Incidental Attitude Formation via the Surveillance Task: A Pre-Registered Replication of Olson and Fazio (2001)

Supplementary Online Materials - Reviewed

# Deviations from the Preregistration

In order to maximize evidential value and transparency, we document all divergences from the preregistration/Stage 1 accepted manuscript to Stage 2 manuscript below.

## Change in terminology from ‘confirmatory’/‘exploratory’ to ‘primary’/‘secondary’ analyses

After writing the Stage 2 manuscript and soliciting comments from the co-authors, there was consensus that the terminology of ‘confirmatory’ vs. ‘exploratory’ analyses was confusing given that all analyses were preregistered (both descriptions and the code implementing them). However, we were also acutely aware of the potential pitfalls of relabeling these analyses given the Registered Report format. We therefore sought advice from Christ Chambers, creator of the Registered Report format and editor for a large number of RR articles to date, about the relative benefits and costs of changing vs. not changing this terminology. His expert opinion was that the term ‘exploratory’ should not be employed within a preregistered analysis. As such, we have changed the Stage 2 manuscript to refer to ‘primary’ analyses (i.e., those that most directly replicate the original Fazio & Olson, 2001 study) versus ‘secondary’ analyses (i.e., those that test the robustness of the EC effect to other exclusion criteria). We felt that this modification to the Stage 1 accepted manuscript was justified on the basis of improving clarity and readability. This change, along with reference to this document, is now footnoted in the manuscript.

## Data collection stopping rule

Due to unforeseen delays, one site was unable to collect data from the specified planned number of participants (100 to 150 per site) within the informally agreed upon timeframe. We provided this lab will additional time insofar as was possible. However, we realized that no maximum timeframe was specified in our preregistration. In order to resolve the situation, we made an updated preregistration that modified our data collection stopping rule (see [osf.io/uyng7](https://osf.io/uyng7)). This updated preregistration (made on 2020-02-11) is discussed in the manuscript. It specified that we would instead use all data collected from all sites, even those who had not met the originally planned sample sizes, and set a hard deadline for data collection after which any and all data from each site would be used (2020-02-19). This also accommodated sites that collected data from more participants than planned in our preregistration. This modification was deemed to be consistent with our meta analytic approach within the preregistered analyses (i.e., even small samples sizes make meaningful contributes as the estimation of the meta effect size, as the uncertainty around all effect sizes is quantified within the meta-analysis models). This decision was driven in large part by the fact that this lab was that of one of the original authors, who we felt it was therefore particularly important to include in the replication.

## Method of calculating confidence intervals

The preregistered implementation of the analyses employed a bootstrapping method to calculate effect sizes at each site prior to meta-analyses. However, due to the change in the data collection stopping rule (see above) one site collected a far smaller than predicted number of participants (*n* = 21). Heterogeneity metrics (e.g., *I*2 and *H*2) were observed to computationally unstable when re-running the analysis script. For the sake of computational reproducibility, we therefore exchanged the bootstrapping method for the arithmetic method throughout. Inspection of the effect sizes and CIs suggested the impact of this decision on the meta effect size estimates and its confidence intervals was less than Hedge’s *g* = 0.01.

## *z* and *p* values for ‘aware’ participants

Due to an oversight, no method to calculate *z* or *p* values was specified or implemented in our written preregistration or preregistered code. The preregistered implementation of the moderator meta-analysis models return values for the difference in effect size between the two subsets (‘aware’ vs. ‘unaware’) and the *p* and *z* values for this difference, but not values for each subset. While the preregistered models are fit for their primary purpose (i.e., testing moderation by awareness), our Stage 1 accepted manuscript stated that we would also report *z* and *p* values for each subset. In order to employ the identical method to how these values were calculated for the ‘unaware’ subset, we therefore fitted (non-moderator) meta-analyses to just the aware subset of participants. The only results reported from these models were the effect sizes, 95% CIs, *z* and *p* values.

## Description of the exclusion criteria

After writing the Stage 2 manuscript and soliciting comments from co-authors, there was consensus that the description of the four exclusion criteria was unclear and confusing. We therefore elected to rewrite this section (pp. 16-19 in the manuscript). Importantly, it is only the *description* *in the manuscript* of what these criteria consisted of and how they were applied that changed. Their implementation did not change between preregistration/Stage 1 acceptance and the Stage 2 manuscript. In fact, the revised descriptions of the criteria in this section are more closely aligned with the actual preregistered protocol and instructions distributed to the sites than the descriptions in the Stage 1 accepted manuscript. We therefore felt this this modification to the Stage 1 accepted manuscript was justified on the basis of improving clarity and readability.

