Obsessive Compulsive visual search: a reexamination of presence-absence asymmetries

Noam Sarna1, Matan Mazor2, & Ruvi Dar1

1 School of Psychological Sciences, Tel Aviv University

2 Department of Experimental Psychology, University of Oxford

Author note

Correspondence concerning this article should be addressed to Noam Sarna, Tel Aviv, Israel 69978. E-mail: [noamsarna@mail.tau.ac.il](mailto:noamsarna@mail.tau.ac.il)

Abstract

In previous research, obsessive-compulsive (OCD) tendencies were associated with longer search times in visual search task. These findings, replicated and extended to a clinical sample, were specific to target-absent trials, with no effect on target-present trials. This selectivity was interpreted as checking behavior in response to mild uncertainty. However, an alternative interpretation is that individuals with high OCD tendencies (OC+) have a specific difficulty with inference about absence. In two large-scale pre-registered online experiments (conceptual replication N= 1007; direct replication N= 226), we sought to replicate the original finding and elucidate its underlying cause: an increased sensitivity to mild uncertainty, or a selective deficiency in inference about absence. Both experiments showed no evidence of prolonged search times in target-absent trials for OC+ individuals. Taken together, our results do not support the notion that inducing mild uncertainty in the form of target absence leads to excessive checking among OC+ individuals.

Obsessive Compulsive visual search: a reexamination of presence-absence asymmetries

# Introduction

Theories on obsessive-compulsive disorder (OCD) emphasize the pivotal role of pathological doubt in the disorder’s phenomenology (Dar, 2004; Dar, Lazarov, & Liberman, 2021; Rasmussen & Eisen, 1989; Reed, 1985) This persistent doubt is reflected in lowered confidence in memory, decision-making, perception and other cognitive functions, which give rise to repetitive checking rituals that, paradoxically, only serve to intensify the doubt (Hout & Kindt, 2003). In the lab, doubt and checking behavior are commonly manifested in slow reaction times (e.g., Sarig, Dar, and Liberman (2012); Banca et al. (2015); Hauser et al. (2017).

The present study focused on the finding that participants with high OCD tendencies (OC+) took more time than those with low OCD tendencies (OC-) to identify when a target was absent from a visual search array, whereas no such difference was observed when the target was present (Toffolo, Hout, Hooge, Engelhard, & Cath, 2013). These findings have been replicated (Toffolo, Hout, Engelhard, Hooge, & Cath, 2014) and extended to a clinical sample, where they were found to be specific to patients with OCD and absent in those suffering from anxiety (Toffolo, Hout, Engelhard, Hooge, & Cath, 2016). In these experiments, checking behavior was operationalized as search time, and high and low uncertainty were operationalized by means of contrasting target-present and target-absent trials. Relatively longer search times for the OC+ group in target-absent trials were interpreted as perseverative checking behavior under mild uncertainty. However, while deciding that a target is absent is indeed commonly accompanied by lower levels of subjective confidence compared to deciding that a target is present (Mazor, Friston, & Fleming, 2020; Mazor, Moran, & Fleming, 2021), these type of decisions about absence, are also qualitatively different from decisions about presence, as they cannot be based on direct perceptual evidence. To determine that a target is absent, one must believe that if the target were present, they would have been able to perceive it: a form of inference that requires counterfactual thinking and reliance on self-knowledge (Mazor, 2021). Therefore, an alternative mechanism behind the longer search times in target-absent trials among OC+ participants could be a specific difficulty with inference about absence, rather than simply heightened sensitivity to uncertainty.

Clinical observations provide some support for the idea that people with OCD struggle with inferences about absence. One example is “Hit-and-run OCD”, in which individuals feel compelled to mentally or physically retrace their driving route to ensure that they did not kill or injure someone while driving (Hyman & Pedrick, 2010). This phenomenon manifests key properties of inference about absence: To conclude that an accident has not happened, a person needs to rely on the belief that if it did happen, they would have noticed it.

This clinical examples raise the possibility that the increased search time for target-absent trials may be due to a specific difficulty in inferring absence rather than a general intolerance of uncertainty. To test this idea, in two pre-registered online studies, we conducted a conceptual replication and a direct replication of the visual search study by Toffolo et al. (2013). Participants high and low in OCD tendencies were presented with visual search displays and asked to decide whether a target was absent or present. Experiment 1 aimed to elucidate whether the increased search times in target-absent trials for OC+ individuals are attributable to a specific difficulty with inference about absence or a general difficulty with handling uncertainty. Following our failure to replicate the original findings in this first experiment, Experiment 2 was designed as a more direct replication of Toffolo et al. (2013), using the exact same stimuli and instructions.

## Experiment 1

In Experiment 1, we sought to dissociate specific difficulties with inference about absence from more general difficulties with uncertainty by introducing an easy target-absent condition. To our surprise, we observed no group differences in target-absent search times, even for search displays that elicit high levels of uncertainty. We therefore focus our report here on this replication failure.

## Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All analysis scripts and anonymized data are available at github.com/Noamsarna/ocd\_visual\_search. The order and timing of experimental events were determined pseudo-randomly by the Mersenne Twister pseudorandom number generator, initialized to ensure registration time-locking (Mazor et al., 2019). A detailed preregistration document for Experiment 1 can be accessed at github.com/Noamsarna/ocd\_visual\_search/tree/main/experiments/Experiment1

# Methods

# Participants

The research was approved by the Research Ethics Committee of Tel-Aviv University (study ID number 0004169-1). One thousand and seven participants were recruited via Prolific (<https://prolific.co/>) and selected based on the following criteria: an acceptance rate above 95%, no participation in previous study pilots, use of non-Safari browsers, and being native English speakers located in the UK. The median completion time for the entire experiment was 14 minutes. Participants were paid £2 for their participation, equivalent to an hourly wage of £8.57. Participants were divided into high (OC+) and low (OC-) OCD tendencies groups based on their scores in the OCI-R (Foa et al., 2002; see below), with the OC+ group consisting of those in the highest quartile of the OCI-R scores distribution and the OC- group comprising those in the lowest quartile of this distribution. The entire sample (n=1007) completed the visual search task. Due to higher-than-expected exclusion rate, and in deviation from our pre-registered plan to collect 250 participants in each group, our final sample included 213 OC+ participants and 220 OC- participants.

The average age of the total sample was 30.41 years (SD = 5.70). Half of the sample identified as female. In terms of ethnicity, the majority (84.00) identified as White, followed by Asian (7%), Black (4%), and Mixed/Other (5%). The predominant nationality was UK (93%). Employment status was predominantly full-time (62%), followed by part-time (16%).

# Visual search task

The visual search task consisted of four blocks, each containing 24 trials of searching for either a closed or an open square. The task began with a practice phase consisting of one block with six trials. Each display was presented for a maximum of 10 seconds or until a response was received. During the practice phase, feedback about accuracy was given after each trial: If the response was correct, the word “Correct!” appeared on the screen for one second; If the response was wrong, the word “Wrong” appeared on the screen for 5 seconds. In the main part of the experiment, no feedback was given, as was the case in the original paradigm (Toffolo et al., 2013). After completing the practice, participants looked for either a closed square among rotated open squares (‘hard search’; Fig. 1, main part, right panel), or for a rotated open square among closed squares (‘easy search’; Fig. 1, main part, left panel). The difference in difficulty between these two search types is due to a search asymmetry for open/closed edges (Treisman & Gormican, 1988). We further manipulated target presence and set size, resulting in a 2X2X2 design (Search type: ‘easy-search’ or ‘hard-search’; Target: present/absent; Set size: 9 or 25). Block order was counterbalanced between participants, and trial order within individual blocks was fully randomized (figure 1).

