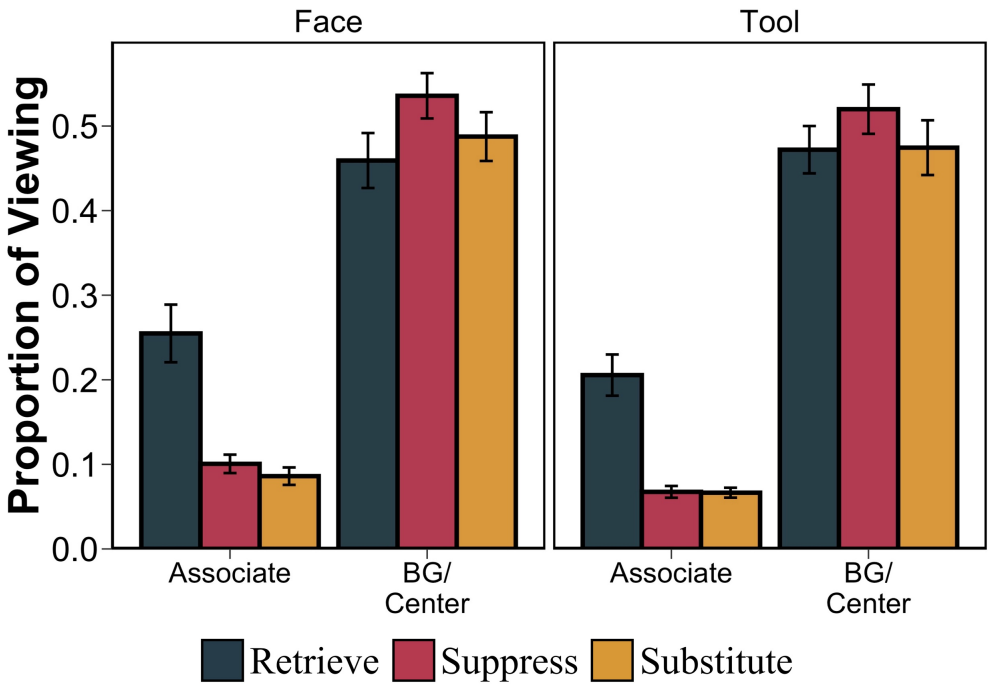
Supplementary Results

Background/Center and Associate Viewing

Comparisons were reported in the manuscript (see "Associate Viewing Relative to Trial-Specific Chance" in the Results Section) that examined whether there were differences in the proportion of total viewing time directed to the Background (including the dots) and Center AOIs across category types and experimental conditions. A figure illustrating these results with data subdivided by category, condition, and AOI type is provided below. As reported in the manuscript, results indicated that a greater proportion of total viewing time was directed to the BG/Center AOI in the suppress conditions than in the retrieve and substitute conditions, which themselves, were not statistically different.

Supplementary Figure 1.



Note. Error bars represent standard error of the mean. BG: Background, which includes the task-relevant dots in the search display.

Proportion of Total Viewing Time Directed to Prioritized Objects

As described in the manuscript (see "Viewing-Based Prioritization" in the Results Section), a single object viewed most often when search displays were presented was identified for every trial as the "prioritized object". This object was then tagged as either the "associate", a non-associate object from the "same category", or a non-associate object from the "other category".

