schwedes@mx.uni-saarland.de

Abstract

It is well known that memory affects eye movements. However, the role of individual eye fixations for recognition memory processes has hardly been investigated. Recent findings show that second fixations are especially relevant for recollection, a process associated with the retrieval of context information, but less for recognition based solely on item familiarity. The aim of the present study was to overcome limitations of a previous study (Schwedes & Wentura, 2019) and to provide further evidence that second fixations are especially relevant for recollection-based recognition. Whereas recollection- and familiarity-based recognition was an unconstrained quasi-experimental variable in a previous study, here we manipulated the depth of stimulus processing in the encoding phase to experimentally manipulate the probability of subsequent item recollection. In the old/new recognition memory test, presentation of test probes was terminated after one or two stimulus fixations. “Old” responses in the recognition test were followed by a remember/know/guess procedure to assess recollection-based versus familiarity-based recognition. We found the expected depth of processing effect, with better recognition and more recollection-based responses after deep encoding. This effect, however, was significantly larger if two fixations instead of just one were allowed. There were no corresponding effects for familiarity-based recognition. Thus, a second fixation seems to play an important role only for recollection-based recognition.

Keywords: recognition memory; recollection; familiarity; eye fixations

Introduction

Eye movement behaviour changes as a function of experience with a stimulus (for a review, see Hannula et al., 2010). For example, in recognition memory tasks, “old” (i.e., studied) stimuli are fixated longer compared to new items. This memory-based effect is already observable in the duration of the first two eye fixations to a stimulus, with longer fixations to known compared to new items (Ryan, Hannula, & Cohen, 2007; Schwedes & Wentura 2012, 2016). Thus, memory seems to influence the duration of very early eye fixations.

However, less is known about the reversed path, that is, the relevance of the first two eye fixations for recognition memory performance and the underlying recognition memory processes. Hsiao and Cottrell (2008) were the first to investigate this issue by restricting the number of allowed test-stimulus fixations in an old/new recognition memory task. They found that recognition performance was already above chance level with only one test-stimulus fixation, but a significant increase in performance if two fixations were allowed. More than two fixations, however, did not further increase performance. Based on a sophisticated presentation technique, they were also able to show that the additional information that is gained from a second fixations plays a significant role in the two-fixation advantage.

Based on this research, Schwedes and Wentura (2019) investigated which type of recognition memory process benefits from a second fixation, thereby boosting performance with two test-stimulus fixations compared to just one. Single- as well as dual-process models of recognition memory agree that recognition memory performance is driven by familiarity and recollection (i.e., see Wixted & Mickes, 2010; Yonelinas, 2002). However, in single process models it is assumed, that both processes are combined in one memory strength signal and participants base their recognition performance on this signal. In contrast, dual process models assume that both processes can be directly accessed and make independent contributions to memory judgements. Although the aim of the current work is not to underpin one of these model classes, the current work proceeds from a dual-models perspective.

In Yonelinas' (1994) dual-process model, familiarity is thought to be a fast signal-detection process that is based on the assessment of a memory-strength signal and associated with the subjective experience that the item has been experienced before (e.g., "I know I have seen this face before."). Recollection, on the other hand, is assumed to be a slower threshold process that is based on the retrieval of context information associated with the earlier encoding of the specific item. It thus manifests in the subjective experience of remembering specific details about the encoding episode (e.g., "I realize this is a colleague I first met at the conference in Boston last year; she told me about her research on recognition memory.").

Schwedes and Wentura found recollection-based but not familiarity-based recognition to increase with two compared to only one test-stimulus fixation. Moreover, using the same technique as Hsiao and Cottrell (2008), they found two factors causally related to the significant increase in recollection-based responding: longer availability of the input, as well as additional information provided at the location of the second fixation. These findings indicated a functional role of second eye fixations for recollection-based but not familiarity-based recognition. They are in line with previous studies reporting a memory effect in second-fixation durations in tasks that require recollection (see Schwedes & Wentura, 2012, 2016), and a higher proportion of recollection after a more dispersed fixation pattern (Kafkas & Montaldi, 2012) or under free-viewing conditions compared to restricted-viewing conditions (Mäntylä & Holm, 2006).

However, Schwedes and Wentura (2019) explored recollection-based versus familiarity-based recognition as an unconstrained quasi-experimental variable. Thus, whether recollection was more or less likely was not under experimental control. In addition, using the remember-know procedure to estimate the proportions of the underlying recognition memory processes is subject to criticism concerning the potential misuse of “remember” responses for highly familiar items by participants. Thus, under a critical view “remember” responses (that should reflect recollection-based recognition) might have only reflected high familiarity. If that was the case in our previous study, a possible interpretation could be that second fixations are not only relevant for recollection but also relevant for high familiarity. To rule this out, it is important to experimentally create conditions that differ in the relative probability of recollection- and familiarity-based recognition. Therefore, in the present study we manipulated the levels of processing (LOP) of the learned material since several studies have revealed that recollection is especially susceptible to LOP manipulations, with more recollection-based recognition after deep compared to shallow encoding (for a review see Yonelinas, 2002). If second fixations are only relevant for recollection-based recognition, the number of allowed fixations should moderate the LOP effect (higher performance for deeply compared to shallowly encoded faces) in the proportion of “remember” responses. That is, a larger difference between deeply and shallowly encoded items should be observable with two compared to only one allowed fixation. However, if “remember” responses only reflect high item familiarity, the LOP manipulation should not result in a different effect regarding number of allowed fixations.

To this end, the present experiment used face stimuli that were presented in a shallow encoding task (gender categorization; see Bower & Karlin, 1974; Wig, Miller, Kingstone & Kelley, 2004) or a deep encoding task (intelligence rating; see Marinkovic et al 2009; Mueller, Bailis, & Goldstein, 1979). In the recognition phase, stimuli were presented for either one or two fixations; this was controlled by an eye-tracking device. In addition to a standard old/new recognition decision, we asked for a remember/know/guess categorization in case of an old response. We expected that recognition performance should be better for deeply encoded items compared to shallowly encoded items, with the boost due to an increased rate of remember judgments (as an index of recollection). The most important prediction, however, was that both the increased recognition performance as well as the increased recollection rate of deeply encoded items should be moderated by the number of fixations. In other words, the effects of the LOP manipulation should be more pronounced if two test-stimulus fixations were allowed compared to just one fixation. No such effects were expected for familiarity-based recognition.

