

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

Abstract

In previous research, obsessive-compulsive (OC) tendencies were associated with longer search times in a visual search task. These findings, which were replicated and extended to a clinical sample, were specific to target-absent trials, with no effect on search times when a target was present in the display. Initially, this selectivity was interpreted as indicative of checking behavior in response to mild uncertainty. However, an alternative interpretation is that individuals with high OC tendencies (OC+) suffer from a more specific difficulty with inferences about absence. In two large-scale pre-registered online experiments (conceptual replication $N = 1004$, direct replication $N = 226$), we sought to replicate the original finding and shed further light on its underlying cause: an increased sensitivity to mild uncertainty, or a selective deficiency in inference about absence. In both experiments, we find no evidence of prolonged search times in target-absent trials for OC+ individuals. Taken together, our findings provide no support for the previously observed higher search times of OC+ participants in target-absent trials. We discuss potential differences relative to previous findings and implications for cognitive and metacognitive theories of OCD.

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; Shapiro, 1965). 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. For instance, in a study by Sarig, Dar, and Liberman (2012), participants with high and low obsessive-compulsive tendencies (OC+; OC- respectively) were asked to judge the exact midpoint on a continuum of colors. In another study by Rotge et al. (2008), patients with OCD had to compare two images and determine whether they were identical or different. In both studies, obsessive-compulsive tendencies were associated with slower reaction times, presumably driven by the need to check. Such findings have been interpreted as indicating that OC+ participants require more evidence when faced with uncertainty in perceptual decision-making tasks (Banca et al., 2015; Hauser et al., 2017).

The present study focused on the finding that OC+ participants took more time than OC- participants 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, and 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.

While decisions about absence are indeed commonly accompanied by lower levels of subjective confidence compared to decisions about presence (Mazor, Friston, & Fleming, 2020; Mazor, Moran, & Fleming, 2021), they 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.

To illustrate the differences between inference about presence and inference about absence, consider an example. Imagine you are going to a party and have planned to meet your friend there. You arrive at a big house with several floors and many rooms, but unfortunately, your phone battery dies so you can't contact your friend. In the "target-present" scenario, you run into your friend and can conclude with confidence that he is present. However, how does the situation unfold in the "target-absent" scenario, where your friend has not in fact arrived at the party? You navigate the space searching for your friend without success. At what point during this process would you conclude that your friend is absent, that is, that if they were there you would have found them by now? As this example shows, the "target-absent" situation not only introduces a heightened level of uncertainty, but also involves a decision-making process that is fundamentally different in nature than in the target-present scenario.

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 exhibits the classic structure of an OCD symptom, involving an obsessional doubt about potential harm (“Did I run someone over?”) and the compulsion to reduce the doubt by checking (American Psychiatric Association, 2013). It also 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. Other clinical examples include situations in which people are obsessed with the idea that they may have posted something offensive online without realizing it, and are compelled to check their entire browsing history.

These clinical examples hint at 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. Disentangling which of these two factors contributes to obsessive compulsive phenomenology is a challenging task due to the robust increase in subjective uncertainty in decisions about absence. Nevertheless, certain judgments about absence can be made with relative certainty, such as deciding that there is no red dot in an array of blue dots (“absence pop-out”; Mazor and Fleming (2022). Here we made use of this fact to distinguish absence-specific versus generic uncertainty effects in the visual search behavior of OC+ individuals.

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). High and low OC participants 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 replicate the previously reported differences in target-absent search times between OC- and OC+ individuals, and to decouple absence- from uncertainty- related effects. By introducing an easy target-absent condition, we intended to dissociate specific difficulties with inference about absence from more general difficulties with uncertainty. 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, aiming our efforts towards reproducing the original observation of OC+ participants showing longer search times than OC- participants in target-absent trials, but not in target-present ones. Interested readers can find a detailed preregistration document including all our original hypotheses for Experiment 1 at osf.io/yxqu9, and all the additional analysis for our preregistered hypotheses can be found in the appendix of this paper.

Methods

Participants

The research was approved by the Research Ethics Committee of Tel-Aviv University (study ID number 0004169-1). One thousand and four participants were recruited via Prolific (<https://prolific.co/>) and selected based on their acceptance rate (>95%) and for being native English speakers, located in the UK. The entire experiment took 14 minutes to complete (median completion time: 14 min.). Participants were paid £2 for their participation, equivalent to an hourly wage of £8.57. Participants were assigned into OC+ group if they were in the highest quartile in the OCI-R scores distribution, and to the OC- group if they were in the lowest quartile of this distribution. 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 details of our preregistered analysis can be accessed at osf.io/yxqu9.