Results from our prioritization analysis, using trial counts, indicated that associates were most often prioritized in the retrieve condition, and that non-associate other-category objects were most often prioritized in the substitute condition. Notably though, it was the case that other-category objects were also prioritized more often than same-category objects in the suppress condition. This means that the *overall pattern* in the data (i.e., lowest number of trials with a prioritized associate, intermediate number of trials with a prioritized same-category object, and highest number of trials with a prioritized other-category object) was similar for suppress trials and substitute trials. One possibility then, is that even in the suppress condition, participants were calling to mind a specific other-category object, contrary to instructions. That said, even though the general pattern was the same for suppress and substitute trials when the dependent variable was the *number* of trials in which a certain object was prioritized, it is possible that the *amount* of viewing time directed to prioritized same- and other-category objects is different between conditions. For instance, it may be the case that, even though a non-associate other-category object was most often prioritized in both the suppress and substitute conditions, more time is spent viewing the prioritized other-category object when substitution is required because participants are actively calling to mind a specific object from the other category (i.e., a specific person, when the associate is a tool) and holding it in mind. Such an outcome would be consistent with the possibility that eye-movements are disproportionately attracted to visual inputs that match what is being actively represented when the search display is presented.

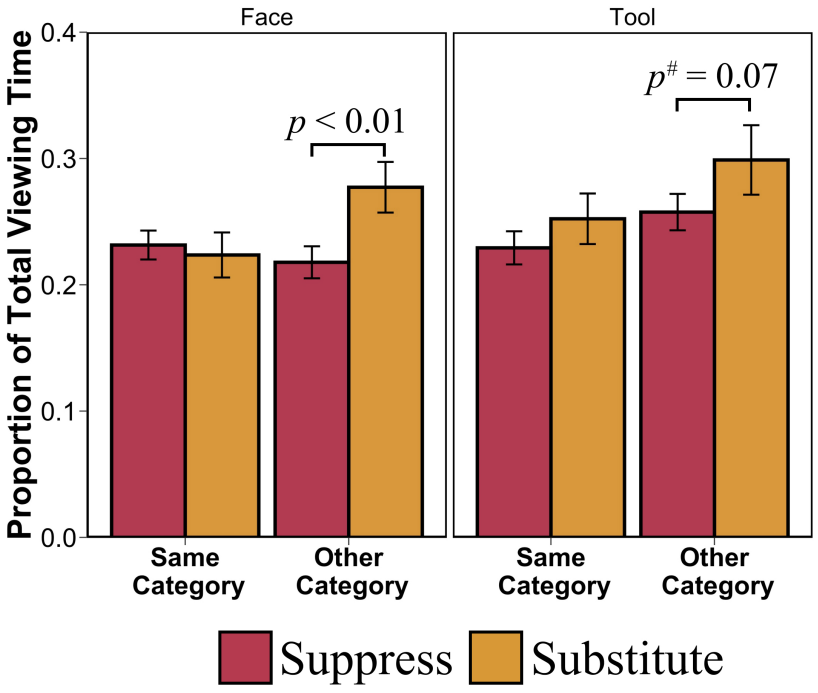
To test this possibility, we examined the proportion of total viewing time directed to prioritized non-associate same- and other-category objects in the suppress and substitute conditions using a repeated-measures ANOVA with factors Condition (Suppress, Substitute), Associate Category (Face, Tool), and Object Type (Same Category, Other Category). Two participants were excluded from this analysis because there were no trials of a given type in one of the condition-object type bins (e.g., for one participant, there were no substitute trials with face associates in which a same-category object was prioritized). Results from the ANOVA indicated that there were significant main effects of Condition, Category, and Object Type, F's ≥ 5.09, p's < 0.05, ηp^2's ≥ 0.18. There was also a significant interaction between Condition and Object Type, $F(1, 23) = 4.65$, $p = 0.04$, $\eta p^2 = 0.17$, and a significant 3-way interaction, $F(1, 23) = 5.71$, $p = 0.03$, $\eta p^2 = 0.20$. The remaining two-way interactions were not significant, F's ≤ 1.83, p's ≥ 0.19, ηp^2's ≤ 0.07.