# Measures

# Obsessive–Compulsive Inventory-Revised (OCI-R; Foa et al., 2002).

The OCI-R is an 18-item self-report measure of OCD symptom severity. Responders are asked to rate their level of distress pertaining to 18 statements in the past month on a five-point scale ranging from 0 (Not at all) to 4 (Extremely). The OCI-R has been shown to have good validity, test-retest reliability and internal consistency in both clinical (Foa et al., 2002) and non-clinical samples (Hajcak, Huppert, Simons, & Foa, 2004).

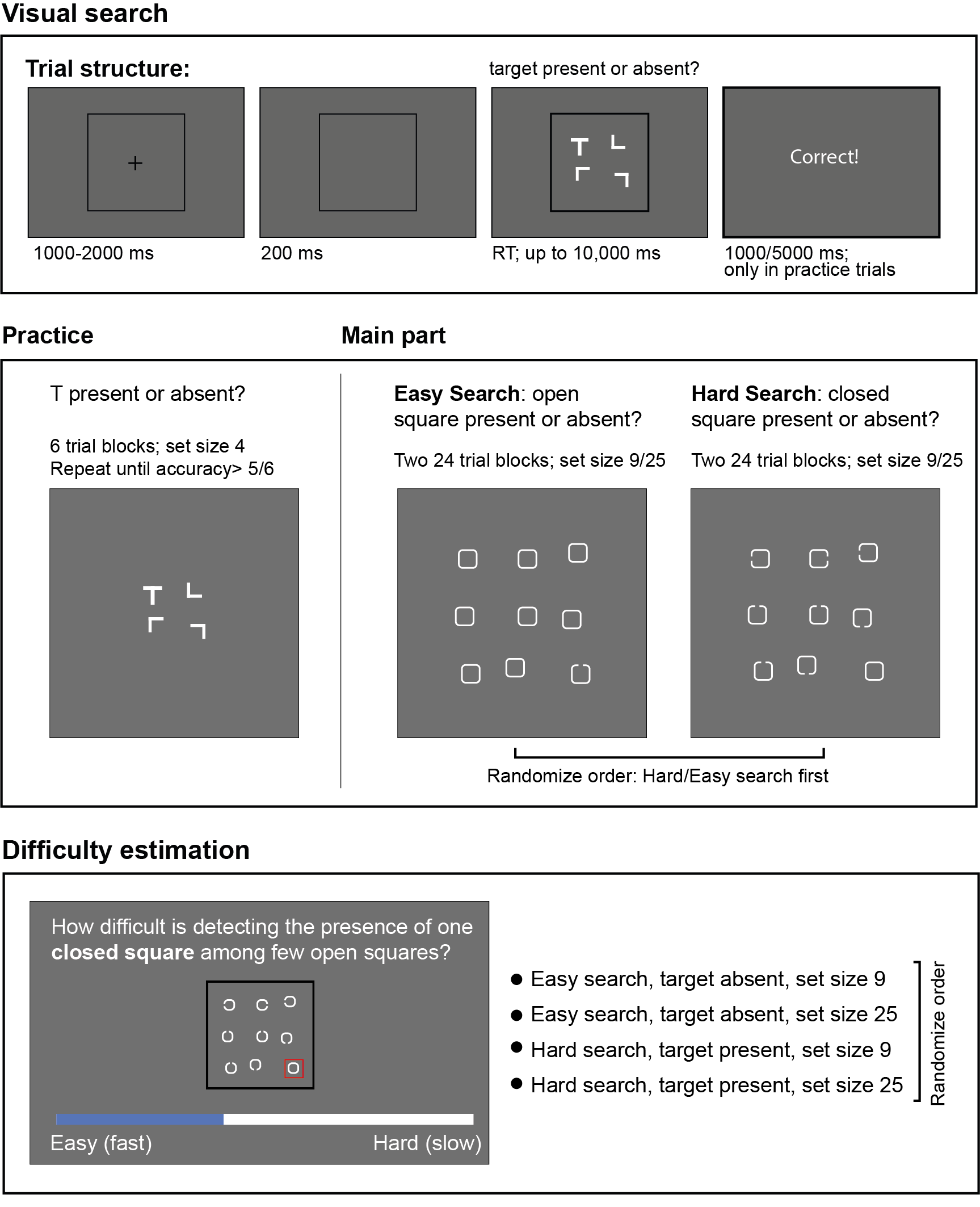
# Depression, Anxiety and Stress Scales-21 (DASS-21; Lovibond & Lovibond, 1995).

The DASS-21 is a 21- item self-report questionnaire that is divided into three seven-item subscales to measure dimensional components of depression, anxiety, and stress. ). Each individual item refers to the respondent’s experiences over the past week and is evaluated on a four-point scale, ranging from 0 (‘the item does not apply to me at all’) to 3 (‘the item applies to me very much or most of the time’). The DASS-21 has shown high reliability, validity, and internal consistency within both clinical groups and community sample (Antony & Bieling, 1998; Henry & Crawford, 2005; Lovibond & Lovibond, 1995). In this study only the depression and anxiety scales were used. We used the depression and anxiety sub-scales to control for non-specific effects associated with OCD tendencies.

## Procedure

A static version of Experiment 1 can be accessed at: <https://noamsarna.github.io/ocd_visual_search/experiments/demos/exp1/>. Participants were first instructed about the experiment’s structure, which comprised three parts: a visual search task, questions about the visual search, and the two inventories: OCI-R and DASS-21. Then, they received written instructions about the visual search task. After completing the visual search task, participants were asked to rate the difficulty of noticing the presence or absence of a certain target among different distractors (more information about this in the appendix). Following the difficulty estimation, participants completed the OCI-R and DASS-21. We included two attention check questions among the OCI-R items, asking participants to select a certain answer (‘If you read this question, check the option ’Not at all’).

**Figure 1** *Overview of experimental Design*



*note* Top panel: each visual search trial started with a centered black fixation cross. Middle panel (Practice): After reading the instructions, participants completed practice trials, searching for a rotated T among rotated L’s in 6-trial blocks until they achieved a minimum accuracy of 0.83 (no more than one error). Middle panel (Main part): The primary experiment comprised 96 trials in four blocks, with the target identity changing after two blocks. Each 24-trial block followed a 2x2 design, manipulating set size (9 or 25) and target presence (present/absent). Bottom panel: Search difficulty estimation: participants used their mouse to rate search difficulty on a continuous scale. In questions about target-present searches, the target was marked with a red square.

## Data Analysis

# Exclusion Criteria

Participants were excluded from the analysis if they made more than 15% errors in the main part of the experiment or for having extremely fast or slow reaction times (below 100 milliseconds) in more than 25% of the trials. Participants were also excluded if they failed one or more of the attention checks. In total, 109 out of 1007 participants were excluded from the analysis. For the remaining participants, error trials and trials with response times below 100 milliseconds were excluded from the response-time analysis.

# Results

We focus our report here on our failure to replicate a group difference in target-absent search times, even for search displays that elicit high levels of uncertainty. All additional analysis from our preregistered hypotheses can be found in the appendix of this paper.