Measures

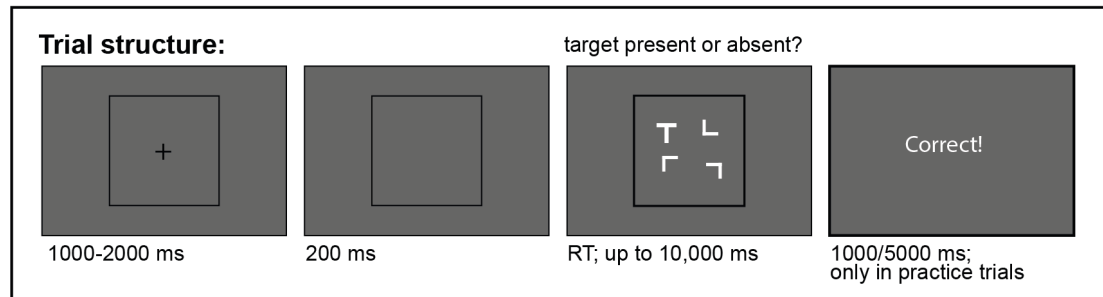
Visual search task

The visual search task described in this study was adapted from Mazor and Fleming (2022), with stimuli intended to replicate the ones used in Toffolo et al. (2013). The task consisted of four blocks, each containing 24 trials of searching for either a closed or an open square. The search items were white on a dark grey background. To ensure that participants understood the task, the experiment 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.

If no response was given within 10 seconds, the next trial immediately appeared. 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. The extended duration of the word “Wrong” was intended to be aversive, to ensure that participants were paying full attention and motivated to provide accurate responses. 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 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: for half of the participants, it was two blocks of ‘easy-search’ followed by two blocks of ‘hard-search’; for the other half, the order was reversed. For all participants, a change in the target was highlighted between the second and third blocks. A message in large font reading “Target changed” was displayed on the screen, followed by a verbal instruction to pay attention to the new target. The trial order within individual blocks was fully randomized (figure 1).