To unpack the 3-way interaction, FDR-corrected t-tests were calculated separately for faces and tools to determine whether there were differences in the proportion of total viewing time across conditions and object types (i.e., same-category suppress, same-category substitute, other-category suppress, other-category substitute). When the associate was a face, the proportion of total viewing time directed to other-category objects was greater for substitute than for suppress trials, $t(23) = 3.46$, $p = 0.02$, $d = 0.71$. A trend in the same direction (i.e., greater other-category object viewing in the substitute than the suppress condition) was also evident prior to correction when the associate was a tool, $t(23) = 1.90$, uncorrected $p = 0.07$, FDR-corrected $p = 0.11$, $d = 0.39$. In contrast, there was no significant difference in the proportion of total viewing time directed to prioritized same-category objects between the suppress and substitute conditions for either category, t's ≤ 1.17, p's ≥ 0.26, d's ≤ 0.24 (see Supplementary Figure 2). This indicates that in addition to being prioritized most often in the substitute condition, other-category objects also attracted more viewing time in the substitute (than the suppress)

condition. This outcome suggests that participants are actively calling to mind, and maintaining, a specific object from the other category in substitute trials but are not (or not as often) doing so in suppress trials. It is possible that sometimes, when the instruction is to suppress (i.e., push the associate out of mind), an exemplar from the other category comes to mind and affects viewing behavior (i.e., participants may not always successfully switch from one instructional set to another across trials), leading to prioritization of the other-category object. Whether and how task switching affects viewing patterns and task compliance could be examined systematically in future work. In sum, these results seem consistent with the proposal that eye-movements are sensitive to the match between active mental representations and visual stimulus information the search display.

Supplementary Figure 2

AOI-based proportion of total viewing time values.



Note. Error bars represent standard error of the mean. p: FDR-corrected p-value. $p^{\#}$: uncorrected p-value.

Final Recognition Conditionalized on Test-and-Encode Performance

In past work using the Think/No-think (TNT) paradigm, accuracy in the final memory test has been analyzed in two ways. In some studies, all trials are included in the analysis, regardless of initial learning (unconditionalized analysis; Hertel & Calcaterra, 2005; Anderson & Green, 2001). In the second approach, analyses are limited to pairs with a correct response prior to the TNT phase (conditionalized analysis; Anderson et al., 2004; Benoit et al., 2014; Detre et al., 2013; Fawcett et al., 2015). While a few studies have reported suppression-induced memory decrements in both cases (Catarino et al., 2015; Yang et al., 2020), others have found that memory in no-think trials is impaired only when using the conditionalized analysis (Hanslmayr et al., 2012). It has been suggested that suppression-induced forgetting effects may be more robust when using this approach because intentional control during the TNT phase is most effective when the encoded information readily comes to mind (Nardo & Anderson, 2023).

Hence, we repeated the analysis on final recognition performance after excluding pairs that were not correctly identified in the hybrid test-and-encode phase. A repeated-measures ANOVA on accuracy revealed a main effect of Associate Category $F(1, 25) = 46.31$, $p < .001$, $\eta_p^2 = .65$, with higher accuracy for scene-face than scene-tool pairs. However, unlike the results from the unconditionalized data, there was no longer a trend for the main effect of Condition, $F(3, 75) = 0.96$, $p = 0.42$, $\eta_p^2 = 0.04$. Additionally, there was no interaction between Associate Category and Condition, $F(3, 75) = 1.94$, $p = 0.13$, $\eta_p^2 = 0.07$. As before, we compared accuracy between conditions separately for faces and tools to examine whether regulation-induced forgetting would be more evident for scene-tool pairs. Accuracy for scene-tool pairs was lower in the substitute condition relative to the retrieve condition prior to correction, $t(25) = 2.17$, uncorrected $p = 0.04$, FDR-corrected $p = 0.35$, $d = 0.43$. No other pairwise differences were significant, t's ≤ 1.72, p's ≥ 0.35, d's ≤ 0.34 (see Supplementary Figure 3A).

