**The Influence of Extinction and Counterconditioning Procedures on Operant Evaluative Conditioning and Intersecting Regularity Effects**

**Hughes, Mattavelli, Hussey, & De Houwer (2020)**

***Deviations from Preregistration***

**Sample size**

We originally indicated in Experiments 1-7 that with an average effect size Cohen’s *d* = 0.50, α = .05, we would need a total of 50 participants per cell to get a power (1 – β) = .80 (i.e., 100 participants in Experiments 1-6 and 300 in Experiment 7). However, after data treatment and exclusions the final sample sizes were a little smaller than anticipated: *n* = 97 in Experiment 1, 94 in Experiment 2, 95 in Experiment 3, 98 in Experiment 4, 95 in Experiment 5, and 90 in Experiment 6 (note: there were 313 participants in Experiment 7). Notably, we carried out a series of (non-pre-registered) meta-analyses to address our study level experimental questions with sufficient power.

**IR memory tests**

In our pre-registered documents we labelled one of the tasks as an “IR memory test”. However, when revising the manuscript, we realized that this task actually assessed for the stimulus-response relations and response-outcome relations that were trained during the acquisition phase and altered during the intervention phase. This task was also delivered twice during the experiment: once after the acquisition training phase and again after the intervention phase (in Experiments 1-3 and 5-7). With this in mind, we thought it more accurate to label the first instantiation of the task as an “acquisition test phase”, and the second as an “extinction test phase”, or a “counterconditioning test phase”, given that it was assessing the relations trained in those respective tasks. In short, the labelling of the same task differed between pre-registration and the final manuscript.

**Influence awareness**

In the pre-registration document we referred to “influence awareness” and indicated that participants were asked three questions: one on demand, another on reactance, and third on believability. It would be more accurate to indicate that they were asked five questions: one on demand, one on influence awareness, one on reactance, and two on believability. The manuscript was revised to reflect this. We also placed the confidence in self-reported ratings measure with the exploratory measures in the manuscript (given that the data was always exploratory and never intended for use in the main manuscript).

**Analyses**

**IAT scoring algorithm**. In our pre-registered document we incorrectly indicated that we would use the D1 IAT scoring algorithm. In the manuscript we used the D2 algorithm instead given that it is a more widely used metric in the IAT literature, and has a simple implementation in a dedicated R package (IATScores).

**t-tests in place of ANOVAs**. We originally proposed that we would carry out a 2 (Stimulus: T1 vs. T2 or O1 vs. O2) x 2 (Training Condition: Intervention vs Acquisition-Only) ANOVA on the self-reported and IAT scores in Experiments 1-6 (and a 2 (Stimulus) x 3 (Training Condition: Extinction vs. Counterconditioning vs Acquisition-Only) ANOVA in Experiment 7). However, we later thought our question could be more parsimoniously answered using a t-test comparing self-reported ratings or IAT scores in the acquisition to the intervention groups (i.e., by controlling for the method factors via counterbalancing and focusing purely on the impact of the intervention on evaluations). The latter analyses were ultimately reported in the final manuscript.

**Behavioral intentions**. Our pre-registered documents omitted reference to the behavioral intentions data, their treatment, and subsequent analyses. In the final document we provide all this information. With respect to treatment, only results from the T1-T2 comparison are reported (i.e., analyses were only carried out on responses that involved participants selecting either T1 or T2 and not on the selection of neither target, both targets, or non-responses). Counts of each response for each study and experiment condition were calculated, which were then used to calculate an odds ratio. *p* values were computed via Fischer’s exact test. Haldane-Anscombe corrections were applied to studies where at least one cell contained zero counts (i.e., counts in all cells were increased by 1). This was done for the outcome (OEC) and target (IR) stimuli.

**Inclusion of training and testing analyses**. During the review process a reviewer asked that we provide information showing that participants paid attention during the learning task and that they were successful in those tasks. We therefore included a new set of analyses – namely – a measure of the mean and SD of accuracy within each phase of the learning task (see Table 1 in the main manuscript) and the percentage of participants who successfully passed the final block of the training and testing phases (see Table 2 in the main manuscript). We also acknowledge that these additions were post-hoc and on the recommendation of the reviewer in the manuscript (see footnote 9).

**Inclusion of meta-analyses**. We also opted to include a series of meta-analyses in order to ask three general questions about our findings that individual studies lacked the power to address or to make general conclusions from: (a) do OEC and IR procedures give rise to evaluations *in general*, (b) are evaluations moderated by extinction or counterconditioning *in general*, and (c) do those effects differ when we exclude participants who failed the learning task? Analyses were conducted using the metafor R package (Viechtbauer, 2010). All models employed a Restricted Maximum Likelihood estimator function. In each case, study was entered as a random intercept in order to acknowledge the non-independence of each study’s outcome variables, and outcome variable type (i.e., IAT, self-reported evaluations, behavioural intentions) was entered as a random slope in order to acknowledge that changes of different magnitudes may be observed between them. Prior to meta-analysis, behavioral intention data were converted from Odds Ratios to Cohen’s *d* scores using the method specified by Hasselblad and Hedges (1995; see also Sánchez-Meca, Marín-Martínez & Chacón-Moscoso, 2003) which has been shown to balance ease of use, bias, and coverage. Meta-analyses were not pre-registered, although the hypotheses assessed within them are similar to the those pre-registered in the individual experiments.

**Sensitivity analyses**. We also opted to include a series of sensitivity analyses to determine if the conclusions we made on the basis of the total sample were congruent or incongruent with those obtained when only the data from those in the pass group was considered (for our rationale here see the main manuscript). These analyses are reported in the Supplementary Materials section and were also not pre-registered.