# Replication of group differences in target-absent RT

To directly replicate group differences in target-absent response time (RT) (Toffolo et al., 2013, 2014, 2016), we focused on the difficult search with the larger set size (set size = 25). We conducted a mixed-effects ANOVA, with mean response time (RT) as the dependent variable, group (OC+ vs. OC-) as a between-subjects variable, and target presence (present vs. absent) as a within-subjects variable. Specifically, we examined the interaction effect testing the hypothesis that the mean RT difference between the OC+ and OC- groups would be significantly more pronounced in target-absent trials. Contrary to our expectations, the analysis did not reveal a significant interaction between group and target presence, , , Cohen’s *d*= 0.12 (figure 2, Exp1). A null result was also obtained in a correlation analysis, pooling data from all participants and treating OCI scores as a continuous variable (see Supplementary materials, pre-registered hypothesis 9). To quantify the evidence for the null we conducted a Bayesian t-test setting the scale at the averaged effect size found in Toffolo et al., (2013, 2014) reflecting a belief that if present, group differences should be negative and moderate in size (Rouder, Speckman, Sun, Morey, & Iverson, 2009). A one-sided Bayesian independent samples t-test produced a Bayes Factor of , providing strong evidence for the null hypothesis of no group differences.

We conducted several additional analyses that examine the interaction between the OC groups and the presence of the target. First, at the group level, we performed multi-level regression, accounting for anxiety and depression. We found no interaction between the OC groups and the presence of the target (preregistered hypothesis 10; , 95% CI , , ). Similarly, when we focused on the initial trials of the task, prior to any accumulated experience (preregistered hypothesis 8), we found no interaction between group and target presence in a mixed-effects ANOVA (, ). Furthermore, at the group level, we observed no significant differences between the groups in their self-reported measures of task difficulty. A group difference in accuracy did reach significance, such that the OC+ group (*M*=0.94) was overall less accurate than the OC- group (*M*=0.95; , ). Importantly, this difference did not replicate in Experiment 2. To extend our analysis to the entire sample, encompassing the four OCI-R quartiles, we replaced the group variable (OC+; OC-) with the full range of OCI-R scores. In this analysis, we still found no interaction between OCI-R scores and the presence of the target (preregistered hypothesis 9; , 95% CI , , . Detailed calculations and results for all these hypotheses are provided in the Appendix for further reference.

## Experiment 2

In Experiment 1, target-absent search times were not significantly slower in OC+ compared to OC- individuals. While this stands in contrast to previous reports (Toffolo et al., 2013, 2014, 2016), our results differed from those of the original study in other respects as well. Most notably, search times in this study (~4.5s for target-absent and ~2.6s for target-present) were overall shorter compared to those in Toffolo et al. (2013) (~5.5 for target-absent and ~3.5s for target-present). We therefore considered the possibility that the task used in Experiment 1 may have been less challenging and potentially insufficient to elicit doubt and trigger checking behavior. To test this, Experiment 2 employed the original stimuli from Toffolo et al. (2013). The preregistered analysis plan can be accessed at the following link: <https://github.com/Noamsarna/ocd_visual_search/tree/main/experiments/Experiment2>. In Experiment 2, we conducted a further power analysis mirroring the methods of Toffolo et al. (2013), utilizing their data and adopting a bootstrap approach to determine an adequately powered sample size, as detailed in the preregistration document for Experiment 2. We employed the Mersenne Twister pseudorandom number generator to ensure that our preregistration preceded data collection (Mazor et al., 2019).

## Method

# Participants

Two hundred twenty-six participants were recruited via Prolific. To maximize statistical power for a group comparison, we invited former participants whose OCI-R scores were in the top or bottom quartile in Exp. 1. In line with our preregistered stopping rule, we kept data collection until we had invited all participants in the first and fourth quartiles from our previous experiment (n=220; n=213, respectively). Participants completed the OCI-R questionnaire again in the present study (the test-retest reliability for the OCI-R yielded a Pearson’s correlation coefficient of , 95% CI , , ), and were assigned to the OC+/OC- groups based on the original cut-off scores from Toffolo et al., 2013 (OCI-R total score ≥ 17 for the OC+ group; OCI-R total score ≤ 5 for the OC- group). Our final sample consisted of 110 OC+ participants and 68 OC- participants. The entire experiment took 12 minutes to complete, and participants were paid £1.8 for their participation, equivalent to an hourly wage of £9.

## Procedure

A static version of Experiment 2 can be accessed here: <https://noamsarna.github.io/ocd_visual_search/experiments/demos/exp2/> Experiment 2 was similar to Experiment 1 with the following exceptions. First, we used the original stimuli from Toffolo et al. (2013). The visual search task consisted of one block of 50 individual search displays, each containing 25 elements. The search task was more challenging due to a larger search grid, which meant larger distances between stimuli, as well as reduced stimulus size. Second, Experiment 2 did not include an assessment of perceived difficulty, comprising only the visual search followed by the same questionnaires as in Experiment 1. Third, to make it identical to Toffolo et al. (2013), practice trials in Experiment 2 (four per block) involved the same stimuli as the main blocks. Fourth, participants were instructed to press the spacebar to move from the fixation cross screen to the search display screen, at which point the search display appeared immediately. Finally, the visual search part of the experiment included only the hard search type: detecting a closed square among open squares.

## Data Analysis

*Exclusion Criteria*

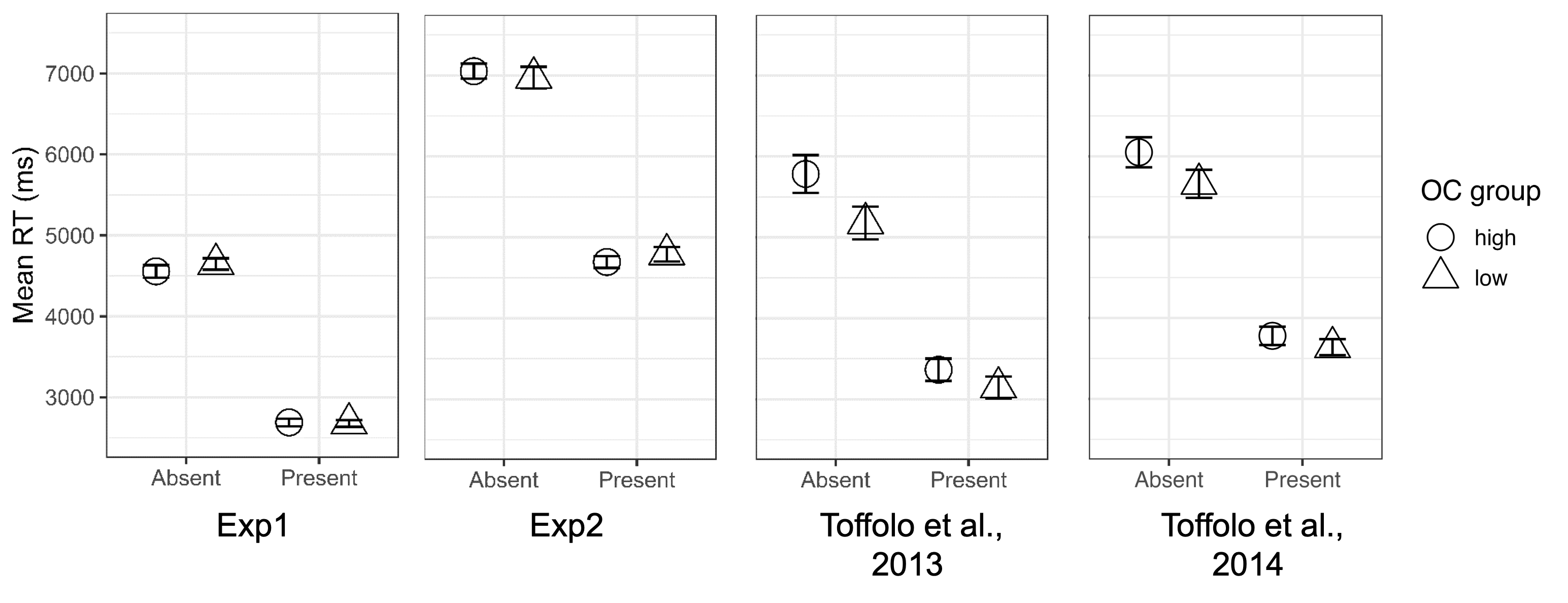
Since Experiment 2 served as a direct replication, we adopted the same rejection criteria as Toffolo et al. (2013), so that participants were excluded if their error count exceeded 2.5 standard deviations from the mean error rate of the entire sample. As in Experiment 1, participants were also excluded from the analysis if they failed to answer correctly one or more attention-check questions.