Method

Participants

Thirty-two undergraduate students from Saarland University took part in the experiment; they received course credit for participating. The data of one participant were discarded due to poor overall recognition performance (i.e., performance at chance level). The data of the remaining participants ($N = 31$; 19 women, 12 men; mean age = 23 years) were analyzed. All participants had normal or corrected-to-normal vision, were native speakers of German, and gave informed written consent at the beginning of the experiment.

The effect of an additional, second test-stimulus fixation on recollection-based recognition in Schwedes and Wentura (2019) was associated with a large effect size of $d_Z = 0.76$. Power analysis for the present experiment assumed a medium-sized effect ($d_Z = 0.5$) in order not to rely on a single estimation and to have enough power for the moderation of the two-fixation advantage by encoding conditions. To detect an effect of $d_Z = 0.5$ with power $1\text{-}\beta = .80$ ($\alpha = .05$, one-tailed), the minimum sample size is 27 (G*Power3; Faul, Erdfelder, Lang, & Buchner, 2007). Testing thirty-two participants facilitated counter-balancing. Actual power with $N = 31$ was $1\text{-}\beta = .86$.

Design and Materials

The materials comprised 128 colored face images (56 men and 56 women) taken from the database of Schwedes and Wentura (2012). All faces were placed against a uniform gray background and measured 174 × 191 pixels (this corresponds to 4.1 × 4.5° of visual angle). Participants were shown half of the faces (28 men and 28 women) in the incidental learning phase. To manipulate encoding depth, this phase was divided into two parts, a gender-categorization task (shallow encoding condition) and an intelligence-rating task (deep encoding condition). The order of encoding tasks was counter-balanced across participants. In the subsequent recognition memory test, the study faces were intermixed with the remaining 56 new faces. To investigate the relevance of the first two test-stimulus fixations for recognition memory processes, half the faces were presented for one fixation and the other half for two fixations. This split was orthogonal to the other factors, resulting in a fully-crossed 2 (face type: old, new) × 2 (fixation number: one, two) × 2 (LOP condition: deep, shallow) within-participants design. The assignment of face stimuli to conditions was counter-balanced across participants.

Apparatus

The eye movements of the participants' dominant eye were recorded with an SMI Hi-Speed Eye-Tracker with a sample rate of 500 Hz and a spatial resolution of 0.01° (manufacturer information). To manipulate test-stimulus presentation times according to the number of executed fixations, the on-line information of terminated fixation events provided by the eye-data recording software iView X^{TM} Hi-Speed were used. The default parameters for fixation detection were used, with a maximal dispersion of 100 pixels (dispersion = [max(x) - min(x)] + [max(y) -min(y)]) and minimum fixation duration of 80 ms. Stimuli were presented with a Windows-based computer on a 17" monitor with a resolution of 1280 × 1024 pixels and a refresh rate of 75 Hz, using the experimental software PsychoPy (Peirce, 2007). Viewing distance was 64 cm.

Procedure

The experiment consisted of four phases: an incidental learning phase with a deep and shallow encoding block; a retention interval; a recognition test; and a follow-up questionnaire. Participants arrived in the lab and signed a consent form, before a standard 13-point calibration of the eye-tracker. Participants were given the cover story that the incidental study phase was an investigation into viewing behavior during face categorization. Each trial began with a 500 ms central fixation cross. After a 50 ms blank screen a face stimulus appeared centrally for 3000 ms, followed by a response screen containing two buttons and a mouse courser. In the shallow encoding condition, the participants' task was to categorize each face as male or female; in the deep encoding condition, the task was to categorize each face as high or low in intelligence. The next trial started immediately after a response was made (see Figure 1A). To account for possible primacy and recency effects, we added four filler trials, two before and two after the experimental study trials, in each categorization task.

After a retention interval of five minutes, during which participants solved some items from a standard intelligence test (i.e., items 41 to 60 of the *Intelligenz-Struktur-Test-Screening* (Liepmann, Beauducel, Brocke, & Nettelnstroth, 2012), an unexpected recognition test followed. Participants were informed that faces of the preceding study phase, intermixed with lures, would appear in the test. Their task was to make an old/new decision for each face. Subsequent to "old" responses, participants were asked for an additional remember/know/guess judgment (Gardiner, & Richardson-Klavehn, 2000; Tulving, 1985). Participants were instructed to give a "remember" response if they could recollect any aspect of the face occurring in the study phase, such as the position of the face in the sequence (i.e., early or late) or any thoughts they had when first viewing the face. If the face evoked familiarity in the absence of conscious recollection, they were instructed to give a “know” response. Whenever they had simply guessed the face to be “old”, they were asked to give a “guess” response. To familiarize participants with the procedure of the recognition test, there were eight practice trials (the four old items were taken from the four filler trials).