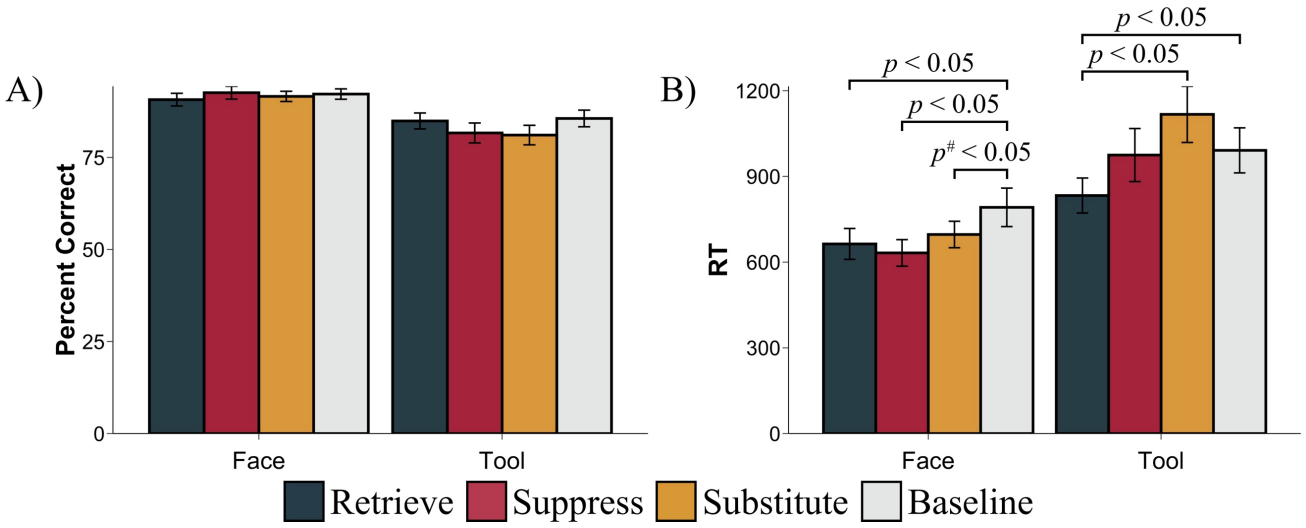
A similar conditionalized analysis on RT limited to correct trials from the post-test recognition phase, revealed main effects of Associate Category, $F(1, 25) = 78.59$, $p < 0.001$, $\eta_p^2 = 0.76$, and Condition, $F(3, 75) = 7.32$, $p < 0.001$, $\eta_p^2 = 0.23$, as well as a significant Associate Category by Condition interaction, $F(3, 75) = 3.08$, $p = 0.03$, $\eta_p^2 = .11$. For trials where the associate was a tool, participants were significantly faster at identifying the associate in the retrieve condition, relative to the baseline and substitute conditions, t's ≥ 3.00, p's $= 0.02$, d's ≥ 0.59. On the other hand, for faces, participants were significantly faster in the retrieve and suppress conditions, relative to the baseline condition, t's ≥ 2.57, p's ≤ 0.05, d's ≥ 0.50. They were also faster in the substitute, relative to the baseline condition, $t(26) = 2.29$, uncorrected $p = 0.03$, $d = 0.45$. However, this difference was not significant after correction (FDR-corrected $p = 0.07$). No other pairwise differences were significant, t's ≤ 1.83, p's ≥ 0.16, d's ≤ 0.36 (Supplementary Figure 3B).

