

## **Empirical tests of political bias in social psychological research**

**Authors: Yoel Inbar, Orly Eitan, Stefan Thau, Eric Uhlmann**

### **A. Hypotheses**

#### *Primary hypotheses*

H1. Overall, abstracts will be rated as describing conservatives more negatively than liberals (evaluative bias).

H2. Overall, abstracts will be rated as explaining conservatives more than liberals (explanatory bias).

#### *Secondary hypotheses*

H3. Politically liberal raters will perceive less liberal bias in research than will be perceived by conservative raters.

H4. Evaluative and explanatory bias against conservatives will be positively correlated at the article level.

#### *A priori exploratory analyses*

Exploratory test 1: Will the degree of political bias in research decrease over time, increase, or stay the same?

Exploratory test 2: How much of the variance in ratings is explained by article versus the partisan perceptions of the raters?

### **B. Methods**

#### **Design**

We collected 846 talk and poster abstracts by searching SPSP conference programs (2003-2013) for abstracts containing politics-relevant keywords. We then asked a previous group of participants recruited on MTurk to classify each abstract as relevant to politics or not. Abstracts classified as politics-relevant by more than 60% of raters were retained for this study, yielding 306 abstracts for the current study.

Participants will be randomly assigned to rate 20 randomly-selected abstracts either for explanatory bias ("To what extent does this research attempt to explain political liberalism [conservatism] or liberal [conservative] ideas?"; 1 = "Not at All," 7 = "A Great Deal"), OR for evaluative bias ("How does this research characterize political liberals [conservatives] or liberal [conservative] ideas?"; 1 = "Extremely Negatively," 4 = "Neutral," 7 = "Extremely Positively"). We will collect separate ratings of bias for/against liberals and conservatives for each abstract; the order of the questions (liberals first vs. conservatives first) will be counterbalanced between subjects. Order and rated bias type (explanatory vs. evaluative) will be the only manipulated between-subjects variables.

Thus each rater will produce 2 ratings per abstract x 20 abstracts, making 40 ratings in all per rater. We will also collect rater demographics: gender, age, income, education, race/ethnicity, and political orientation (social, economic, overall, and partisan affiliation). Finally, we will collect English proficiency (native speaker or if not native speaker, at what age English was learned), and ask participants to report whether they read the abstracts carefully (stating that there is no penalty for answering “no.”).

### **Planned sample**

This experiment will be run on Amazon’s Mechanical Turk. We will restrict the recruitment to a) people in the U.S.; b) with a HIT approval rate of 95% or higher; c) who have completed at least 100 HITs.

Each abstract will be rated by 160 workers, who will each rate 20 abstracts (with the exception of additional one group of 160 who will rate only 6, because we have 306 abstracts total). Thus, we will recruit  $160 * (300/20) = 2,400 + 160 = 2,560$  participants.

Participants will be excluded if any of the following are true:

1. They indicate in response to a yes/no question at the end of the survey that they did not read the abstracts carefully.
2. They indicate that they learned to speak English less than 10 years ago

A power analysis conducted with the PANGAEA (Power Analysis for General Anova Designs; Westfall, 2015) web app shows power of over .99 to detect an effect size of  $d = .2$  for the two primary hypothesis (H1 and H2).

## **C. Analysis plan**

### **Confirmatory analyses**

Two tailed tests will be used throughout. We test H1-H3 using multilevel regression models. For illustration, R code (lme4) will be given, although analyses may be carried out in another program.

For H1-H3, data will be in long format with the target (liberals vs. conservative) dummy-coded. For most analyses, the focal test will be of this dummy or interactions involving it.

#### *H1: Evaluative bias*

We will fit a mixed model with evaluative bias as the DV, a target fixed effect, and random effects of rater and abstract:

```
eval ~ target + (1|rater) + (1|abstract)
```

The target dummy fixed effect (are conservatives described more negatively than liberals) will test H1.

#### *H2: Explanatory bias*

We will fit a mixed model with explanatory bias as the DV, a target fixed effect, and random effects of rater and abstract:

```
explan ~ target + (1|rater) + (1|abstract)
```

The target dummy fixed effect (does research explain conservatives more than liberals) will test H2.

The unit of effect size for assessing the amount of evaluative and explanatory bias in the abstracts (H1 and H2) will be Cohen's  $f^2$ , computed using the formula:

$$f^2 = \frac{R^2}{1 - R^2}$$

By convention, a small effect would be  $f^2 = .02$ , a medium effect  $f^2 = .15$ , and a large effect  $f^2 = .35$ .

### *H3: The effect of rater politics on ratings of bias*

For both type of bias, we will fit a mixed model that includes rater politics and the interaction of politics and target as fixed effects. This interaction (does the perception that conservatives are explained more, or more negatively, depend on rater politics?) will be the test of H3. These analyses will exclude participants who do not rate themselves on political orientation.

We will use self-rated social liberalism/conservatism as our primary measure of politics, but we will also conduct robustness analyses using a) overall self-rated liberalism/conservatism; b) a composite of social liberalism/conservatism, economic liberalism/conservatism, and overall liberalism/conservatism; c) political party identification. Robustness analyses a) and b) will exclude any participants who chose "don't know/not political" "libertarian" or "other" on the overall liberalism/conservatism question; robustness analysis c) will exclude any participants who chose "I have no party preference" or "I support another political party."

The models are specified as:

```
eval ~ target + poli.social + target*poli.social + (1|rater) +  
(1|abstract)  
explan ~ target + poli.social + target*poli.social + (1|rater)  
+ (1|abstract)
```

### *H4: Are evaluative and explanatory bias correlated at the abstract level?*

Because separate groups will rate evaluative and explanatory bias for each abstract, we will average these ratings for each abstract, producing one (average) rating of evaluative bias and one of explanatory bias for each abstract. We will then correlate these two sets of ratings across the 306 abstracts. A significant positive correlation coefficient will be support for the hypothesis.

## Exploratory Analyses

*Exploratory H1: Will the degree of political bias in research decrease over time, increase, or stay the same?*

To test this question we will add fixed effects for abstract year. The target\*year interaction will test the hypothesis:

```
eval ~ target + year + year*target + (1|rater) + (1|abstract)
explan ~ target + year + year*target + (1|rater) +
(1|abstract)
```

*Exploratory H2: How much of the variance in ratings is explained by abstract, and how much is explained by rater ideology?*

In order to answer this question we will compare variance explained ( $R^2$ ) for three models: a) a baseline model without a random term for abstract; b) with a random term for abstract but no term for rater ideology; c) a random term for abstract and fixed terms for rater ideology and the ideology x target interaction. The change in variance explained between a) and b) is the variance explained by abstract; the change between b) and c) is the variance explained by rater ideology.

Model A: `eval ~ target + (1|rater)`

Model B: `eval ~ target + (1|rater) + (1|abstract)`

Model C: `eval ~ target + poli.social + target*poli.social + (1|rater) + (1|abstract)`

## Final Questions

We attest that no data have been collected for this project.