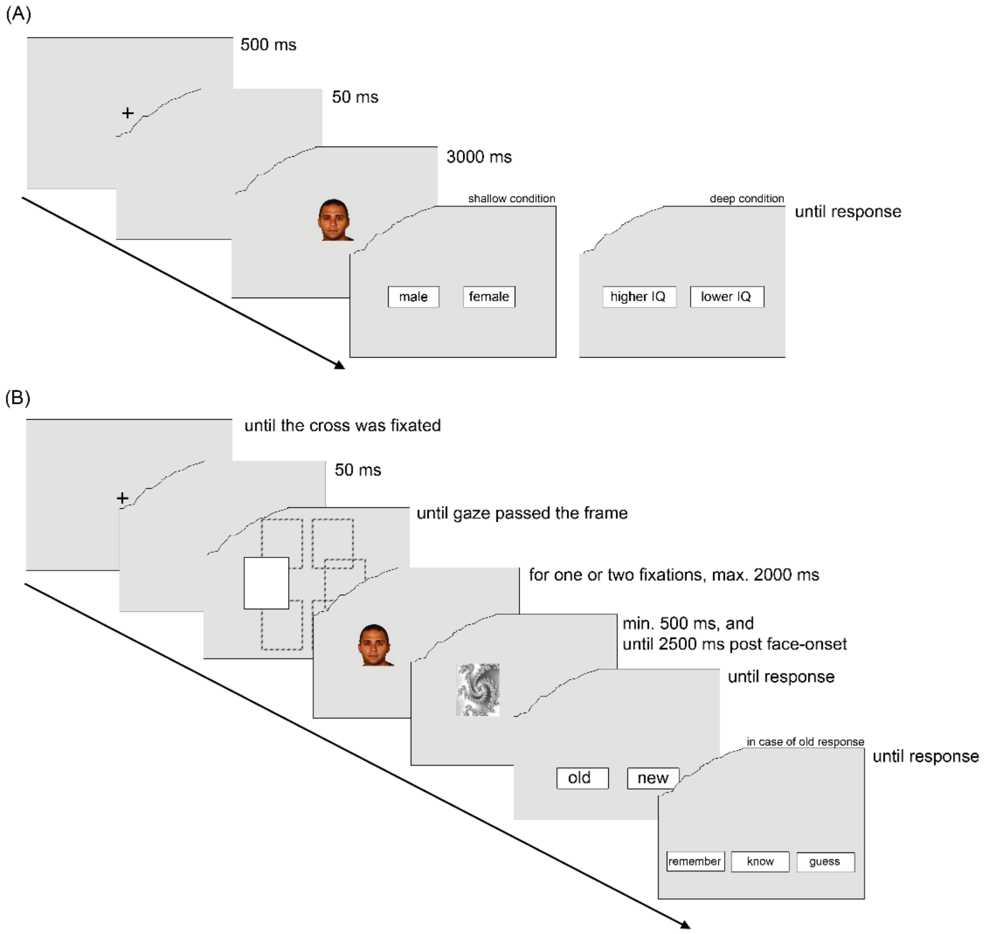

Figure 1. An illustration of the trial sequence in the incidental study phase (A) and the recognition test (B).

The procedure of the recognition test was the same as in Schwedes and Wentura (2019, Exp. 1). A trial started with a central fixation cross. When it was fixated the experimenter pressed the space bar (if a drift correction was needed, it was implemented at this stage). After a 50 ms blank screen, a frame appeared randomly in one of six possible locations, indicating where the face would appear (see Figure 1B). Participants were instructed to direct their gaze inside the frame to see the face. As soon as gaze was detected to fall inside the frame, the face was presented. Thus, parafoveal pre-processing of the face was precluded. The face remained on screen until the end of the last permitted fixation for the particular trial (i.e., one or two) had been registered on the face. The maximum presentation time was set to 2,000 ms to avoid abnormal presentation times in the case of staring. The face was then replaced by a mask to destroy the retinal afterimage. The mask stayed on screen until 2,500 ms post stimulus onset, with a minimum duration of 500 ms. Subsequently, two response buttons and a mouse cursor appeared, and participants' task was to indicate if the just presented face was "old" (studied) or "new". In case of an "old" response, a new screen with three response buttons appeared and participants made a remember/know/guess judgment. The blank-screen inter-trial interval was 50 ms (see Figure 1B).

Data Preparation

The overall recognition memory performance was assessed with the discrimination measure *Pr*, calculated by subtracting the probability of an "old" response to a new face, *p*(false alarm [FA]) from the probability of an "old" response to a study face, *p*(hit). Since we had no block-wise testing for deeply and shallowly encoded items, the same FA rate was subtracted from deeply and shallowly encoded stimuli correctly classified as "old".[1] Thus, the difference in *Pr* between the LOP conditions represents the difference in hits.

---

[1] Table A1 (Online Resource 1) reports all hit and FA probabilities separately for each fixation condition.

The corresponding measure for the response bias, *Br*, was calculated by dividing *p*(FA) by 1 – (*p*[hit] – *p*[FA]) (see Snodgrass & Corwin, 1988). We used the correction introduced by Snodgrass and Corwin (1988) to deal with the problem of a division by zero in some cases when calculating the response bias. That is, hit, FA, CR, and miss probabilities were calculated by adding 0.5 to the numerator and 1 to the denominator. To be consistent, we applied the correction to all data reported here, as recommended by Snodgrass and Corwin.

To analyze the underlying recognition memory processes, we estimated the probabilities of “remember”, “know”, and “guess” responses that followed hits, separately for each cell of the 2 (fixation number) × 2 (LOP condition) design, as well as the probabilities of the different responses that followed false alarms, separately for each fixation condition. For example, the probability of a correct “old” response to a deeply encoded face followed by a “remember” response was calculated by dividing the number of deeply encoded old faces that attracted a “remember” response by the total number of deeply encoded old faces.

As in our previous study (Schwedes & Wentura, 2019), the probability of a “remember” response to an old item (corrected for FAs) was used as an estimate of recollection-based recognition performance. Correspondingly, we decided to use the probability of a “know” response to an old item (corrected for FAs) as an estimate of familiarity-based recognition performance, rather than using the independence remember/know (IRK) method introduced by Yonelinas and Jacoby (1995). This decision was based on the known inability of the IRK method to produce reliable estimates of familiarity under experimental conditions that primarily affect recollection with only a small or no effect on familiarity—as in the present case (see, e.g., Gardiner & Richardson-Klavehn, 2000; Richardson-Klavehn et al., 1996, for a detailed discussion of this point).