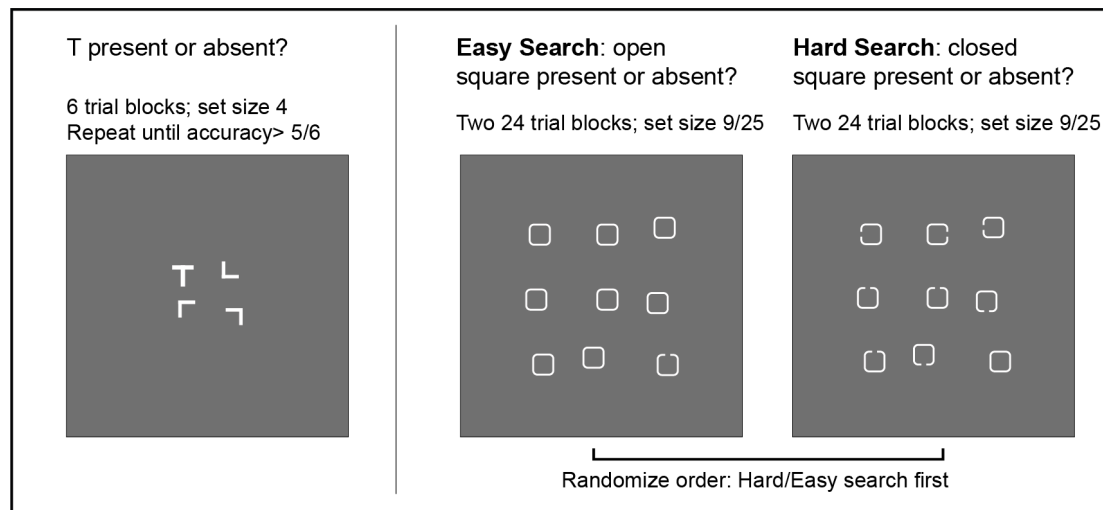
ABSENCE ASYMMETRIES

Visual search



Practice

Main part



Difficulty estimation

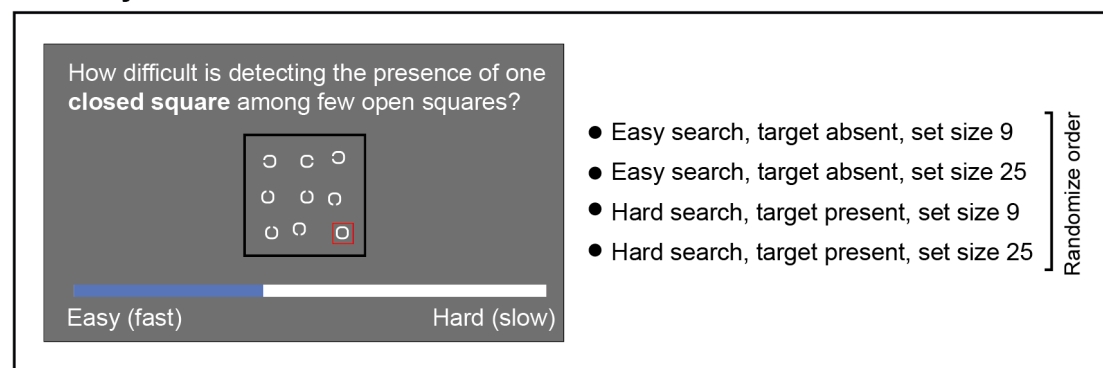


Figure 1 - Overview of experimental Design. Top panel: each visual search trial started with a centered black fixation cross. Middle panel (Practice): After reading the instructions, participants completed practice trials, searching for a rotated T among rotated L's in 6-trial blocks until they achieved 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.

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 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'). Respondents are asked to reflect upon the relevance of each statement to their experiences over the past week. 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.

Procedure

Experiment Structure. 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: the visual search task, questions about the visual search task, and the two inventories: OCI-R and DASS-21. Then, they received written instructions about the visual search task.

Specifically, they were asked to report, as accurately and quickly as possible, whether a target stimulus was present or absent by pressing the J and F keys, respectively. To familiarize participants with the task structure and response keys, they were first presented with six practice trials, in which the target stimulus was a rotated T, and distractors were rotated Ls. Practice trials were delivered in one block of six trials, and the main part of the experiment started only once participants responded correctly on at least five trials.

After completing two of the four visual search blocks, participants were encouraged to take a short break. The main part of the experiment was followed by questions designed to evaluate the subjective difficulty of this part of the 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').

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, Mazor, & Mukamel, 2019).

Data Analysis

Exclusion Criteria

Participants were excluded 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 or above 5 seconds) in more than 25% of the trials. Participants were also excluded from the analysis if they failed one or more of the attention checks. In total, 106 out of 1004 participants were excluded from the analysis. For the remaining participants, error trials and trials with response times below 100 milliseconds or above 5 seconds were excluded from the response-time analysis.

Hypotheses and analysis plan

As detailed above, Experiment 1 was initially designed to test several hypotheses about the behavior of individuals high on obsessive-compulsive tendencies (OC+) in a visual search paradigm. Specifically, Experiment 1 aimed to distinguish between general difficulties with uncertainty and specific issues related to making inferences about absence in OC+ individuals. In the following analysis we focus only on our attempt to replicate the results of Toffolo and colleagues (2013). Results related to our preregistered hypotheses can be found in the appendix.

Transparency and openness

We used R (Version 4.3.1; R Core Team, 2021) and the R-packages: *lvmisc* (Version 0.1.1; Miscellaneous 2021), *broom* (Version 1.0.5; Robinson, Hayes, & Couch, 2021), *caret* (Version 6.0.94; Kuhn, 2021), *citr* (Aust, 2019), *cowplot* (Version 1.1.1; Wilke, 2020), *devtools* (Wickham, Hester, Chang, & Bryan, 2022), *dplyr* (Version 1.1.2; Wickham, Francois, Henry, & Muller, 2021), *ggplot2* (Version 3.4.3; Wickham, 2016), *ggpubr* (Version 0.6.0; Kassambara, 2020), *lme4* (Version 1.1.34; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *papaja* (Version 0.1.1; Aust & Barth, 2022) and *tidyverse* (Version 2.0.0; Wickham et al., 2019) for all our analyses. A detailed preregistration document for Experiment 1 can be accessed at osf.io/yxqu9. All analysis scripts and anonymized data are available at github.com/Noamsarna/ocd_visual_search.

