# 02/27/2023 Amendments

**Authors**

Added author (MCF)

**Research Questions**

* In addition to collecting ratings of the extent to which they have the motivation and opportunity to adjust responses, participants will rate the extent to which they (a) believe the experimenter’s hypothesis and (b) believe others will adjust their responses to fit to the experimenter’s hypothesis.

**Hypotheses**

* For Hypothesis 1, Coles et al. (2022) predicts that demand effects will be moderated by belief.
* Tests of higher-order interactions (and their decompositions) will be considered exploratory because we are convinced that the measures and models will be too noisy to yield reliable conclusions.

**Explanation of existing data and sample size**

* Specified that the current number of observations (k = 232, s = 34) is likely to change as we double-check coding decisions, refine inclusion criteria, and discover additional relevant records.

**Measured variables**

* Plans to collect post-hoc ratings of motivation and opportunity are no longer described as preliminary. Instead, it is contingent on the availability of resources.
* We will also collect belief and prediction ratings from 100-250 participants.
* Motivation will be measured on the following scale: -3 = Extremely motivated to adjust response to be inconsistent with the researchers' stated hypothesis; 3 = Extremely motivated to adjust response to be consistent with the researcher's stated hypothesis
* Ability to adjust responses will be measured (0 = “extremely incapable” to 4 = “extremely capable)
* Belief in the experimenter’s hypothesis will be measured (0 = “extremely incapable” to 4 = “extremely capable)
* Predictions about whether other participants will confirm the researcher’s hypothesis will be collected (-3 = “extremely likely to adjust responses to be inconsistent” to 3 = “extremely likely to adjust responses to be consistent”)
* Each participant will review 10 vignettes. After each vignette, they will be asked to identify the researcher’s stated hypothesis. We plan to exclude ratings in cases where the correct hypothesis is not identified.
* We will not include nil-hypothesis comparisons in our vignette analyses because our coding strategy cannot accommodate the potential moderating role of motivation and belief in this condition. For example, imagine that a participant is (a) told that an intervention will not impact mood (nil demand), and (b) is extremely motivated to disconfirm the hypothesis. Relative to a control condition, this participant could disconfirm the hypothesis by either increasing or decreasing their mood report. Thus, even if motivation does moderate the effects of demand characteristics, we would not expect a systematic pattern to emerge with our coding scheme.
* For two of these vignettes (the positive- and nil-hypothesis conditions in Coles et al., 2022, Study 2), participants will complete the actual study procedures. This means they will pose happy and neutral facial expressions across two blocks and subsequently self-report their emotional experience. These data could allow us to examine – in an exploratory manner – the extent to which hypothetical ratings match participants’ actual responses in an experiment with explicit demand characteristics.

**Statistical Models**

* We originally planned to analyze our data using meta-analysis with robust variance estimates, but were convinced that three-level meta-regression would allow us to better describe sources of variability. In general, this did not substantially change our parameter estimates or conclusions.
* To estimate the overall effect size, we will fit an intercept-only 3LMA model. For moderator analyses, continuous and dummy-coded categorical moderators will be separately entered into the model. For categorical moderators, we will use the models to estimate overall effect sizes within each subgroup of the moderator.
* We originally planned to conduct PET-PEESE with aggregated dependent effect sizes, but later decided to follow advice from a simulation study and use PET-PEESE in a three-level meta-analysis (Rodgers & Pustejovsky, 2021). We will report both in the manuscript if their results diverge.

# Summer, 2024 Amendments

**Authors**

Added author (MW)

**Research Questions**

* Are demand effects moderated by the quality of the record, as assessed using a modified version of the Downs and black (1998) checklist?

**Data Collection Procedures**

* Added two unplanned literature searches based on reviewer feedback. First, we repeated our original search to identify records published since January 2022. Second, we performed an additional APA PsychInfo search using the following search terms: “participant role” OR “demand effects” OR “good subject effect” OR “expectancy effect” OR “evaluative apprehension”
* Found additional unpublished records

**Inclusion Criteria**

* We will exclude records wherein (a) the predicted effect of awareness is unclear, (b) the population or intervention was clinical, and/or (c) the DV described in the demand manipulation was not actually measured.

**Measured Variables**

* Will transform Cohen’s d to Hedge’s g based on reviewer feedback
* Will use the proportion of participants who passed an attention check as a proxy for receptivity
* Will add the Coles et al. (Study 2) data to the meta-analysis, but not describe in the manuscript because reviewer impressions were not generally positive.
* We will collect more vignette ratings because a reviewer pointed out that the measures may be unreliable (thus necessitating more data to converge on accurate point estimates). For sample size planning, we estimated how many observations would be needed to decrease the length of the confidence intervals to 1 for ratings of motivation to adjust responses, opportunity to adjust responses, belief in the experimenter’s hypothesis, and predictions about whether other participants will respond to demand characteristics. We did so using the presize R package and point estimates of the M and SD of the four measures.

**Statistical Models**

* Will use cluster-robust estimation procedures for estimating overall effects, testing for moderators, estimating subgroup effects, and conducting precision effect tests per a reviewer request