Results

The data and scripts for data aggregation and data analysis are available via OSF (https://osf.io/tv3j6/?view_only=0557275b85ba4080b5b58a06b81b3b9e). We discarded trials that contained a blink during test-stimulus presentation (0.5 % of all trials in the test phase), trials without fixations on the face stimulus (1.4 % of the remaining trials in the test phase), as well as trials with incomplete second fixations, that is, where the summed duration of the first and second fixations exceeded 2,000 ms (i.e., where the input was replaced by the mask during the second fixation; 0.2 % of all trials in the test phase). Based on these criteria, a total of 2.0 % of trials were excluded. The mean duration of first and second fixations across allowed number of fixations, executed fixation, stimulus Type and response are listed in Table 1. Unless otherwise noted, all effects referred to as statistically significant throughout the text are associated with *p* values less than .05, two-tailed. Results regarding the locations of first and second fixations are reported in *Appendix A*.

Table 1
Mean duration of first and second fixations (*SD* in parentheses) across allowed number of fixations, executed fixation, stimulus Type and response.

			Response			
Allowed Fixations	Executed Fixation	Stimulus Type	new	old		
				remember	know	guess
One	First	old-deep	170 (32)	366 (245)	221 (87)	174 (39)
		old-shallow	242 (104)	259 (124)	264 (125)	210 (132)
		new	246 (60)	316 (272)	201 (61)	203 (94)

Two	First	old-deep	219 (143)	237 (76)	200 (60)	210 (70)
		old-shallow	245 (119)	225 (74)	209 (119)	193 (68)
		new	241 (64)	219 (67)	254 (164)	233 (131)
	Second	old-deep	715 (394)	876 (257)	775 (295)	565 (519)
		old-shallow	824 (360)	1007 (327)	809 (334)	734 (517)
		new	812 (169)	578 (362)	874 (340)	658 (418)

Overall Recognition Memory Performance

Analyses of variance. Figure 2 shows mean recognition memory performance, *Pr,* as a function of fixation number and LOP condition. We ran a 2 (fixation number: one vs. two) × 2 (LOP condition: deep vs. shallow) repeated measures ANOVA with *Pr* as the dependent variable to check whether we observe the common pattern of better recognition performance after deep compared to shallow encoding and, most important, whether this effect is moderated by the number of allowed fixations. The analysis yielded a significant main effect of fixation number, $F(1,30) = 75.47$, $p < .001$, $\eta_p^2 = .716$, indicating that performance increased when two test-stimulus fixations were allowed compared to only one. The main effect of LOP condition was significant as well, $F(1,30) = 31.13$, $p < .001$, $\eta_p^2 = .509$. As expected, recognition performance for deeply encoded faces was

Table 2
Mean probability (*SD* in parentheses) of hits and FAs as a function of fixation condition.

	Allowed Fixations	
	One	Two
p(hit_{deep})	.63 (.15)	.83 (.12)
p($hit_{shallow}$)	.52 (.18)	.64 (.15)
p(FA)	.28 (.13)	.22 (.13)
p($hitR_{deep}$)	.32 (.15)	.57 (.17)
p($hitK_{deep}$)	.27 (.13)	.26 (.12)

p($hitG_{deep}$)	.12 (.09)	.07 (.06)
p($hitR_{shallow}$)	.16 (.10)	.32 (.16)
p($hitK_{shallow}$)	.31 (.15)	.29 (.12)
p($hitG_{shallow}$)	.12 (.07)	.10 (.07)
p(*FAR*)	.04 (.04)	.04 (.03)
p(*FAK*)	.17 (.10)	.15 (.09)
p(*FAG*)	.11 (.10)	.07 (.07)

Note: $p(hit\_R_{deep})$ is for example the abbreviation for the probability of a correct "old" response to a deeply encoded face followed by a "remember" response.

better than for shallowly encoded faces. In line with our hypothesis, the effects were qualified by a fixation number × LOP condition interaction, $F(1,30) = 4.11$, $p = .052$, $\eta_p{}^2 = .120$.[2] The LOP effect was stronger if two fixations were allowed compared to only one fixation (see Figure 2). Both LOP effects were significant, $t(30) = 5.71$, $p < .001$, $d_Z = 1.02$ for two fixations, and $t(30) = 3.28$, $p = .003$, $d_Z = 0.59$ for one fixation. The mean probability of hits and FAs as a function of fixation condition are listed in Table 2. The analyses of the response bias, *Br*, can be found in Appendix B.

---

[2] Note that this interaction test is equivalent to a paired samples *t*-test (with t = squareroot(F)) that compares the dependent variable "deep advantage" (i.e., the difference between *Pr* for deeply encoded items minus *Pr* for shallowly encoded items) for one and two fixations; this *t*-test, however, allows for adequate one-tailed testing, $t(30) = 2.03$, $p = .026$ (one-tailed), $d_Z = .36$.

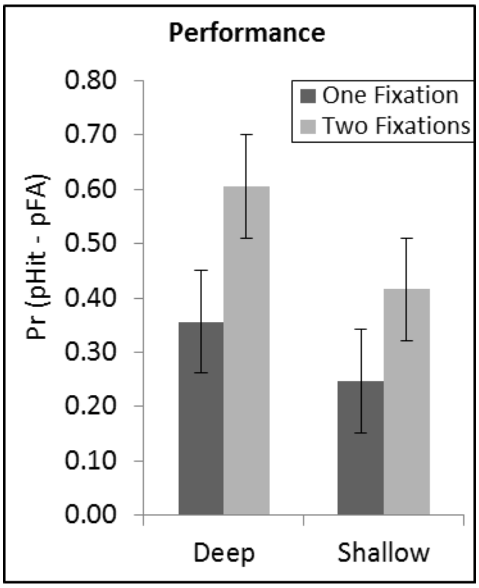

Figure 2. Performance (in *Pr*) across fixation-number and levels-of-processing conditions. Error bars are 95 % within-subject confidence intervals (Jarmasz & Hollands, 2009) for the interaction effect.