Results

Replication of group differences in target-absent RT

To directly replicate group differences in target-absent RTs (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 tested for an interaction between group and target presence, followed by a pattern wherein 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, $F(1,431) = 1.62, p = .203$ (figure 2, Exp1; preregistered hypothesis 3).

Similarly, when we used our entire sample (the four quartiles), and replaced the group variable (OC+; OC-) with the complete OCI-R scale, we found no interaction between OCI-R score and the presence of the target (preregistered hypothesis 9; $\hat{\beta} = -0.07, 95\% CI [-2.48, 2.35], t(941.54) = -0.05, p = .957$). No significant group differences were observed when accounting for anxiety and depression either (preregistered hypothesis 10; $\hat{\beta} = 8.62, 95\% CI [-21.50, 38.74], t(463.53) = 0.56, p = .575$). There were also no significant RT differences between groups during the task's initial trials, prior to the accumulation of any experience (preregistered hypothesis 8; $F(1,360) = 0.93, p = .336$). Lastly, we observed no significant differences between the groups in their self-reported measures of task difficulty. 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 experiment differed from the original study in several respects. Most notably, search times in this study ($\sim 4.5s$ for target-absent and $\sim 2.6s$ for target-present) were overall shorter compared to those in Toffolo et al. (2013) (~ 5.5 for target-absent and $\sim 3.5s$ 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. In order to directly investigate this hypothesis, Experiment 2 employed the original

stimuli from Toffolo et al. (2013) in an online sample to ascertain whether the failure to reproduce the effect resulted from an excessively easy task.

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 $r = .87, p < 0.001$), and were assigned to the OC+/OC- groups based on the 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.

Material

Visual search task

We used the original stimuli from Toffolo et al. (2013), which were kindly provided to us by Marieke Toffolo. The visual search task consisted of one block of 50 individual search displays, each containing 25 elements. Half of the search displays were target-absent trials, in

which 25 squares with a gap in one of the four edges were presented, and the other half were target-present trials, in which 24 open squares were presented along with one closed square: the target. As in Experiment 1, the experiment began with a practice phase consisting of four search displays (two target-absent and two target-present). In the practice phase, participants received feedback on their response accuracy and could move to the next part of the experiment only after successfully completing all four trials. The practice phase was repeated until the participant's performance was perfect or until it was repeated more than three times, at which point the experiment terminated. In the practice and test phase each trial lasted for a maximum of 10 seconds or until a response was received. If no response was given within 10 seconds, the next trial was presented. In the main part of the experiment, no feedback was given.

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, 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 contain an assessment of perceived difficulty, comprising only a 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 involved the same stimuli as the main blocks. Fourth, participants were instructed to press the spacebar to move 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.

The preregistered analysis plan can be accessed at the following link: <https://osf.io/8a5mr>. As in Experiment 1, we employed the Mersenne Twister pseudorandom number generator to ensure that our preregistration preceded data collection (Mazor et al., 2019). 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

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, $t(174) = 37.53$, $p < .001$)

Direct replication

The primary aim of Experiment 2 was to examine the interaction between group (OC+ vs. OC-) and condition (target-absent vs. target-present). Following the analysis performed in Toffolo et al., (2013), we conducted a one-tailed t-test using the difference in search times (mean RT absent - mean RT present) as the dependent variable and group (OC+ vs. OC-) as the independent variable. 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 revealed no significant differences between the groups, $t(144.88) = 1.41, p = .081$, providing no evidence for the expected interaction. Notably, the numeric trend of the interaction in our sample was driven by OC+ shorter response time than OC- in target-present trials, rather than by longer response time 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).