## Non-preregistered analyses

All non-preregistered analyses are clearly marked in both the code implementation and the manuscript. These fully reported in the manuscript.

# Sample Size and Characteristics

Table S1 below details the sample size and sample characteristics at each site and percent of exclusions for each of the contingency awareness/recollective memory exclusion criteria.

Table S1

Sample size, sample characteristics, and percent of exclusions for each of the contingency awareness/recollective memory exclusion criteria, as a function of data-collection site.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **Age** | | **Gender** | | | | **Percent excluded** | | | | |
| **Site** | ***n* manual exclusions** | ***n* for analysis** | **Mean** | **SD** | **Female** | **Male** | **Other identity** | **Did not answer** | **Surveillance task performance** | **Olson & Fazio (2001)** | **Olson & Fazio (2001) modified** | **Bar-Anan et al. (2010)** | **Bar-Anan et al. (2010) modified** |
| **Balas** | 6 | 100 | 26.5 | 4.7 | 57 | 43 | 0 | 0 | 3.0 | 2.1 | 19.6 | 41.2 | 16.5 |
| **Mierop** | 1 | 99 | 21.7 | 4.2 | 66 | 33 | 0 | 0 | 2.0 | 8.2 | 17.5 | 43.3 | 21.7 |
| **Gast** | 0 | 120 | 23.6 | 7.2 | 91 | 26 | 1 | 2 | 2.5 | 6.0 | 26.4 | 49.4 | 24.7 |
| **Gawronski** | 0 | 155 | 18.9 | 1.1 | 113 | 41 | 1 | 0 | 2.6 | 7.2 | 74.1 | 51.2 | 30.2 |
| **Hütter** | 2 | 148 | 22.7 | 6.2 | 109 | 39 | 0 | 0 | 1.4 | 18.4 | 41.6 | 57.3 | 43.6 |
| **Kurdi** | 0 | 151 | 19.3 | 1.3 | 120 | 31 | 0 | 0 | 1.3 | 8.0 | 21.4 | 39.4 | 21.4 |
| **Moran** | 1 | 99 | 20.0 | 3.2 | 75 | 24 | 0 | 0 | 1.0 | 2.0 | 28.6 | 46.9 | 27.6 |
| **Olson** | 0 | 21 | 20.0 | 0.0 | 10 | 11 | 0 | 0 | 0.0 | 9.5 | 28.6 | 42.9 | 33.3 |
| **Douglas** | 0 | 148 | 18.6 | 0.8 | 98 | 50 | 0 | 0 | 2.0 | 6.9 | 19.9 | 58.2 | 35.6 |
| **Stahl** | 0 | 100 | 21.7 | 5.1 | 80 | 20 | 0 | 0 | 3.0 | 13.4 | 32.0 | 54.6 | 35.1 |
| **Unkelbach** | 0 | 142 | 23.6 | 7.0 | 82 | 57 | 1 | 2 | 1.4 | 10.0 | 36.3 | 51.2 | 29.9 |
| **Vadillo** | 0 | 195 | 19.9 | 3.0 | 166 | 25 | 3 | 1 | 1.5 | 1.0 | 15.0 | 39.3 | 12.9 |

*Note.* Each lab is identified by the last name of the corresponding author. *n* manual exclusions: exclusions made manually before the analysis due to incomplete data file (1 case at Moran’s site, 2 cases at Hütter’s site), technical problems (4 cases at Balas’s site), unusual participant behaviour (1 case at Balas’s site), participant eligibility (1 case at Balas’s site), and data recoding issues (1 case at Mierop’s site). *n* for analysis: represents the sample size after the manual exclusions. Age and gender are characteristics are calculated from the sample for analysis after manual exclusions. Percent excluded surveillance task performance: percent of exclusions based on the number of errors made during the surveillance task (percentage accuracy < mean – 3 SD per site). Percent excluded for Olson & Fazio (2001), Olson & Fazio (2001) modified, Bar-Anan et al. (2010), and Bar-Anan et al. (2010) modified represent the percent of the sample excluded *after* surveillance task exclusions had been excluded. These mirror the way these exclusions have been reported in the manuscript.