Linear mixed model analyses. The number of allowed fixations may influence recognition memory performance not only through the amount of encoding time but also through the amount of input information (since the second fixation will be on a different part of the stimulus than the first fixation). To disentangle these influences we ran a multilevel logistic regression using the function *glmer* of the *lme4* library of R (Bates, Maechler, Bolker, & Walker, 2015). In all reported analyses, the participants were classed as random factor. We report random-slopes model (see Barr, Levy, Scheepers, & Tily, 2013). We dummy-coded old-shallowly encoded (D_1) and old-deeply encoded items (D_2) with the reference condition of new items. In a first step of the hierarchical regression, we predicted the old/new item response from (a) the objective old/new status of the item (i.e., by D_1 and D_2), (b) the number of allowed fixations (i.e. , the amount of input information), and (c) the interaction terms (i.e., D_1 × number of fixations and D_2 × number of fixations). As can be seen in the left panel of Table 3 both D_1 and D_2 had a positive and significant weight, showing that “old”/”new” responses, to a large extent, reflect the objective old/new status of the items. Both interaction terms were significant as well, indicating that the basic weights of D_1 and D_2 increased with two fixations (and decreased with only one fixation since number of fixations was a centered variable). This increase was stronger for deeply encoded items compared to shallowly encoded items. Thus, this result corroborates the results of our conventional analysis.

In a second step, we added total presentation duration (i.e., the amount of viewing time: the duration of the first fixation for the one-fixation condition and the sum of first and second fixation durations for the two-fixation condition) and the corresponding interaction terms with D_1 and D_2 as additional predictors. If number of fixations influences recognition performance only through the extension of encoding time, the interaction terms of D_1 and D_2 with fixation number should no longer be significant, whereas the interaction terms of D_1 and D_2 with total duration should be significant.[3] If, however, the number of fixations influences recognition performance also through the amount of input information, the interaction terms of D_1 and D_2 with fixation number should remain significant even if they “compete” with the interaction terms of D_1 and D_2 with duration. As can be seen in the right panel of Table 3, we observed the latter: the interaction terms of D_1 and D_2 with fixation number were still significant. However, the levels of processing effect was partially due to duration, as can be seen (a) in the decreased weight for the interaction term of D_2 and fixation number in the second analysis relative to the first, and (b) in the significant weight for the

---

[3] The correlation between total duration and number of fixations was *r* = .56. That is, the degree of collinearity is not so extreme that it would preclude this analysis from the outset.

interaction term of D_2 and duration. The weight for the interaction term of D_2 and fixation number was, however, still numerically larger than the weight for the interaction term of D_1 and fixation number.

Table 3
Results of the multilevel logistic regression analysis with item-related response (“old” = 1 vs. “new” = 0) as the dependent variable.

	Step 1				Step 2			
Fixed Factor	Weight	*SE*	*z*	*p*	Weight	*SE*	*z*	*p*
Intercept	-1.215	0.123	-9.87	< .001	-1.227	0.125	-9.84	< .001
D1 (shallow)	1.559	0.147	10.63	< .001	1.576	0.147	10.69	< .001
D2 (deep)	2.488	0.162	15.33	< .001	2.568	0.171	15.03	< .001
NFix	-0.203	0.076	-2.67	.008	-0.154	0.092	-1.67	.095
Duration					-0.107	0.093	-1.15	.250
D1 × NFix	0.457	0.101	4.53	< .001	0.397	0.123	3.22	.001
D2 × NFix	0.801	0.125	6.43	< .001	0.586	0.153	3.83	< .001
D1 × Duration					0.140	0.143	0.98	.329
D2 × Duration					0.484	0.192	2.52	.012

Note. D1: shallow = 1, deep = 0, new = 0; D2: deep = 1, shallow = 0, new = 0; NFix: fixation number, with one fixation = -1, two fixations = 1; duration = total presentation duration (z-standardized)

Recognition Memory Processes: Familiarity and Recollection

To test our focal hypothesis—that the effect of the LOP manipulation on recollection-based recognition should be more pronounced if two test-stimulus fixations were allowed rather than just one—we ran a 2 (fixation number: one vs. two) × 2 (LOP condition: deep vs. shallow) repeated measures ANOVA with Pr_{remember} values, calculated as $Pr_{\text{remember}} = p(\text{hit}_{\text{remember}}) - p(\text{FA}_{\text{remember}})$, as the dependent variable. The analysis revealed significant main effects of fixation number, $F(1,30) = 78.45$, $p < .001$, $\eta_p^2 = .723$, as well as LOP condition, $F(1,30) = 50.22$, $p < .001$, $\eta_p^2 = .626$. Most important, main effects were qualified by a Fixation Number × LOP condition interaction, $F(1,30) = 8.46$, $p = .007$, $\eta_p^2 = .220$. There were significant LOP effects in Pr_{remember} values for both the two-fixation condition, $t(30) = 6.81$, $p < .001$, $d_Z = 1.22$, and the one-fixation condition, $t(30) = 5.33$, $p < .001$, $d_Z = 0.96$. As hypothesized and indicated by the interaction effect, the LOP effect in recollection-based recognition was larger if two fixations were allowed compared to only one fixation.