# Results

*Direct replication*

In contrast to Toffolo et al. (2013), where presence-absence differences in reaction time were more pronounced among OC+ participants, in our replication sample the one-tailed t-test of the interaction contrast (using the difference in search times as a DV), was not significant, , , Cohen’s *d*=0.22 providing no evidence for the expected interaction. Notably, the numeric trend of the interaction in our sample was driven by shorter RT in the OC+ as compared to the OC- in target-present trials, rather than by longer RT for target-absent responses (Figure 2, Exp2). This pattern is different from that reported by Toffolo et al. (2013), where OC+ participants were slower in both search types, but particularly in target-absent searches (Figure 2, Toffolo et al., 2013). Unlike in Experiment 1, we observed no differences in accuracy between the groups (OC+: *M*=0.83, OC-: *M*=0.83; , ). Finally, a one-sided Bayesian independent samples t-test produced a Bayes Factor of , providing moderate evidence for the null hypothesis of no group differences.

**Figure 2** *Results from Experiment 1, Experiment 2 and Toffolo et al., 2013, 2014*



*note* Mean reaction times for target-absent and target-present trials (X-axis). Error bars represent the standard error of the mean. Shapes represent the OC groups: Circle for OC+; Triangle for OC-.

# Discussion

In two preregistered, large-sample studies, we found no evidence of prolonged search time among OC+ participants in target-absent trials, contrary to previous findings by Toffolo and colleagues (2013, 2014, 2016).

The most notable difference between our experiments and those conducted by Toffolo and colleagues (2013, 2014, 2016) lies in our use of an online setting versus their use of in-person lab experiments. Completing tasks online as opposed to a laboratory setting generates more “technical noise”, that is, unexplained variance driven by technical variation. However, previous studies suggest that such noise has minimal impact on RT differences in perceptual tasks. In a study comparing RT distributions from a lab-based Matlab and an online JavaScript experiment, the results revealed near-identical RTs between the two setups (Leeuw & Motz, 2016). The JavaScript experiment showed a consistent delay of around 25 milliseconds, which had minimal impact on the sensitivity to RT changes due to experimental manipulations.

Furthermore, in our study, participants completed the visual search task using a range of computers and displays, rather than in a controlled lab environment with a fixed screen as in Toffolo et al. (2013, 2014, 2016). Yet, simulation studies have demonstrated minimal impact of technical variance on statistical power and the precision of effect size estimates (Brand & Bradley, 2012). Key behavioral findings in psychology, including those observed in the Stroop and flanker tasks, as well as effects reliant on much smaller time constants, like attentional blink and subliminal priming, have been successfully replicated in web-based studies (Crump, McDonnell, & Gureckis, 2013). Specifically, a recent online visual search study reported significant RT variations between experimental conditions, with a focus on smaller time constants than those anticipated in our paradigm (Mazor & Fleming, 2022). Particularly strong evidence for the comparability of lab-based vs. web-based findings comes from a study which utilized a fully randomized design for reaction time effects (Hilbig, 2016). The results showed that a word frequency effect (manifested in different RT) was comparable in magnitude across all three conditions. Taken together, these studies show that whereas some variations between settings in RT exist, they are minor, especially when the outcome measure is RT alterations due to experimental manipulations.

Additional differences between our research and Toffolo’s studies, which could interact with OCD tendencies, are anonymity and demographic variations. It is plausible that the anonymity afforded by online studies could lead to participants feeling less responsible for study outcomes than identifiable psychology students who meet experimenters in person. Moreover, participants in Toffolo’s studies were monitored by an eye-tracker camera, a factor that has been suggested to reduce reliance on internal cues such as metacognitive experiences (Noah, Schul, & Mayo, 2018). Given the sensitivity of OC+ individuals to personal responsibility (Salkovskis, 1985), and the heightened sense of anonymity in online studies, the transition to an online setting may have attenuated group differences in checking behavior.

Our failure to find an association between obsessive-compulsive tendencies and inference about absence may appear inconsistent with well-known clinical manifestations of OCD, such as those observed in “Hit-and-Run OCD”. However, our experimental operationalization of inference about absence differed from these clinical manifestations in two important ways. First, we did not manipulate perceived responsibility, a key feature of “Hit-and-Run OCD” and one that is posited to play a key role in OCD more generally (Salkovskis, 1985). Second, in the clinical example of “Hit-and-Run OCD”, the compulsion is associated more with a recollection of an event rather than its direct experience. Indeed, most findings related to reduced confidence in individuals with OCD have been observed in relation to memory rather than perception (for review see Dar, Sarna, Yardeni, and Lazarov (2022). More research is needed to elucidate the interaction of these two features with inference about absence in OCD.

Finally, this replication attempt puts into action several key features of replicable science (Tackett et al., 2017). Our study included detailed preregistration with hypotheses, power analysis, analysis plan and exclusion criteria. We used the preregistration time-locking tool (Mazor, Mazor, & Mukamel, 2019), thereby guaranteeing that our registration preceded the data collection process. Furthermore, our study represents the first independent replication attempt. Lastly, we have made our raw data, analysis scripts, and task codes publicly available. Beyond a contribution to the experimental literature on OCD, we hope this report may serve as a helpful reference for reproducible and open clinical psychological science.

## Conclusion

The presented findings diverge from those of previous studies by Toffolo and her colleagues (2013, 2014, 2016), as we were unable to replicate the effect of prolonged search time for OC+ participants in target-absent trials. To the very least, this replication failure indicates that the original effect may be constrained to a specific setting, thus limiting its generalizability to other contexts. More broadly, our results advocate for the application of open science practices in clinical psychology research, to foster methodological integrity and ensure the reliability of findings.

# References

Antony, M. M., & Bieling, P. J. (1998). *Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample*.