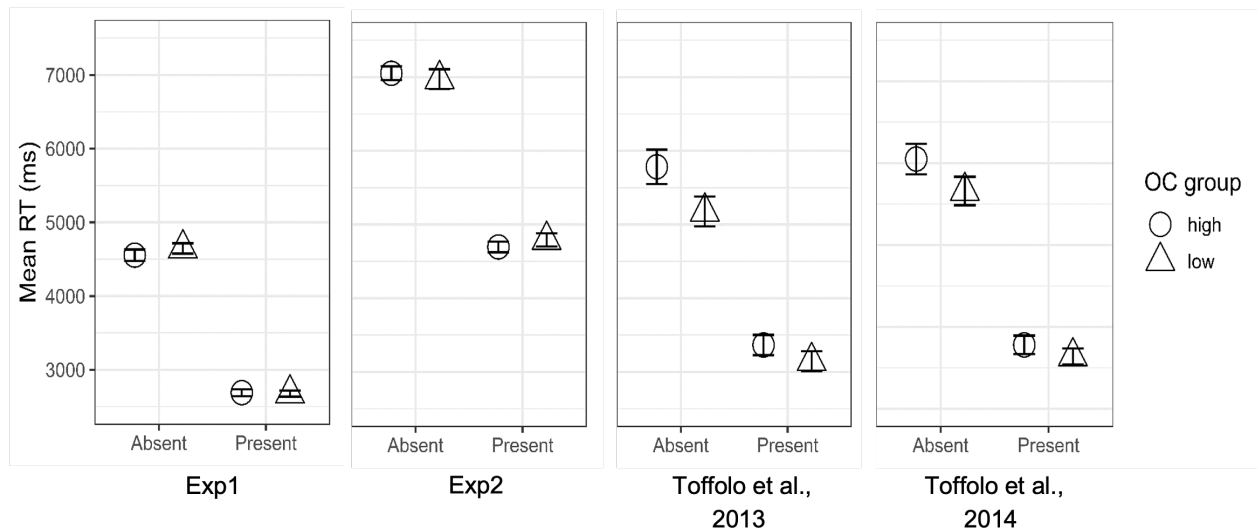


Figure 2 - Results from Experiment 1 (preregistered H3), Experiment 2 (preregistered H2) and Toffolo et al., 2013, 2014. Mean reaction times for target-absent and target-present trials (X-axis). Error bars represent the standard error of the mean. Shapes represent the OC groups: Circle for OC+; Triangle for OC-.

Discussion

In two preregistered, large-sample studies, we found no evidence of prolonged search time among OC+ participants in target-absent trials. Experiment 1 was a conceptual replication, ensuring that the basic structure of the task remained consistent with Toffolo et al., (2013), while varying the uncertainty level and manipulating stimulus set size. No significant differences emerged between the search times of OC+ and OC- participants across any conditions. Subsequently, in Experiment 2, we meticulously replicated the original study methodology and design, employing identical stimuli, task structure, and instructions. Nevertheless, Experiment 2 yielded no evidence of group differences in search times. Collectively, these null results are in disagreement with the previous findings by Toffolo and colleagues (2013, 2014, 2016). In discussing these results below, we examine in detail all the potential disparities between our studies and the original studies we sought to replicate.

Differences between the two studies

Experiment 1 was a conceptual replication attempt. Our main objective was to disentangle whether the prolonged search time of OC+ participants observed in target-absent trials was primarily due to a general sensitivity to uncertainty (Tolin, Abramowitz, Brigidi, & Foa, 2003) or alternatively reflected a more specific difficulty with inference about absence (Mazor, 2021). However, no group differences materialized in any of the tested conditions, and consequently there were no differential patterns to account for. We considered the possibility that search difficulty in Experiment 1 might have been insufficient to elicit doubt and trigger repetitive checking (reflected in longer RT). In comparison to Toffolo et al., (2013), in Experiment 1 we used a smaller and denser search grid, with thicker stimulus outlines, resulting in an overall easier

task. Indeed, mean search times in Experiment 1 were faster by about one second relative to those in Toffolo et al. (2013). In addition, the grid in Experiment 1 was small and organized enough to theoretically allow sequential search (going over the grid line by line, which could also reduce the level of uncertainty). With these differences in mind, we designed Experiment 2, in which we aimed to replicate the original study design as close as possible. In Experiment 2 we used the same stimuli, task structure and instructions as in Toffolo et al., (2013). In accordance with Experiment 1, however, we again found no group differences between OC+ and OC- participants in search time. In this experiment, the overall (collapsing over groups) mean search time was slower than both Toffolo et al., (2013) and Toffolo et al., (2014), and more similar to the mean search time in the clinical OCD sample in Toffolo et al., (2016), ruling out the possibility that insufficient task difficulty might explain the lack of differences between the groups (see Fig. 2).

Online vs. in person design

The most notable difference between our experiments and those conducted by Toffolo et al., (2013, 2014, 2016) lies in our use of an online setting versus their use of in-person data collection. Completing tasks online, as opposed to in a laboratory setting, generates more technical noise: unexplained variance driven by technical variation. This technical noise can be categorized into two main types: noise stemming from disparities in software and technology and noise resulting from technical and situational heterogeneity, such as variable computers, monitors, or distances from the screen (Hilbig, 2016). In the following paragraphs, we consider the potential impact of these two factors on the different results observed.

