Olfactory

David Kane

1/26/2019

Data Introduction

There are two id variables: `ideval_n` (which I rename to `e_id`) identifies the evaluators and `type_n` (which I rename as `t_id`) identifies the targets. In the R data frame which I grabbed from the Dataverse, they come in as ordered factors. That is clearly wrong, but I doubt that this was an issue in the published paper.

The data set includes 2499 rows. There were 21 "target" individuals whose smell was evaluated by 119 evaluators. The 2499 rows is a result of 21 times 119. However, it *looks* like the target individuals *also* served as evaluators. I don't see this mentioned in the paper, but what else would explain why some of the values for `e_id` are also present in `t_id`? Without *any* codebook with clear variable explanations, it is tough to know.

The 304 missing values for the response variable (a measure of attractiveness) are a bit of a mystery. The paper discusses dropping one target because he wore the pads for two days instead of one, but my sense is that none of his data should be here. The missing values for `attractive` are spread not-exactly-evenly across all 21 targets. This number of missing is consistent with the regression results, which only show 2,195 observations, but I am still curious about the reason for the missing values. A proper forensic examination would look more closely at this.

Targets

We can replicate the SI-1 table about targets perfectly. Ought to explore the `gt` package in order to make these look nicer.

Evaluators

We are one off when trying to replicate the data for evaluators. We have one extra male Conservative.

```
##
##          Conservative Liberal Sum
##   Female           17      49  66
##   Male             27      26  53
##   Sum              44      75 119
```

My *guess* is that this is a coding mistake in which the original paper drops an evaluator with a `politicalIdeo` value of 4, instead of classifying him as a Conservative. I doubt it matters.

UPDATE: I contacted Dustin Tingley. He confirms that the published result is mistaken and that my numbers are correct.

Figure 1

Here are some rough versions of the component parts of Figure 1.

Looking by eyeball, these seem to match the published plots. Note the single evaluator with a `politicalIdeo` value of 4. I think that this individual is mistakenly dropped from Table SI-1 but is present in these figures.