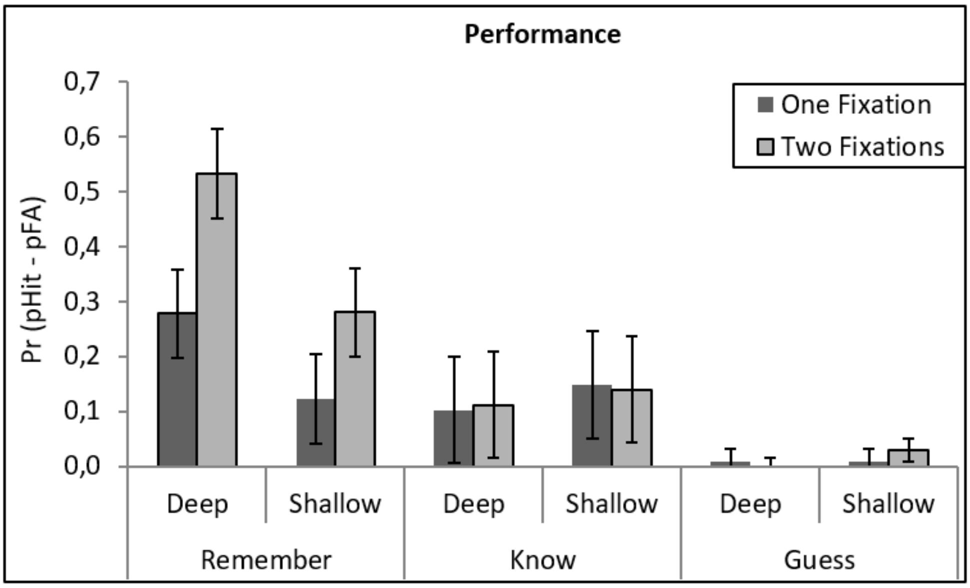

Figure 3. Performance across fixation-number, levels-of-processing conditions and remember/know/guess responses. Error bars are 95 % within-subject confidence intervals (Jarmasz & Hollands, 2009) for the interaction effect.

Regarding familiarity-based recognition, a 2 (fixation number: one vs. two) × 2 (LOP condition: deep vs. shallow) repeated measures ANOVA with Pr_{know} values, calculated as $Pr_{\text{know}} = p(\text{hit}_{\text{know}}) - p(\text{FA}_{\text{know}})$ as the dependent variable yielded no significant effects, $F < 1$ for fixation number, $F(1,30) = 2.66$, p = .113 for LOP condition, and $F < 1$ for the interaction (see Figure 3). Thus, familiarity-based recognition did not differ for deeply and shallowly encoded items, or across fixation conditions.

Discussion

The aim of this experiment was to provide evidence that second fixations of a to-be-remembered stimulus are especially relevant for recollection-based but not familiarity-based recognition. We addressed this by deep versus shallow encoding conditions to experimentally manipulate the probability of item recollection. This allowed us to compare the influence of the first two fixations across experimentally created conditions that facilitated recollection-based recognition (i.e., deep encoding) or familiarity-based recognition (i.e., shallow encoding).

First of all, our results fulfill the basic preconditions for our central hypothesis: (1) In line with previous findings (see Craik & Lockhart, 1972; Craik & Tulving 2004), we found better overall recognition performance for deeply compared to shallowly encoded items. (2) In addition, performance increased if two fixations (compared to only one fixation) were allowed (see Hsiao & Cottrell, 2008; Schwedes & Wentura, 2019). Given this backdrop, we found the higher performance for deeply encoded items was stronger in the two-fixation condition. Most important and in line with our central hypothesis, this interaction effect was reflected in the rate of “remember” (i.e. recollection-based) responses (corrected for false alarms) without an effect in the rate of “know” (i.e. familiarity-based) responses.

A huge range of previous studies have shown that deep encoding especially results in more recollection-based recognition (see Yonelinas, 2002, for review). One criticism with regard to the study by Schwedes and Wentura (2019) was that the study does not rule out that high item familiarity has been the key process that caused the increased rate in “remember” responses with two allowed fixations. However, if that would have been valid, the LOP manipulations should not have moderated the remember rate depending on the allowed numbers of fixations.

This pattern of results supports the notion that “remember” responses reflect especially recollection-based recognition, as has been shown by Evans and Wilding (2012), and underpins previous assumptions, that second fixations are especially relevant for recollection but not familiarity. In addition, the present study rules out a further alternative interpretation of our previous findings (Schwedes & Wentura, 2019), namely that some facial stimuli may tend to evoke “remember” responses and also – accidentally – happen to be associated with an increase in memory performance from the first to the second fixation. Such an interpretation in terms of an item effect is ruled out by the present study, as recollection rate was varied through an experimental manipulation in a balanced design.

There are two possible explanations for why second fixations are especially relevant for recollection-based recognition. One explanation is time-based. Dual-process models of recognition memory assume that the computation of recollection takes longer than the computation of familiarity. This is supported by the temporal occurrence of event-related potentials (ERPs) associated with familiarity (the early mid frontal old/new effect, about 300-500 ms post stimulus onset) and recollection (the parietal old/new effect, about 400-800 ms post stimulus onset; for a review see Rugg & Curran, 2007). It follows that the “second fixation advantage” may be due to the simple fact that the input is available for longer in the two-fixation condition compared to the one-fixation condition.

However, previous results tend to contradict this possible explanation. Hsiao and Cottrell (2008) found lower recognition memory performance in a two-fixation condition that provided the same input during first and second fixations compared to a two-fixation condition that provided input from two gaze locations. Schwedes and Wentura (2019) replicated this finding and found that this effect was especially pronounced for recollection-based recognition. Further evidence comes from the linear mixed model analyses of the present study: Fixation number remained a significant moderator of recognition performance even if total fixation duration was entered as a competitive moderator.

Thus, from a dual-process model perspective the role of the first two eye fixations might be as follows: the information provided by the first fixation activates stored representations that overlap with the incoming information. If no further fixations are possible (as in the one-fixation condition of the present study), it is assessed whether the “resonance” of the input with a stored representation, which results in a familiarity signal, exceeds a decision criterion or not. If so, a familiarity-based old response is given. However, if a second fixation can be directed to a different stimulus part, the memory representation that is maximally activated after the first fixation can be considered a hypothesis regarding the identity of the viewed stimulus. The second fixation is then a test of this hypothesis, to verify the identity of the stimulus and obtain access to contextual information related to that stimulus. For deeply encoded items, the stored memory representation might be more complete and contains more details than for shallowly encoded ones. Therefore, the possibility that the input of a second fixation (that has a higher uniqueness for deeply encoded items) can confirm the stimulus-identity hypothesis produced on the basis of the first fixation should be higher, as the input will map onto more details of the stored representation.

A more general perspective to interpret the benefit of recollection from second fixations of deep encoded items that is also compatible with single-process accounts is as follows: For deeply encoded items the input of second fixations can be seen as a retrieval cue with higher specificity. Since retrieval cues with higher specificity especially improve memory performance of deep encoded items (see Moscovitch & Criak, 1976), we can see a stronger second fixation advantage for deeply compared to shallowly encoded faces[4].