First, differences in technical noise can be instigated by differences in software: Matlab in previous studies of visual search in OCD, versus JavaScript in ours. In an effort to shed light on

ABSENCE ASYMMETRIES

this specific matter, Leeuw and Motz (2016) carried out a study comparing the response time distributions derived from a lab-based psychophysics toolbox (Matlab) and an online experiment powered by JavaScript. The results revealed near-identical response times between the JavaScript and Matlab experiments, with a negligible and constant delay for the JavaScript engine of approximately 25 milliseconds. Therefore, Leeuw and Motz (2016) concluded that the two setups demonstrated equal sensitivity to response time alterations resulting from experimental manipulations.

Second, participants in our sample completed the visual search task in their own homes, using a range of computers, displays, and keyboards, rather than in a controlled lab environment with a fixed screen as in Toffolo et al. (2013, 2014, 2016). Importantly, a consistent finding across research on online studies is that technical variation does not pose a significant issue (Hilbig, 2016). 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).

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). Participants were randomly assigned to complete a lexical decision task either in a lab setting with standard software for psychological experiment, in a lab using a browser version, or on the web using their own computer and space. The results showed that the words 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 resulting from experimental manipulations. Even if we consider the constant delay of JavaScript compared to Matlab tasks (~25ms, Leeuw and Motz (2016)), and add to it to the difference between a lab setting with E-prime and a web setting with a browser (34 ms, Table 2 in Hilbig (2016)) we get a difference of ~60 ms, an order of magnitude smaller than the difference between OC+ and OC- participant in target-absent searches reported by Toffolo et al., (2013), which was ~600 ms.

Comparison of accuracy

One discrepancy between our dataset and the one reported by Toffolo et al., (2013) is worth highlighting. The mean error rate across all trials in Toffolo et al. (2013) ($M=5.87$, $SD=5.96$) was lower than in our Experiment 2 ($M=8.51$, $SD=4.01$). To examine whether less accurate or possibly less engaged participants could have skewed our results, we reevaluated the data from Experiment 2, this time including only participants whose error rate fell below the raw error rate exclusion criterion (19 errors out of 50 trials) employed by Toffolo et al. (2013). Despite applying this strict exclusion criterion, the crucial interaction between group and condition did not reach statistical significance, as elaborated in the supplementary results.

Feeling observed

An interesting perspective on our failure to replicate a well-documented psychological finding comes from a recent paper discussing the failed attempts to replicate the widely recognized facial-feedback hypothesis (Noah, Schul, & Mayo, 2018b). According to this paper, both the original study and its replication failures were valid, with the difference between the two

outcomes attributed to the presence of a camera. The theory proposed to elucidate this disparity suggests that individuals alter their behavior when they perceive themselves to be observed by others. The theory posits that individuals, feeling watched, adopt an external perspective on themselves, thus shifting their reliance from solely internal information to information perceivable by potential observers. In a series of studies supporting this theory, the researchers showed that when people are in an ‘objective state’, where they think of themselves from the perspective of an external observer, they are less influenced by subjective experiences such as metacognitive experiences (Noah, Schul, & Mayo, 2018a). This hypothesis introduces the possibility that the eye tracker used in Toffolo’s experiments, while clearly different from a conventional camera, may have created a sense of being observed, leading participants to lessen their reliance on internal cues.

As previously noted, the psychological processes required for making target-present and target-absent decisions are distinct, with interesting links to the dichotomy of feeling observed versus not. Decisions regarding target presence can be readily justified to an outsider (“Look at the second row, the third item on the left; there it is!”) in a way that is not true of decisions about target absence (“Look at the grid; do you see how the target isn’t there?”). Therefore, if individuals indeed adopt an external perspective of themselves when feeling observed, this may impact their decision-making to different extents as a function of target presence or absence. We leave it for future studies to explore this hypothetical link and its potential interaction with obsessive-compulsive tendencies.