Banca, P., Vestergaard, M. D., Rankov, V., Baek, K., Mitchell, S., Lapa, T., … Voon, V. (2015). Evidence accumulation in obsessive-compulsive disorder: The role of uncertainty and monetary reward on perceptual decision-making thresholds. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *40*(5), 1192–1202. <https://doi.org/10.1038/npp.2014.303>

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*(3), 350–357. <https://doi.org/10.1177/0894439311415604>

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon’s mechanical turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>

Dar, R. (2004). Elucidating the mechanism of uncertainty and doubt in obsessive-compulsive checkers. *Journal of Behavior Therapy and Experimental Psychiatry*, *35*(2), 153–163. <https://doi.org/10.1016/j.jbtep.2004.04.006>

Dar, R., Lazarov, A., & Liberman, N. (2021). Seeking proxies for internal states (SPIS): Towards a novel model of obsessive-compulsive disorder. *Behaviour Research and Therapy*, *147*, 103987.

Dar, R., Sarna, N., Yardeni, G., & Lazarov, A. (2022). Are people with obsessive-compulsive disorder under-confident in their memory and perception? A review and meta-analysis. *Psychological Medicine*, *52*(13), 2404–2412. <https://doi.org/10.1017/S0033291722001908>

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485.

Hajcak, G., Huppert, J. D., Simons, R. F., & Foa, E. B. (2004). Psychometric properties of the OCI-r in a college sample. *Behaviour Research and Therapy*, *42*(1), 115–123. <https://doi.org/10.1016/j.brat.2003.08.002>

Hauser, T. U., Allen, M., NSPN Consortium, Bullmore, E. T., Goodyer, I., Fonagy, P., … Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*(1), 6614. <https://doi.org/10.1038/s41598-017-06116-z>

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239. <https://doi.org/10.1348/014466505X29657>

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. <https://doi.org/10.3758/s13428-015-0678-9>

Hout, M. van den, & Kindt, M. (2003). Repeated checking causes memory distrust. *Behaviour Research and Therapy*, *41*(3), 301–316.

Hyman, B. M., & Pedrick, C. (2010). *The OCD workbook: Your guide to breaking free from obsessive-compulsive disorder*. New Harbinger Publications.

Leeuw, J. R. de, & Motz, B. A. (2016). Psychophysics in a web browser? Comparing response times collected with JavaScript and psychophysics toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. <https://doi.org/10.3758/s13428-015-0567-2>

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.

Mazor, M. (2021). *Inference about absence as a window into the mental self-model*.

Mazor, M., & Fleming, S. M. (2022). Efficient search termination without task experience. *Journal of Experimental Psychology: General*.

Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife*, *9*, e53900.

Mazor, M., Mazor, N., & Mukamel, R. (2019). A novel tool for time-locking study plans to results. *European Journal of Neuroscience*, *49*(9), 1149–1156.

Mazor, M., Moran, R., & Fleming, S. M. (2021). Stage 2 registered report: Metacognitive asymmetries in visual perception. *Neuroscience of Consciousness*, *2021*(1), niab025. <https://doi.org/10.1093/nc/niab025>

Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology*, *114*(5), 657–664. <https://doi.org/10.1037/pspa0000121>

Rasmussen, S. A., & Eisen, J. L. (1989). *Clinical features and phenomenology of obsessive compulsive disorder*. SLACK Incorporated Thorofare, NJ.

Reed, G. F. (1985). *Obsessional experience and compulsive behavior: A cognitive-structural approach*. Academic Press. Retrieved from <https://books.google.co.il/books?id=AGHXzgEACAAJ>

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>

Salkovskis, P. M. (1985). Obsessional-compulsive problems: A cognitive-behavioural analysis. *Behaviour Research and Therapy*, *23*(5), 571–583.

Sarig, S., Dar, R., & Liberman, N. (2012). Obsessive-compulsive tendencies are related to indecisiveness and reliance on feedback in a neutral color judgment task. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 692–697.

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., … Shrout, P. E. (2017). It’s time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, *12*(5), 742–756. <https://doi.org/10.1177/1745691617690042>

Toffolo, M., Hout, M. A. van den, Engelhard, I. M., Hooge, I. T. C., & Cath, D. C. (2014). Uncertainty, checking, and intolerance of uncertainty in subclinical obsessive compulsive disorder: An extended replication. *Journal of Obsessive-Compulsive and Related Disorders*, *3*(4), 338–344. <https://doi.org/10.1016/j.jocrd.2014.08.004>

Toffolo, M., Hout, M. A. van den, Engelhard, I. M., Hooge, I. T., & Cath, D. C. (2016). Patients with obsessive-compulsive disorder check excessively in response to mild uncertainty. *Behavior Therapy*, *47*(4), 550–559.

Toffolo, M., Hout, M. A. van den, Hooge, I. T. C., Engelhard, I. M., & Cath, D. C. (2013). Mild uncertainty promotes checking behavior in subclinical obsessive-compulsive disorder. *Clinical Psychological Science*, *1*(2), 103–109. <https://doi.org/10.1177/2167702612472487>

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15.

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-outcome knowledge dissociates from behavior in obsessive-compulsive disorder following contingency degradation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 200–209.

# (APPENDIX) Appendix

# Supplementary methods, Experiment 1

## Search termination without task experience

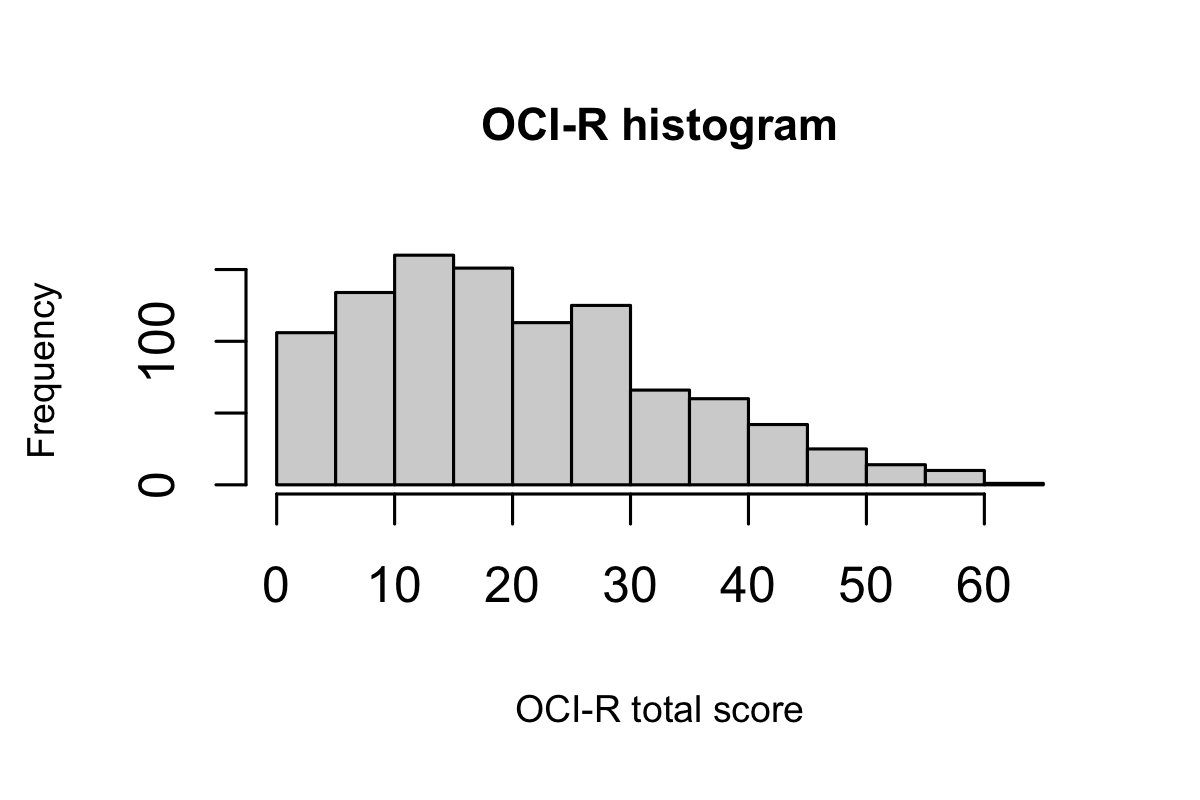
To assess participants’ search termination without task experience, i.e., decisions about target absence made before ever experiencing target-present trials (Mazor & Fleming, 2022), the first two trials of the first and third blocks were always target-absent trials (one of each set size), presented in randomized order. This was unknown to the participants.