The notion that the conditions relevant for recollection to occur (i.e., information provided by second fixations) are met later in time, could also explain the later occurrence of recollection-related ERPs. The mean duration of first and second fixations to familiar faces in our earlier study (Schwedes & Wentura, 2012)—where eye movements were recorded while participants viewed studied faces that were intermixed with new faces under recognition conditions—were 266 ms (76 ms *SD*) and 412 ms (162 ms *SD*), respectively. Thus, second fixations terminated on average 678 ms post stimulus onset. This maps onto the time window of the parietal old/new effect associated with recollection (approx. 400-800 ms; Rugg & Curran, 2007) reasonably well. Thus, recollection may not occur later because the process itself is slow, but because the requirements for recollection-based processing are fulfilled later.

Conclusion

In summary, beyond the well-known effect of memory on eye movement behavior, the present study provides further evidence for the

---

[4] We thank the anonymous Reviewer 2 for his hint to this alternative interpretation.

reversed relation: The possibility to execute a second fixation to a stimulus seems to be especially relevant to recollect that stimulus. These findings suggest that a full-fledged theory of recognition should incorporate the interplay between memory activations and eye-gaze behavior. On a more pragmatic note, these findings should be considered in research that is interested in recollection but uses experimental settings that might affect the execution of a second stimulus fixation, like constrained presentation times.

Compliance with Ethical Standards

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

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

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*, 751-757.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684

Craik, F. I., & Tulving, E. (2004). Depth of processing and the retention of words in episodic memory. In D.A. Balota & E.J. Marsh (Eds.), *Cognitive Psychology: Key Readings in Cognition* (pp. 296-308). New York: Psychology Press

Evans, L. H., & Wilding, E. L. (2012). Recollection and Familiarity Make Independent Contributions to Memory Judgments. *The Journal of Neuroscience, 32*, 7253–7257

Faul, F., Erdfelder, E., Lang, A.-G, & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 229–224). New York: Oxford University Press

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4,* 1-16.

Hsiao, J. H., & Cottrell, G. (2008).Two fixations suffice in face recognition. *Psychological Science, 19,* 998-1006.

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63,* 124–138.

Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia, 50,* 3080–3093.

Liepmann, D., Beauducel, A., Brocke, B. & Nettelnstroth, W. (2012). *Intelligenz-Struktur-Test – Screening – (IST-Screening)*. Manual. Göttingen: Hogrefe.

Mäntylä, T., & Holm, L. (2006). Gaze control and recollective experience in face recognition. Visual Cognition, 13, 365–386.

Marinkovic, K., Oscar-Berman, M., Urbam, T., O’Reilly, C. E., Howard, J. A., Sawyer, K., & Harris, G. J. (2009).Alcoholism and dampened temporal limbic activation to emotional faces. *Alcoholism, Clinical and Experimental Research, 33,* 1880–1892.

Mueller, J. H., Bailis, K. L., Goldstein, A. G. (1979). Depth of processing and anxiety in facial recognition. *British Journal of Psychology, 70,* 511–515

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8–13.

Richardson-Klavehn, A., Gardiner, J. M., & Java, R. I. (1996). Memory: Task dissociations, process dissociations and dissociations of consciousness. In

G. Underwood (Ed.), *Implicit Cognition* (pp. 85–158). New York: Oxford University Press.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*, 251–257.

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15,* 508-525.

Schwedes, C., & Wentura, D. (2012). The revealing glance: eye gaze behavior to concealed information. *Memory & Cognition, 40,* 642-651.

Schwedes, C., & Wentura, D. (2016). Through the eyes to memory: Fixation durations as an early indirect index of concealed knowledge. *Memory & Cognition, 44,* 1244-1258.

Schwedes, C., & Wentura, D. (2019). The relevance of the first two eye fixations for recognition memory processes. *Memory.* Advance online publication. doi.org/10.1080/09658211.2019.1567789

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology, 117*, 34-50

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.

Wig, G. S., Miller, M. B., Kingstone, A., & Kelley, W. M. (2004). Separable routes to human memory formation: Dissociating task and material contributions in the prefrontal cortex. *Journal of Cognitive Neuroscience, 16*, 139-148.

Wixted JT, Mickes L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117,* 1025–1054

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 20*, 1341–1354.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34*, 622–643.

Author note

Charlotte Schwedes and Dirk Wentura, Saarland University, Faculty of Human and Business Sciences, Department of Psychology, Germany. Demian Scherer, University of Münster, Institute for Psychology in Education, Germany

We thank Ullrich Ecker for valuable comments on an earlier version of this manuscript. Correspondence concerning this article may be addressed to Charlotte Schwedes, Saarland University, Faculty of Human and Business Sciences, Department of Psychology, Building A2.4, D-66123 Saarbrücken (e-mail: c.schwedes@mx.uni-saarland.de)

Appendix A

Below we provide information about the loci of first and second fixations. Note, since these analyses were not the focus of the reported study, the used materials were not normed, that is, there was small variance in the exact position of the eyes and the nose. To report the loci of first and second fixations, we used a normalization procedure that post hoc minimized the differences in fixation locations due to the differences in the used material.

Normalization procedure

In a first step, we generated an average face of all faces, by averaging each pixel over all faces. We then took the x and y position of the right eye, the left eye, and the tip of the nose from this average face (target face) and from each of the faces used in the experiment (hereinafter source faces). To norm the fixation locations (x and y position) of the source face, we calculated four scaling factors for each source face. One for the distance in the y-dimension between left eye and the tip of the nose, one for the distance in the y-dimension between right eye and the tip of the nose, one for the distance in the x-dimension between left eye and the tip of the nose, and one for the distance in the x-dimension between right eye and the tip of the nose. Each scaling factor was computed by the relation of the respective distance in the source face and the target face. The x and y values of each fixation were then multiplied by the corresponding scaling factor. For example, for the x position of a fixation that landed to the right of the nose, the x scaling factor for the right side (for the distance in the x-dimension between right eye and the tip of the nose) was used and vice versa. We used these normalized data for visualization and analyzes of the loci.