Overall, these results indicate that when analyses exclude pairs with incorrect test-and-encode responses, there is no effect of memory regulation on final recognition accuracy. That said, it is likely that this conditionalized analysis includes just a subset of the pairs that were successfully encoded in our experiment. As a reminder, when an incorrect response was made during the test-and-encode phase, the original pair was re-presented, and participants were explicitly instructed to use this final opportunity to encode the pair. Therefore, it is highly probable that some pairs eliminated from the conditionalized analysis were encoded successfully (after participants made incorrect responses). These pairs (learned after an incorrect response), together with a subset of pairs that elicited correct responses but did not have particularly strong memory representations, may then have been driving the effects of regulation on post-test recognition performance in our original analysis. This would be consistent with observations that successful memory regulation varies as a function of memory strength (Newman & Norman, 2010; Norman et al., 2007; see also Marsh & Anderson 2022), as is outlined in our discussion section. Results

based on reaction time data were unchanged in the conditionalized analysis, although now, the significant differences survived correction for multiple comparisons. Specifically, while retrieval of scene-tool pairs in the regulation and search phase led to faster identification of the associate in the final recognition phase relative to the baseline pairs, this was not the case in the suppress and substitute conditions. On the other hand, recognition of face associates was faster for all pairs presented in the regulation and search phase, suggesting that even brief retrieval of the pairs may have strengthened memory in trials where the associate was a face.

Supplementary Figure 3.

Accuracy and RT in the post-test recognition phase conditionalized on hybrid accuracy.



Note. Error bars represent standard error of the mean. *p*: FDR-corrected *p*-value. $p^{\#}$: uncorrected *p*-value. RT: Reaction time

Post Experimental Questionnaire (PEQ)

The PEQ was included to assess whether participants were following instructions (retrieve, suppress, substitute) upon presentation of the scene cue and during the delay. Recent work suggests that

the extent to which participants comply with no-think instructions in the TNT task is an important determinant of subsequent suppression-induced forgetting (Liu et al., 2022). Thus, we computed a "non-compliance score" in the suppress condition, based on the procedure in Liu et al. (2021), using responses to the first two questions in the PEQ:

After the scene was presented I:

1. Made sure I still knew the associated face/tool first, and then tried to not think of this associated face/tool.
2. Tried not to think of the associated face/tool, but then after the trial was over, I made sure I still remembered the associated face/tool.

Participants reported frequency of these behaviors on a 5-point scale (0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Frequently, 4 = Very Frequently), and the non-compliance score was computed as the sum of the responses to these questions. A score of 0 would indicate that participants never deliberately called the associate to mind in the suppress condition, whereas a sum of 8 would indicate that they did so at both the beginning and the end of the trial. A similar score in the substitute condition was not meaningful because participants needed to call the associate to mind at the start of the trial in order to select a substitute from the other category.

Results from the suppress condition indicated that, while participants were generally following instructions, compliance was not perfect (Range: 2 – 7, M = 4.08, SD = 1.23; see Supplementary Table 1 for details). Thus, it is possible that the modest effects of suppression-induced forgetting in our experiment may have been due to some non-compliance (or difficulty with compliance) following the instructional cues.

Importantly though, it is also possible that the PEQ questions, taken from a previously published study (Hertel & Calcaterra, 2005; see also Nardo & Anderson 2023), were not sufficiently adapted to the procedures of our experiment. In contrast to the study that originally used this PEQ, participants in our experiment were constantly switching between *three* experimental conditions, and for two of these (i.e., retrieve and substitute) they *had* to retrieve the associate to comply with task instructions. It seems likely that they were also doing this on suppress trials, before actively attempting to regulate retrieval (i.e., push the target out of awareness). If they thought this approach (for suppress trials) was consistent with task demands, that would explain why several participants indicated in response to Question 1, above, that they *frequently* or *very frequently* made sure they knew the associate before attempting to suppress it (see Supplementary Table 1). That this may have been the case is suggested by responses to Question 2. Participants were much less likely to agree that they were purposely retrieving the associate at the end of the trial (see Supplementary Table 1). Therefore, whether or not responses to Question 1 indicate lack of compliance or not is difficult to say with certainty. More generally, even if these questions reasonably (and consistently) measure compliance across participants, due to low statistical power, we could not systematically compare memory performance between compliant and non-compliant participants, or exclude non-compliant participants based on the recommendations outlined in Nardo & Anderson (2023). Under these circumstances, it might be considered surprising that we saw any effects of memory regulation in our data.