## Estimation of search difficulty

After the visual search task, participants were asked to rate the difficulty of detecting a specific target, or inferring its absence, among various distractors. Each question referred to a picture of a trial with a red square around the target if it was present, or no marking at all if the target was absent (so that participants were not actively searching for the target in the difficulty estimation part). Participants then rated the perceived difficulty on a continues scale, dragging the curser from left to right (see figure 1; difficulty estimation). They rated four different trial types: detecting the presence of one closed square among many open squares, detecting the presence of one closed square among few open, detecting the absence of one open square among many closed squares and detecting the absence of one open square few closed squares. We focused on these four trials as this was our most intriguing hypothesis, comparing easy target-absent trials to more challenging target-present trials.

## OCI-R Psychometric properties

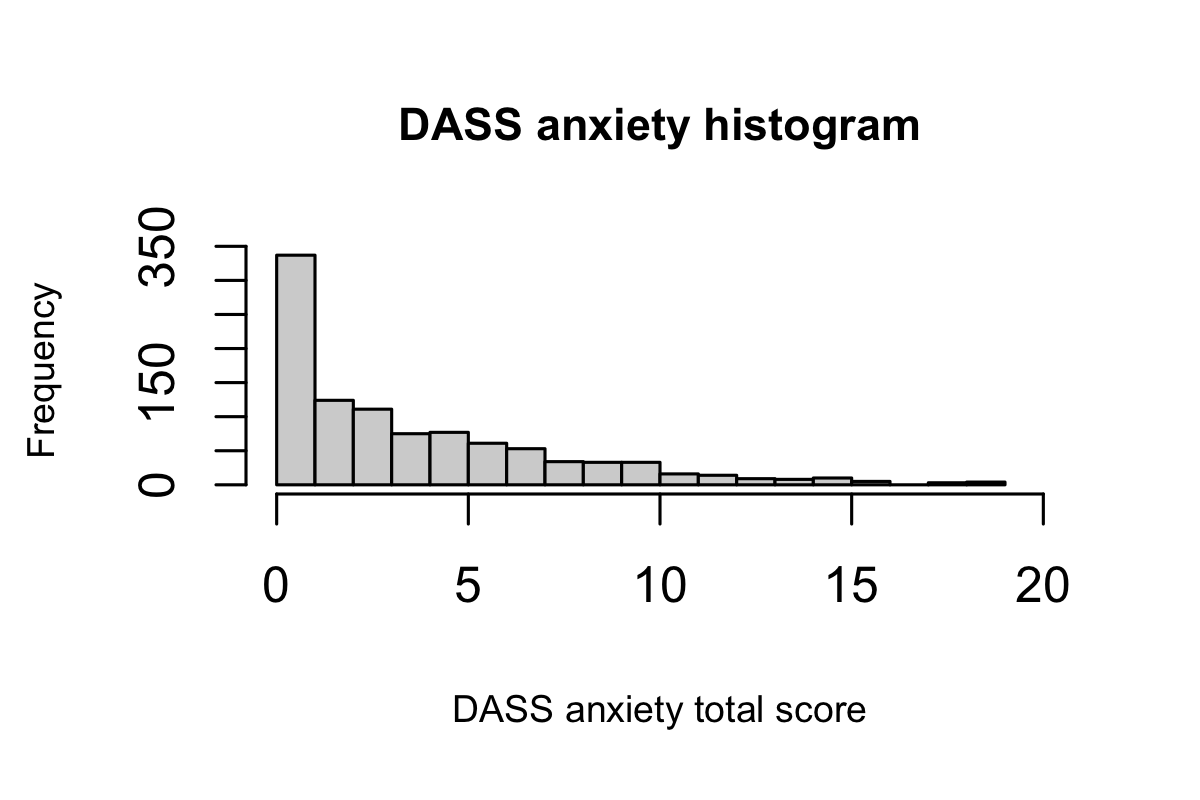
**Figure A1** *Histogram for the OCI-R total score*



The OCI-R demonstrated excellent internal consistency in our sample, with a Cronbach’s alpha of 0.91. The distribution of scores showed a slight positive skewness 0.65, indicating a modest asymmetry with a tail extending towards higher scores. Additionally, the kurtosis of 2.90 suggests a mesokurtic distribution, closely resembling the normal distribution in terms of its peak and tail characteristics. OCI-R cutoff scores for the low OC group (OC-) were defined as OCI-R < 11, and OCI-R cutoff scores for the high OC group were determined as OCI-R > 28.