Inference about absence in the clinic and in the lab

Having analyzed the disparities between our experiments and those we attempted to replicate, we turn to revisit the earlier discussion on clinical manifestation of inference about absence in OCD. Despite a promising theoretical relationship and a cluster of related clinical manifestations ('hit and run OCD'), we found no effect of obsessive-compulsive tendencies on inference about absence. This failure may be rooted in the notion of perceived responsibility. Salkovskis (1985) posits that individuals with OCD frequently demonstrate an inflated sense of responsibility, perceiving themselves as key players in harm prevention. This heightened sense of responsibility gives rise to compulsive behavior directed at mitigating potential threats (Salkovskis, 1985). In 'hit and run OCD' behavior, the compulsion does not directly aim to prevent future harm but rather to verify that it has not occurred and possibly avoid secondary damage. Importantly, responsibility was not manipulated in the visual search task utilized in our study, as missing the target had virtually no adverse effects.

Furthermore, in the clinical example of "hit and run", the compulsion is associated more with a recollection of an event rather than its direct experience. For example, an individual may obsess over the possibility that they may have run someone over without realizing, which primarily involves their memory of the event. 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)). This implies that cognitive biases and distortions in OCD could predominantly affect memory functions, more than perception. Future research exploring inference issues about absence in OCD may consider focusing on memory and incorporate a condition that involves personal responsibility.

Before concluding, we wish to examine our study through the lens of replicability. In an influential publication, Tackett et al. (2017) strongly advocated for broadening the discussion on replicability in psychology, emphasizing the need for clinical psychologists to rigorously implement replicability practices. Our attempt at replication embedded several of the pivotal elements underscored by Tackett et al. (2017). Firstly, prior to data collection, we documented our hypotheses, analysis plan, and criteria for data exclusion, ensuring these decisions were established prior to the collection and review of data. Furthermore, we utilized the preregistration time-locking tool as detailed by Mazar et al. (2019), thereby guaranteeing that our registration indeed preceded the data collection process. Secondly, while the effect of prolonged search time for OC+ has been replicated twice by Toffolo and her colleague (2014, 2016), our study represents the first independent replication attempt. Thirdly, in both experiments, we conducted a power analysis to determine our sample size, and in Experiment 2, adhered to the same criteria for determining OC+ and OC- used by Toffolo et al., (2013) to minimize population deviations. Lastly, in the interest of promoting research transparency and reproducibility, we have made our raw data, analysis scripts, and task codes publicly available. By incorporating these practices, we have strived to enhance the robustness of our study, hoping it may serve as a helpful reference for future research into visual search patterns in OCD.

Conclusion

The presented findings diverge from those of previous studies, as we were unable to replicate the effect of prolonged search time for high 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, this study serves to advocate for the application of open science practices in clinical psychology research,

consistent with those widely utilized within the realms of social and cognitive psychology, in order 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.*
- Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations.* Retrieved from <https://github.com/crsh/citr>
- Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown.* Retrieved from <https://github.com/crsh/papaja>
- 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>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 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>

Diagnostic and statistical manual of mental disorders (DSM-5®). (2013). American Psychiatric Pub.

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.
- Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots*. Retrieved from <https://CRAN.R-project.org/package=ggpubr>
- Kuhn, M. (2021). *Caret: Classification and regression training*. Retrieved from <https://CRAN.R-project.org/package=caret>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

<https://doi.org/10.18637/jss.v082.i13>

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. (2018a). Thinking of oneself as an object of observation reduces reliance on metacognitive information. *Journal of Experimental Psychology: General*, 147(7), 1023–1042. <https://doi.org/10.1037/xge0000440>

Noah, T., Schul, Y., & Mayo, R. (2018b). 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>

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York: Springer. <https://doi.org/10.1007/b98882>

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

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>

Robinson, D., Hayes, A., & Couch, S. (2021). *Broom: Convert statistical objects into tidy tibbles*.

Rotge, J. Y., Clair, A. H., Jaafari, N., Hantouche, E. G., Pelissolo, A., Goillandeau, M., ... Aouizerate, B. (2008). A challenging task for assessment of checking behaviors in obsessive–compulsive disorder. *Acta Psychiatrica Scandinavica*, 117(6), 465–473. <https://doi.org/10.1111/j.1600-0447.2008.01173.x>

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.

Shapiro, D. (1965). *Neurotic styles*.

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>

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., & Foa, E. B. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *Journal of Anxiety Disorders*, 17(2), 233–242.

