Summary of Updates

On 10/11/2022, we added details about our plans to collect post-hoc ratings of (a) motivation to adjust responses, (b) ability to adjust responses, (c) belief in the communicated hypotheses, and (d) the predicted likelihood that other participants would adjust their responses. We also made amendments to our analysis strategy (e.g., switching from robust variance estimates to three-level meta-analysis). Changes are denoted in green.

Title

A meta-analysis of the effects of experimentally-manipulated demand characteristics

Authors

Nicholas A. Coles & Michael Frank

Research Questions

Overall

What is the overall effect of explicit demand characteristic manipulations on their referenced outcome(s)-of-interest?

To what extent are the effects of explicit demand characteristics manipulations consistent across studies?

Study Characteristic Moderators

We will examine if the effect of explicit demand characteristics manipulations are moderated by any of the following variables:

- Whether participants are students (vs. online workers or a mix)

- Whether participants are paid (vs. volunteers)

- Whether the study was conducted in-person (vs. online)

- Whether the effect size compares positive demand characteristics information, negative demand characteristics information, nil demand characteristics information, and/or no demand characteristics information (i.e., a control group)

- Whether the effect size compares two vs. one demand characteristics manipulation

- Whether the effect size comparison is between- or within-subjects

- Whether the study is published

Conceptual Moderators

In a follow-up survey, we will collect post-hoc ratings of three potential moderators:

- Motivation: Perceived motivation to confirm the experimenter's hypothesis

- Opportunity: Perceived opportunity to control the outcome-of-interest

- Belief: Belief in the experimenter's hypothesis

Participant prediction

In the same follow-up survey, we will collect post-hoc ratings of the extent to which people believe that *other* participants will respond to the demand characteristics manipulations.

Hypotheses

This work is mostly exploratory. However, for the conceptual moderators, we are interested in testing the following hypothesis:

H1: The effect of demand characteristic manipulations will be moderated by motivation and opportunity (as predicted by Coles et al. 2022; Rosnow and Rosenthal, 1997) and belief in the experimenter's hypothesis (as predicted by Coles et al. 2022).

We originally intended to test the following hypotheses in a confirmatory fashion. However–before the necessary data were collected–we were convinced that the measures and models would be too noisy to yield reliable conclusions. Thus, we now plan to examine them in an exploratory manner.

E1: For the effect of demand characteristics manipulations, there will be a two-way interaction between motivation and opportunity. The simple effect of motivation will be larger when opportunity is high vs. low (as predicted by Coles et al. 2022; Rosnow and Rosenthal, 1997)

E2: When participants are not motivated to confirm or disconfirm the experimenter's hypothesis, the effect of demand characteristics manipulations will be:

- E2.1: Zero (as predicted by Rosnow and Rosenthal, 1997), or

- E2.1: Positive and non-zero (as predicted by Coles et al. 2022)

E3: When participants do not have the opportunity to adjust their responses, the effect of demand characteristic manipulations will be:

- E3.0: Zero (as predicted by Rosnow and Rosenthal, 1997), or

- E3.1: Positive and non-zero (as predicted by Coles et al. 2022)

Data collection procedures

Our literature search strategy was developed in consultation with a librarian at Stanford University. To identify relevant articles, we searched APA PsycInfo on January 12, 2022 using relatively broad search terms: "demand characteristics" OR "hypothesis awareness". This yielded 850 records.

We also released calls for unpublished studies on the Society for Personality and Social Psychology Open Forum; Nicholas Coles' personal Twitter; and the Facebook Psychological Methods Discussion Group and PsychMAP group. This yielded 3 additional records.

Inclusion criteria

To be included in the meta-analysis, records must have met the following inclusion criteria:

- The record documented primary research (as opposed to a review of other findings)

- The researcher manipulated what participants were told about the hypothesized relationship between an independent and dependent variable. (See below for minor exceptions.)

- Participants' responses on the dependent variable described (or strongly implied) in the demand characteristics manipulation were measured.

- Information necessary for computing effect sizes was included.

In a few scenarios, we excluded observations where the researcher's manipulation of demand characteristics was confounded. For example, Sigall, Aronson, and Van Hoose (1970) told one group of participants that they (a) expected an increase in number copying behavior, but (b) this increase would be indicative of a personality disorder. This condition intentionally confounds hypothesis awareness with evaluation apprehension and was thus excluded.

Explanation of existing data and sample size

Thus far, we have coded 232 effect sizes from 34 studies. However, this number will likely change as we double-check coding decisions, refine inclusion criteria, and discover additional relevant records. (E.g., librarians are searching for several records that we indicated as potentially relevant during the record screening stage, but had difficulty accessing; Reviewers may also recommend additional searches.)