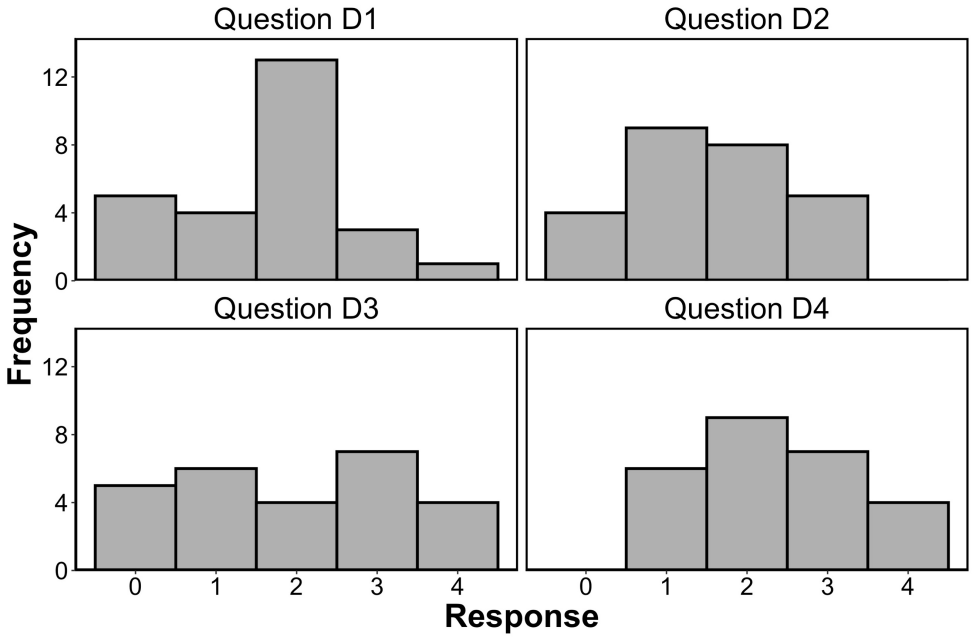
The PEQ was also used to determine if participants were deliberately directing their eyes to certain objects upon the presentation of the search display and this part of the survey was prepared explicitly for the purposes of the current study. The instructions given to participants stressed that the cue (retrieve, suppress or substitute) only pertained to the object they were to call to mind following the presentation of the scene, and during the delay period. Participants were told that the instructional cue

had no bearing on the task they performed when the search display was presented. To confirm that participants followed these instructions, we asked in the PEQ if they directed their eyes to specific objects when the search display was presented (Questions D1 – D4 in Supplementary Table 1). As can be seen in Supplementary Figure 4, participants indicated that on most trials, they did not deliberately alter their viewing based on the instructional cue (retrieve, suppress, substitute).

In the final open-ended question, four participants reported that they directed viewing to the associate in the retrieve condition, and a specific object from the other category on substitute trials. To ensure that the eye movement results were not driven by these participants, we repeated the analyses without those participants. The pattern of results from the eye movement data remained unchanged. Specifically, proportion of total viewing time directed to the associate was reduced in the suppress and substitute conditions compared to the retrieve condition, and a specific object from the other category as the associate was most often prioritized in the substitute condition. This suggests that the process by which retrieved information held active in mind attracts viewing, may not require deliberate control.

Supplementary Figure 4.

Histogram of Responses to Questions D1-4 in the PEQ.



Note. Responses from the questions pertaining to deliberate manipulation of eye movements are plotted here. Responses to other questions on strategies used by participants to control retrieval are reported in Supplementary Table 1.

Supplementary Table 1.

Responses to Post-experimental Questionnaire.