[https://doi.org/10.1016/S0887-6185\(02\)00182-2](https://doi.org/10.1016/S0887-6185(02)00182-2)

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.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron*, 96(2), 348–354.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
Retrieved from <https://ggplot2.tidyverse.org>

Wickham, H. (2021). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., Francois, R., Henry, L., & Muller, K. (2021). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2022). *Devtools: Tools to make developing r packages easier*. Retrieved from <https://CRAN.R-project.org/package=devtools>

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=cowplot>

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 continuous scale, dragging the cursor 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.

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 osf.io/yxqu9. 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, $t(1,795) = -57.77, p < .001$. 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, $t(897) = -3.34, p = .001$ (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 $F(1,431) = 0.00$, $p = .993$, 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. However, given our failure to replicate a group difference in the hard search (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, $F(1,431) = 1.25$, $p = .265$.

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). 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 compare model performance using their Akaike Information Criterion (AIC), with lower AIC values indicating a better fit of the model to the data (Burnham & Anderson, 2004). 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.

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.

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 $t(361) = -13.43$, $p < .001$. 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 $F(1,360) = 0.93$, $p = .336$.

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 ($\hat{\beta} = -0.07$, 95% CI $[-2.48, 2.35]$, $t(941.52) = -0.05$, $p = .957$).

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 ($\hat{\beta} = 8.62$, 95% CI $[-21.50, 38.74]$, $t(463.53) = 0.56$, $p = .575$).

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, 2017). 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 A1).

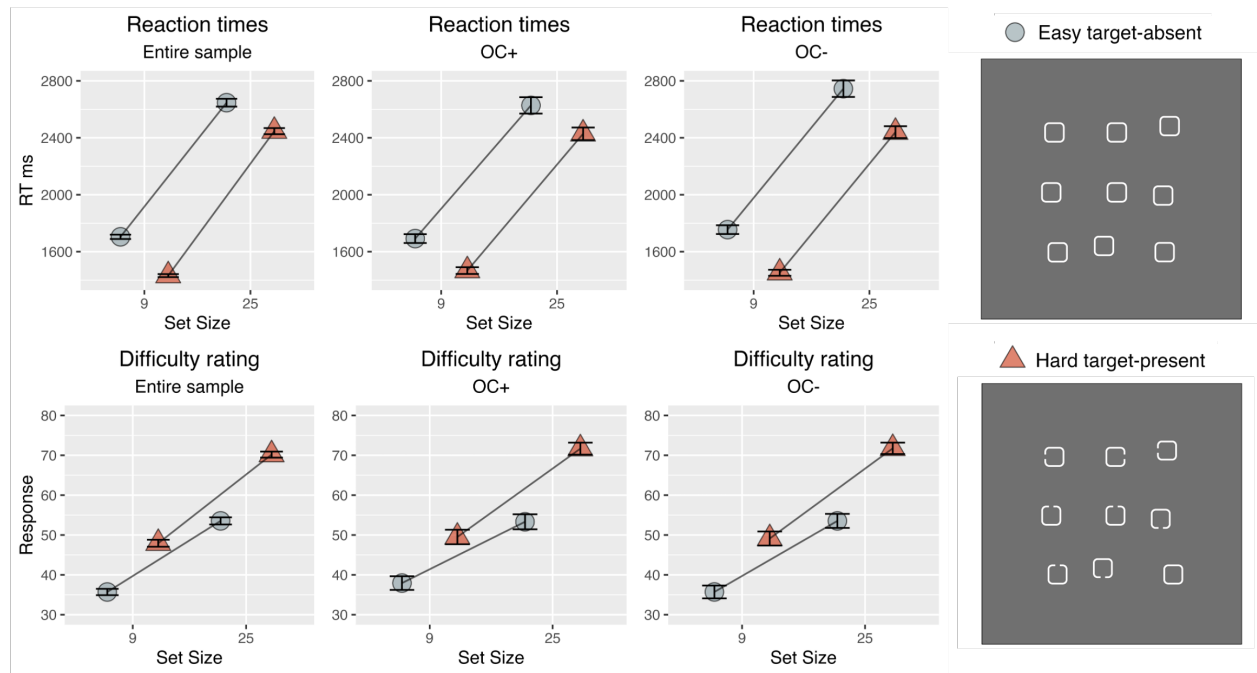


Figure A1 - Relationship between perceived difficulty and reaction times (RT) for high and low OC participants. Rows display mean RT (top-row) and mean estimated difficulty (bottom-row). Columns represent results for the entire sample (left column) the OC+ group (Middle column) and OC- (Right column). 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

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, $t(142.07) = 1.47$, $p = .072$, 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.