To pilot data processing and analysis procedures, a small subset of the data were analyzed at an earlier stage in the project. At this stage, we found preliminary evidence of a medium-sized, highly-heterogeneous, positive effect of demand characteristics. ~~However, we have not yet examined the full dataset or tested moderators or hypotheses-of-interest.~~ A full dataset of effect sizes were analyzed before making analysis plan amendments denoted in green text. (We note, though, the implications of those changes.)

Measured variables

Effect Size: Cohen's standardized *d*

For studies with between-subject designs and continuous measures:

- *M* and *SD* reported
    - Formula: Cooper, Hedges, & Valentine, 2009; p. 226.
- *t*-values reported
    - Formula: Cooper, Hedges, & Valentine, 2009; p. 228
- *F*-values reported
    - Formula: Cooper, Hedges, & Valentine, 2009; p. 228
- *p*-values reported
    - Formula: Cooper, Hedges, & Valentine, 2009; p. 228

For studies with between-subject designs and categorical measures:

- Formulas:
    - Borenstein et al. 2011; p. 36; Equation 5.8
    - Borenstein et al. 2011; p. 36; Equation 5.9
    - Borenstein et al. 2011; p. 47; Equation 7.1

For studies with within-subject designs:

- *M* and *SD* reported
    - Formulas:
        - Cooper, Hedges, & Valentine, 2009; p. 229
        - http://handbook.cochrane.org/chapter_16/16_4_6_1_mean_differences.htm
- *t*-values reported
    - Formula: Cooper, Hedges, & Valentine, 2009, p. 229
- *F*-values reported:
    - Formula: Cooper, Hedges, & Valentine, 2009; p. 229

Study Characteristic Moderators

- Whether participants are students (yes, no)

- Whether participants are paid (yes, no)

- Whether the study was conducted online (yes, no)

- Type of demand characteristic comparison

  - Positive demand (participants told about hypothesized effect that amplifies relationship between IV and DV) vs. control (participants not told a hypothesis)

  - Positive demand vs. nil demand (participant told that the researcher does *not* expect a relationship between IV and DV)

  - Positive demand vs. negative demand (participants told about hypothesized effect that reverses relationship between IV and DV)

  - Negative demand vs. control

  - Negative demand vs. nil demand

  - Control vs. nil demand

- Whether one or two demand characteristic conditions are being compared (one, two)

- Comparison design (between-subjects, within-subjects)

- Whether the study is published (yes, no)

Conceptual Moderators and Participant Predictions

Depending on the availability of resources, we plan to collect *post-hoc* measures of motivation, opportunity, belief, and predictions through a new set of participants (100-250). ~~Here, we describe our *preliminary plans*.~~

For each demand characteristics condition in each study, this new set of participants will review key study details and report the extent to which they:

1. Would be motivated to provide hypothesis-consistent responses (-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),
2. Would be able to adjust their responses on the outcome-of-interest (0 = "extremely incapable" to 4 = "extremely capable)
3. Would believe the experimenter's hypothesis (-3 ="strong disbelief" to 3 = "strong belief")

4. Expect other participants to confirm the researcher's hypothesis (-3 = "extremely likely to adjust responses to be inconsistent" to 3 = "extremely likely to adjust responses to be consistent").

Participants will be randomly assigned to review 10 of these vignettes. After each vignette, they will be asked to identify the researcher's stated hypothesis–and we plan to exclude ratings in cases where the correct hypothesis is not identified.

For each condition, scores will be computed by averaging participant ratings. For the control conditions, we will assume these values are zero because participants were not given information about the experimenter's hypothesis. For each effect size, we will sum the scores for the two conditions being compared. Doing so allows us to accommodate the fact that some studies do not use a control group, instead comparing two demand characteristics conditions.

Of note, we will not include nil-hypothesis comparisons in our analyses because our coding strategy cannnot 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 will 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.

Meta-regression with robust variance estimates

For overall effect size estimates and subgroup analyses (i.e., analyses split by potential moderators), we plan to use random three-level meta-analysis ~~meta-regression with robust variance estimates (Hedges, Tipton, & Johnson, 2010)~~. 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.

To calculate Cohen's *d* for within-subjects design we need the correlation between the pre- and post- measures. This correlation is rarely reported, so we will assume a correlation of *r* = .50. Nonetheless, for our *overall effect size analysis,* we will perform sensitivity analyses that test the following values: *r* = .10, .30, .70, .90.

Subgroup analyses

Publication bias

To estimate, correct, and evaluate the impact of publication bias, we will use three approaches: (1) PET-PEESE analyses (Stanley and Doucouliagos, 2013), (2) three-parameter selection modeling (Vevea & Hedges, 1995), and (3) the Mathur and VanderWeele (2020) sensitivity analyses.

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. To handle dependencies for the three-parameter selection modeling, we used aggregated dependent effect sizes (using the Borenstein, 2009 approach). The default assumed correlation between dependent effect sizes in the Borenstein (2009) aggregation method is .50. We will use this default value to inform our conclusion, but re-run the analyses with the following values: *r* = .10, .30, .70, .90.

Inference criteria

p < .05