Results

As can be seen from Figure A1, first fixations (plot B) are more distributed than second fixations (plot C). Plot A also shows, that the typical locations of a first and a second fixation differ between individuals.

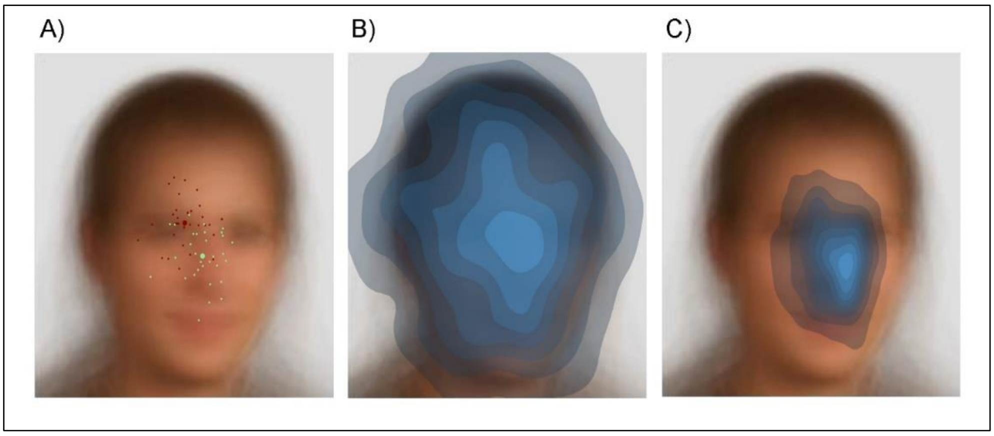

Figure A1. Plot A shows the grand mean (big dot) of the location of the first (red) and second (green) fixation and the mean locations of first and second fixations of each participant (small dots). Plot B shows the 2d density distribution of all single first fixations and plot C the 2d density distribution of all single second fixations.

To analyze whether the x and y dimensions between first and second fixations differ significantly, we run two multilevel linear regression (using the lmer function of the lme4 package). The difference of the x-dimension and y-dimension values, respectively, of first and second fixation seved as the dependent variable (comparable to Hsiao & Cottrell, 2008), participants as random factor, and the test for the regression constant as the decisive result. The results show that first and second fixations differ significantly in the x ($t = 7.87$, $p < .001$) and y ($t = -11.16$, $p < .001$) dimension. Second fixations are more right (closer to the nose) and lower (more in the center of the face) than first fixations.

In addition, we analyzed, whether the (Euclidian) distance between first and second fixation is predictive for, first, the correctness of the response after two allowed fixations and, second, for the probability of a remember (vs. know) response.

Regarding the correctness of the response after two allowed fixations, we ran a multilevel logistic regression (using the glmer function of the lme4 package; with participants as random factor incl. random slopes) with the correctness of the response as the dependent variable and the objective old/new status of the face as well as the distance between first and second fixation as the predictors with participants as random factor. The results show no influence of the distance between first and second fixations on the correctness of the response (main effect distance: z= 1.142, p = .253; interaction effect: z= -0.051, p = .959).

Regarding the response type (remember vs. know response), we ran a multilevel logistic regression (using the glmer function of the lme4 package; with participants as random factor incl. random slopes) with the type of response (remember=1, know=0) as the dependent variable and the distance as the predictor with participants as random factor. Only trials with old stimuli were analyzed. The results show no influence of the distance between first and second fixations on the response type (main effect distance: z= 0.75, p = .454). These results show that the distance between first and second fixations is not predictive for the correctness of the response and not predictive for the remember /know distinction.

To provide information whether second fixations land closer to diagnostic face parts for deeply encoded faces, we, first, ran two multilevel linear regressions to test whether locations for deeply and shollowly encoded faces differ at all: One with the x position as the dependent variable and a second with y positions as the dependent variable and in both models LOP (deep=1 and shallow=-1) as the predictor and participants as random factor. Regarding the x coordinates, LOP had no significant influence, t=-0.36, p=.724. The same was the case for the y dimension, t=1.49, p=.137. Thus, the location of second fixations do not differ for deeply and shallowly encoded items. Hence, by inference it cannot be the case that second fixations land closer to diagnostic face parts for deeply encoded faces.

In a last analysis we looked whether second fixation locations (x and y coordinates) differ between „remember“ and „know“ responses. Therefore, we again ran two multilevel linear regressions (using the lmer function of the lme4 package). One with the x position as the dependent variable and a second one with y positions as the dependent variable and in both models the remember/know response (remember=1 and know=-1) as the predictor and participant as random factor. Regarding the x coordinates, the remember/know response had no significant influence, t=0.95, p=.350. The same was the case for the y dimension[5], t=1.17, p=.242. Thus, the locations of second fixations do not differ between remember and know responses.

---

[5] This model resulted in a singular fit. Therefore, we reduced the number of variance-covariance parameters by running a random intercept model. The results were essentially the same (t=1.21, p=.226).

Appendix B

Since the LOP effects represent the difference in the hit rates (see *data preparation*), we only report effects in responses bias that include the factor fixation number. A 2 (fixation number: one vs. two) × 2 (LOP condition: deep vs. shallow) ANOVA for repeated measures with *Br* as the dependent variable yielded a main effect of fixation number that missed the criterion of significance, $F(1,30) = 3.29$, $p = .080$, $\eta_p^2 = .099$ as well as a significant fixation number × LOP interaction effect, $F(1,30) = 11.57$, $p = .002$, $\eta_p^2 = .278$. For deeply encoded items, the bias became more liberal when two fixations were allowed, $t(30) = 3.00$, $p = .005$, $d_Z = .54$. That is, after deep encoding participants are more cautious in responding “old” if only one stimulus information is available. With an additional input their reservation to respond “old” declines. This was not the case for shallowly encoded items, $t(30) = 0.15$, $p = .882$.