A. Please answer the following questions to indicate the extent to which you utilized each strategy when attempting to ***suppress*** items. After the scene was presented I:	Never	Rarely	Sometimes	Frequently	Very frequently
1. Made sure I still knew the associated face/tool first, and then tried to not think of this associated face/tool	0 (0%)	3 (11.54%)	8 (30.77%)	10 (38.46%)	5 (19.23%)
2. Tried not to think of the associated face/tool, but then after the trial was over, I made sure I still remembered the associated face/tool.	6 (23.08%)	8 (30.77%)	8 (30.77%)	3 (11.54%)	1 (3.85%)
3. Kept myself from thinking about the associated face/tool by keeping my mind completely blank	5 (19.23%)	6 (23.08%)	5 (19.23%)	7 (26.92%)	3 (11.54%)

B. Please answer the following questions to indicate the extent to which you utilized each strategy when attempting to ***substitute*** items. After the scene was presented I:

	Never	Rarely	Sometimes	Frequently	Very Frequently
1. Made sure I still knew the associated face/tool first, and then tried to call to mind a substitute tool/face	0 (0%)	2 (7.69%)	0 (0%)	10 (38.46%)	14 (53.85%)
2. Tried not to think of the associated face/tool, and immediately called to mind a substitute tool/face	9 (34.62%)	7 (26.92%)	6 (23.08%)	3 (11.54%)	1 (3.85%)
3. Always used the same face/tool for substitution (for example -- always used the hammer when trying to substitute for a face; always used Obama when trying to substitute for a tool)	7 (26.92%)	6 (23.08%)	1 (3.85%)	7 (26.92%)	5 (19.23%)
4. Made pairs between objects that I would use for substitution (for example -- always use the hammer when trying to substitute for Obama; always use Rupert Grint when trying to substitute for the scissors)	10 (38.46%)	9 (34.62%)	4 (15.38%)	3 (11.54%)	0 (0%)

C. Please answer the following questions to indicate the extent to which you utilized each strategy when attempting to ***retrieve*** items.

After the scene was presented I:

	Never	Rarely	Sometimes	Frequently	Very Frequently
1. Made sure I still knew the associated face/tool first, and then kept it in mind until the search display was presented	0 (0%)	1 (3.85%)	3 (11.54%)	4 (15.38%)	18 (69.23%)
2. Continued to try to retrieve the associate until the search display was presented if it did not immediately come to mind	1 (3.85%)	2 (7.69%)	6 (23.08%)	10 (38.46%)	7 (26.92%)
3. Called the associated face/tool to mind, but then thought about something else until the search display was presented	7 (26.92%)	11 (42.31%)	5 (19.23%)	2 (7.69%)	1 (3.85%)
4. Deliberately looked for (moved my eyes to) the associated face/tool in the search display on every trial (whether the instructions were to *retrieve*, *suppress*, or *substitute*)	5 (19.23%)	4 (15.38%)	13 (50%)	3 (11.54%)	1 (3.85%)

D. Please answer the following questions to indicate the extent to which you performed the following behaviors ***when the search display was presented***. After the search display was presented I:

	Never	Rarely	Sometimes	Frequently	Very Frequently
1. Deliberately avoided looking for (moving my eyes to) the associated face/tool in the search display regardless of trial type (whether instructions were to *retrieve*, *suppress*, or *substitute*)	4 (15.38%)	9 (34.62%)	8 (30.77%)	5 (19.23%)	0 (0%)
2. Deliberately looked for (moved my eyes to) the face/tool in the search display that I had used for substitution in the *Substitute* trials	5 (19.23%)	6 (23.08%)	4 (15.38%)	7 (26.92%)	4 (15.38%)
3. Simply looked at the objects, without regard for whether they were associates, substitutes, etc. until the search target (the lone circle) was revealed	0 (0%)	6 (23.08%)	9 (34.62%)	7 (26.92%)	4 (15.38%)

Note. For all questions, participants responded on a 5-point scale (0 - Never, 1 – Rarely, 2 – Sometimes, 3 – Frequently, 4 – Very Frequently)