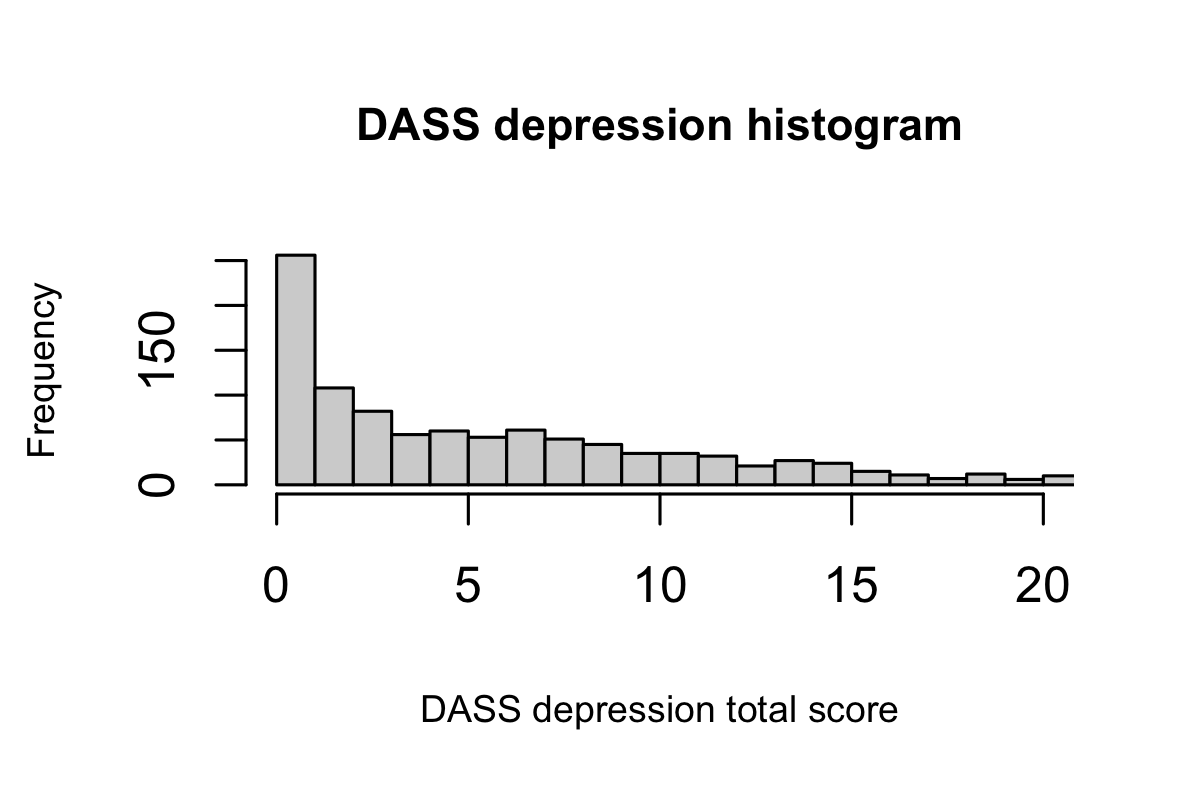
## DASS Psychometric properties

**Figure A2** *Histogram for the DASS anxiety sub-scale*



The DASS anxiety sub-scale demonstrated excellent internal consistency in our sample, with a Cronbach’s alpha of 0.85.

**Figure A3** *Histogram for the DASS depression sub-scale*



The DASS depression sub-scale demonstrated excellent internal consistency in our sample, with a Cronbach’s alpha of 0.93.

# Supplementary results, Experiment 1

In this section we present the results for all of our preregistered hypotheses. A detailed preregistration document including all our original hypotheses for Experiment 1 at github.com/Noamsarna/ocd\_visual\_search/tree/main/experiments/Experiment1. For all of our hypotheses regarding search time provided in this section, we used search slopes as our dependent measure. Subject-wise search slopes were extracted for each combination of search type (easy/hard search) and presence of the target (present/absent), resulting in four search slopes estimates for each participant – easy-present; hard-present; easy-absent; hard-absent. These slopes were computed by fitting a linear regression model to predict reaction time as a function of set size, with one intercept and one set-size term.

## Hypotheses 1 and 2 - Task validation

To validate our paradigm structure and to assess whether we successfully created an easier search by leveraging search asymmetries (i.e., switching between the target and distractors), we first examined the difference in slopes between the two search types (easy/hard), regardless of target presence (pre-registered hypothesis H1). As anticipated, a one-tailed paired t-test demonstrated a steeper slope for the difficult search 100 ms/item compared to the easy search 38 ms/item, , . Furthermore, a one-tailed paired t-test revealed that target-present slopes in the hard search 63 ms/item were steeper than target-absent slopes in the easy search 59 ms/item, , (pre-registered hypothesis H2). The initial two control comparisons served to validate that we successfully designed a target-absent condition that was easier than a target-present condition, thereby experimentally decoupling decision certainty from target presence, and enabling to measure their independent effects on search time as a function of obsessive-compulsive tendencies.

## Hypothesis 3 – Direct replication

See main text.

## Hypothesis 4: Extension of Toffolo et al. (2013), search slopes

Following the previous analysis which addressed the replication of Toffolo and colleagues, we conducted the same mixed-effects ANOVA, with group (OC+/OC-) as between-subject and target presence (present/absent) as within-subject variables, but with search slopes (reaction time as a function of set size) as a dependent variable. Consistent with our findings for Hypothesis 3, a mixed-effects ANOVA revealed no statistically significant interaction between group and target presence , , again providing no evidence for selective slowing of OC+ as compared to OC- individuals in target-absent trials.

## Hypothesis 5: Low-uncertainty inference about absence

Originally, our primary focus was on this hypothesis, examining the effect of obsessive-compulsive tendencies on target-absent search times in an easy, low-uncertainty, search setting (set size =9). However, given our failure to replicate a group difference in the hard search (set size =25; Hypotheses 3 and 4), a significant result here seemed unlikely. Indeed, a mixed-effects ANOVA with group as a between-subjects variable and target presence as a within-subjects variable did not reveal a significant interaction between group and target presence in the easy search, , .

## Hypothesis 6: Model comparison

Our experimental design aimed to differentiate between difficulties arising from higher uncertainty and those arising from the absence of the target. To achieve this, we used two search types: a hard search (searching for a closed square among open squares), and an easy search (searching for an open square among closed squares). Additionally, we manipulated the presence or absence of the target. To determine whether uncertainty or absence had a greater impact on OC+ search time, we constructed two competing regression models that differed only in their interaction terms (table A1). Beyond the search type and the target presence predictors, the first model (M1) included the interaction between group and search type, while the second model (M2) included the interaction between group and target presence (table A1). Additionally, both models included random slopes for search type and target presence for each participant.Since the two models differ only in their last interaction effect, their complexity (that is, the number of fitted coefficients) is the same, which allowed us to compare these models directly. We report that in deviation from our pre-registered plan, we did not compare model accuracy using leave-one-out cross-validation (LOOCV), but instead compared model performance using the Akaike Information Criterion (AIC), with lower AIC values indicating a better fit of the model to the data (Burnham & Anderson, 2004). We chose the AIC over LOOCV because, at the time of performing the analysis, the R implementation of LOOCV did not support the incorporation of random factors in its models. The model comparison table shows that both models have the similar AIC values (difference in AIC scores < 3) suggesting that the interaction between group and target presence does not explain search time to a greater extent than the interaction between group and search type.

(#tab:table\_1)

*Table A1 – Model comparison table*

| Model | AIC | scaled\_AIC | formula |
| --- | --- | --- | --- |
| m1 | 26,248.49 | 0.00 | estimate 1 + OCI\_group + search\_type \* target\_present + search\_type \* OCI\_group |
| m2 | 26,248.36 | -0.13 | estimate 1 + OCI\_group + search\_type \* target\_present + target\_present \* OCI\_group |

*Note.* The two models (M1, M2) are concerning hypothesis 6. AIC value of M1 was scaled to zero, to make the differences clearer. Differences in AIC greater than 10 are considered as decisive evidence supporting the model with the lower numerical value.

## Hypotheses 7 and 8 - First trials

In this analysis, we focused on the two first trials of each search type. These trials are special in that by the time of making these responses, participants haven’t yet experienced target presence. First, we compared the slopes of the two types of searches (easy/hard search) to see if they were different already in the first trials, using a two-tailed paired t-test (H7). This allowed us to isolate pre-specified search termination strategies from ones that were acquired throughout the course of the experiment. In line with our preregistration, prior to the analysis, we corrected for trial and block order effects by subtracting the mean response time of each trial position from the individual’s RT, and then adding the subject’s mean response time. We found a significant difference between the slope of the easy search (62 ms/item) and the slope of the hard search (132 ms/item) in the first trials of each block type , . This significant difference between the search slopes indicates that subjects’ search termination rule was sensitive to the difference between the two search types prior to having experience with the easy and hard search tasks.