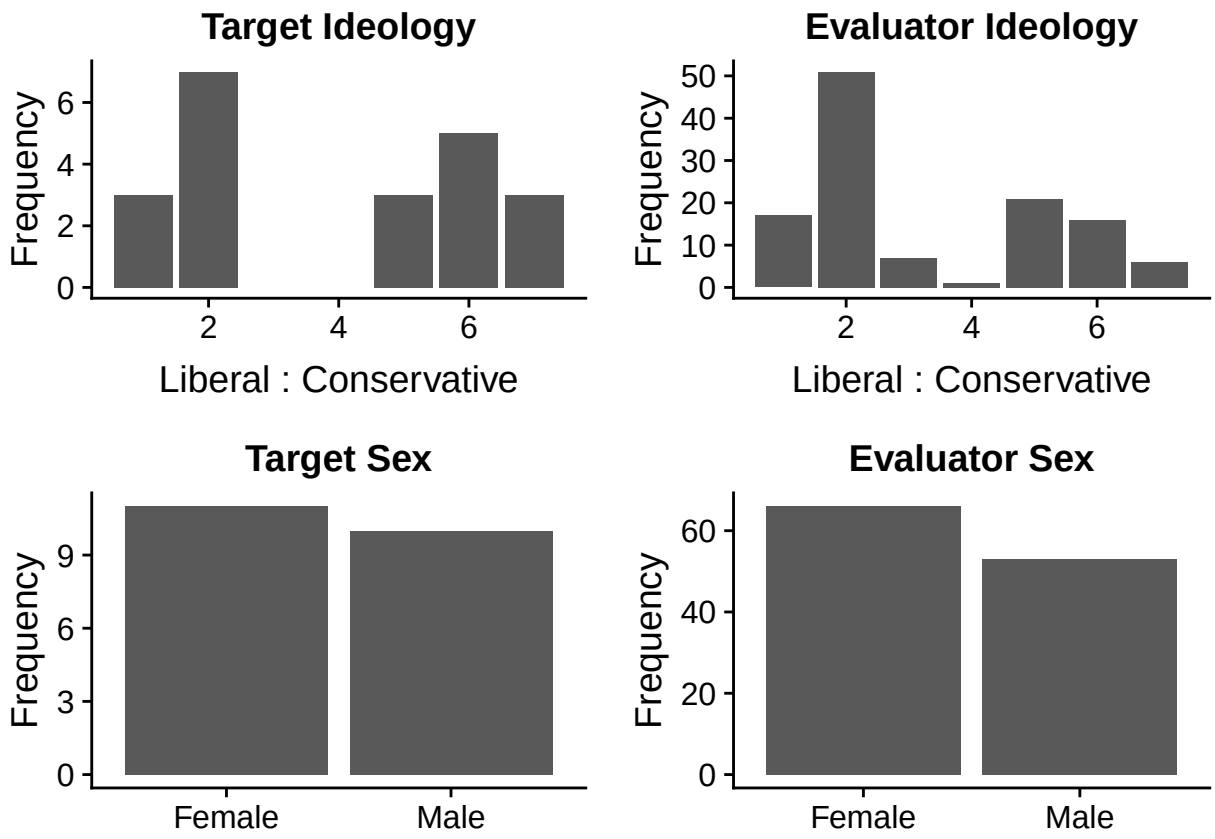

Figure 1: *Note*: The sexual categories are binary, and the ideological ones exist on a 7-point scale ranging from *very liberal* (1) to *very conservative* (7).

The `cowplot` package would be helpful in grouping these plots together. We might also mess around with the x-axis tick labels to see all 7 values. Not sure what the point is of making the gender bars more narrow.

Table 1

For now, I don't bother with clustered standard errors.

```
mod_1 <- lm(data = x, attractive ~ same_ideology + same_sex +
             e_ideology + t_ideology +
             e_gender   + t_gender +
             e_attractiveness + t_attractiveness)


mod_2 <- lm(data = x, attractive ~ neg_abs_ideo_diff + same_sex +
              e_ideo_scale + t_ideo_scale +
              e_gender + t_gender +
              e_attractiveness + t_attractiveness)

mod_3 <- lm(data = x, attractive ~ neg_abs_ideo_diff + same_sex + t_id + e_id)
```

Note how closely the key result, the coefficient and standard error for `neg_abs_ideo_diff` matches the published version in Model 3.

Table 1: Odor Attraction as a Function of Ideological Similarity

	attractive		
	Model 1	**Model 2**	**Model 3**
Same Ideology	0.085		
	(0.052)		
-Abs. Ideology Diff.		0.021	0.023
		(0.014)	(0.015)
Same Sex	-0.144^{***}	-0.143^{***}	-0.143^{***}
	(0.051)	(0.051)	(0.052)
Conservative Eval.	0.006		
	(0.054)		
Conservative Target	-0.020		
	(0.052)		
Ideology of Eval.		-0.001	
		(0.014)	
Ideology of Target		0.006	
		(0.012)	
Male Evaluator	-0.00003	0.0004	
	(0.052)	(0.052)	
Male Target	-0.017	-0.014	
	(0.053)	(0.053)	
Avg. Target Attract	0.999^{***}	0.999^{***}	
	(0.046)	(0.046)	
Avg Eval. Attract	0.999^{***}	1.001^{***}	
	(0.040)	(0.041)	
N	2195	2195	2195
R-squared	0.343	0.342	0.343
Adj. R-squared	0.340	0.340	0.302
Residual Std. Error	1.177 (df = 2186)	1.177 (df = 2186)	1.210 (df = 2067)
F Statistic	142.424^{***} (df = 8; 2186)	142.316^{***} (df = 8; 2186)	8.485^{***} (df = 127; 2067)

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$