We next asked whether this effect of search type on target-absent search times was similar for OC+ and OC- individuals (H8). Using a mixed-effects ANOVA with slope as a dependent variable and group and search type as the two between- and within-subject independent variables, we tested for the interaction of group (OC+/OC-) with search type in the first trials. We did not find a significant interaction between search type and group on first-trial slopes , .

## Hypothesis 9: Correlations between OC tendencies and search slopes

In order to strengthen the validity of our results, we reanalyzed our main hypothesis (H6, M2) incorporating the full spectrum of OC scores (raw OCI total scores). We employed a mixed-effects regression model, which was analogous to M2 from Hypothesis 6, but with the key difference of employing the complete range of total OCI scores, as opposed to focusing solely on the top and bottom quartiles. As this revised model included all participants, not just those with high and low OC scores, it was fitted on a different dataset than M2 in Hypothesis 6. Due to this fundamental difference in the underlying data, a direct comparison of these models would not be appropriate. Consequently, our focus was on the interaction term involving the full range of OCI scores. The interaction between search type and total OCI score was found to be non-significant (, 95% CI , , ).

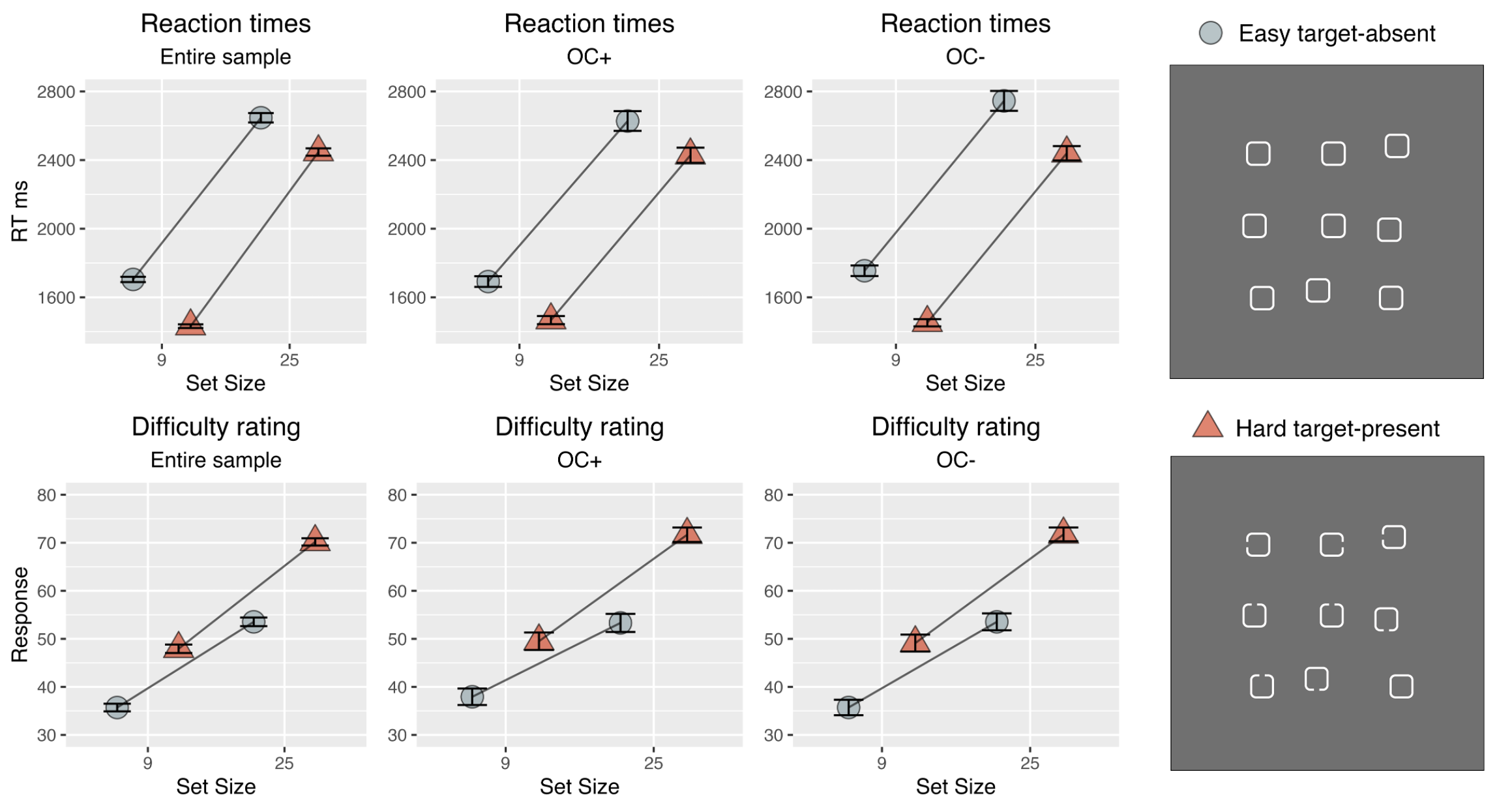
## Hypothesis 10: Controlling for depression and anxiety

Finally, in order to test the possible effects of depression and anxiety, we added DASS subscales of depression and anxiety scores as predictors into our model. We used a mixed-effects regression model identical to the second model in hypothesis 6, except for the addition of depression and anxiety as predictors. The interaction between group and target presence remained below the significance threshold after adding depression and anxiety to the model (, 95% CI , , ).

## Exploratory analysis - Measuring the perceived difficulty of participants

In order to compare explicit metacognitive knowledge with behavior, we collected perceived difficulty ratings for the different search tasks. A discrepancy between perceived difficulty and task performance in the OC+ group would suggest a dissociation between action and knowledge, in line with previous findings (Vaghi et al., 2019). We examined two search types: target-absent easy searches and target-present hard searches, with set sizes of 9 and 25. For both groups, discrepancies between perceived difficulty and mean response time were observed in the same direction: despite overall slower search times in easy target-absent searches, these searches were perceived as easier than hard target-present trials (figure A4).

**Figure 4A** *Relationship between perceived difficulty and reaction times (RT) for high and low OC participants*



*note* Rows display mean RT (top-row) and mean estimated difficulty (bottom-row). Columns represent results for the entire sample (left column) the OC+ group (Middle column) and OC- (Right column). Reaction times and difficulty ratings for easy target-absent and hard target-present trials appear in grey and red, respectively. Error bars represent the standard error of the mean. Right side of the figure shows illustration for easy target absent and hard target present searches with the small set size (9).

# Supplementary results, Experiment 2

## Hypothesis 1 - Task validation

To validate our paradigm and demonstrate that target-absent searches are more difficult, we tested for an effect of condition (target-absent vs. target-present) on mean search times, collapsed across both groups. Our results revealed the expected difference, whereby the mean response time for target-absent responses (M=7,023 ms, SD=1,019 ms) was significantly longer than the mean response time for target present responses (M= 4,733 ms, SD= 751 ms, , )

## Reanalysis of Experiment 2, preregistered hypothesis 2

Here we reanalyze Experiment 2, preregistered hypothesis 2, this time including only participants who had error rate smaller than the actual error rate used in Toffolo et al. (2013) (number of error <19). The one-tailed t-test revealed no significant differences between the groups, , , providing no evidence for the expected interaction. Here too, a numeric interaction in the expected direction was driven by a descriptive difference in the timing of target-present responses, rather than by target-absent responses.