In the following, we show how we conducted our instrumental variable regressions, including annotated outputs (the R code is in blue).

Pre-Study

To test the strength of the instrument, we use the ivmodel package (Kang et al., 2020). The section "First Stage Regression Result" shows the *F* statistic of the excluded instrument (condition). → see the yellow frame

```
> #------------------
> #Strength of the instrument
> library(ivmodel)
> # without controls
> iv1.fit = ivmodelFormula(performance ~ SL |
+                            condition, data=data_complete,
+                          heteroSE = TRUE)
>
> summary(iv1.fit)

Call:
ivmodel(Y = Y, D = D, Z = Z, intercept = intercept, beta0 = beta0,
    alpha = alpha, k = k, manyweakSE = manyweakSE, heteroSE = heteroSE,
    clusterID = clusterID, deltarange = deltarange, na.action = na.action)
sample size: 166
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

First Stage Regression Result:

F=16.14579, df1=1, df2=164, p-value is 8.9175e-05
R-squared=0.08962624,   Adjusted R-squared=0.08407518
Residual standard error: 0.8499023 on 165 degrees of freedom
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Coefficients of k-Class Estimators:

             k Estimate Std. Error t value Pr(>|t|)
OLS     0.0000  -4.4039     2.9742  -1.481    0.141
Fuller  0.9939  -6.0254     9.4362  -0.639    0.524
TSLS    1.0000  -6.1364    10.0507  -0.611    0.542
LIML    1.0000  -6.1364    10.0507  -0.611    0.542
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Alternative tests for the treatment effect under H_0: beta=0.

Anderson-Rubin test (under F distribution):
F=0.3651036, df1=1, df2=164, p-value is 0.54652
95 percent confidence interval:
 [-29.6752157093531, 16.1908340589841]

Conditional Likelihood Ratio test (under Normal approximation):
Test Stat=0.3651036, p-value is 0.54652
95 percent confidence interval:
 [-29.6753153974172, 16.1909237483481]
```

The *F*-value is significant, $F(1, 164) = 16.15$ ($p = .000$). To indicate the relevance of our instrument, it has to be compared to the critical value by Stock and Yogo (2005) and should be at least larger than the critical value of 8.96 (allowing for a 15% bias). Our value is above the critical value of 8.96 and even close to the strictest critical value of 16.38 (allowing for a 10% bias).

Additionally, control variables can be included. They must be added in the first- and second-stage regression:

```
> #------------
> # with controls
> iv2.fit = ivmodelFormula(performance ~ SL + AGRE + sex + age + edu_dummy3 + edu_du
mmy4 + edu_dummy5|
+                          condition + AGRE + sex + age + edu_dummy3 + edu_dummy4
 + edu_dummy5, data=data_complete,
+                        heteroSE = TRUE)
>
> summary(iv2.fit)

Call:
ivmodel(Y = Y, D = D, Z = Z, X = X, intercept = intercept, beta0 = beta0,
    alpha = alpha, k = k, manyweakSE = manyweakSE, heteroSE = heteroSE,
    clusterID = clusterID, deltarange = deltarange, na.action = na.action)
sample size: 166
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

First Stage Regression Result:

F=20.18639, df1=1, df2=158, p-value is 1.3506e-05
R-squared=0.1132881,   Adjusted R-squared=0.107676
Residual standard error: 0.8349928 on 159 degrees of freedom
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Coefficients of k-Class Estimators:

             k Estimate Std. Error t value Pr(>|t|)
OLS     0.0000  -5.0943     2.8989  -1.757   0.0808 .
Fuller  0.9937  -4.0126     8.2684  -0.485   0.6281
LIML    1.0000  -3.9518     8.7197  -0.453   0.6510
TSLS    1.0000  -3.9518     8.7197  -0.453   0.6510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Alternative tests for the treatment effect under H_0: beta=0.

Anderson-Rubin test (under F distribution):
F=0.1901577, df1=1, df2=158, p-value is 0.66338
95 percent confidence interval:
 [-23.4175538851349, 16.1312228593601]

Conditional Likelihood Ratio test (under Normal approximation):
Test Stat=0.1901577, p-value is 0.66338
95 percent confidence interval:
 [-23.417470342871, 16.1311339346413]
```

In our case, including the control variables leads to an even stronger even stronger relevance of our instrument, $F(1, 158) = 20.19$ ($p = .000$), exceeding even the strictest critical value of 16.38 by Stock and Yogo (2005). Therefore, we argue that our manipulation should be an appropriate instrument.

With the ivmodel package, we can also calculate the appropriate sample size for our main experiment. We use the IVsize function and conduct it for the 2SLS approach with and without control variables, a power of .90, and assuming a true effect on performance of half a standard deviation (coef. = 17.31; see Table S3.3 for the standard deviation of performance). The resulting minimal sample size was $N = 589$ (without controls; with controls, the required sample size was smaller: $N = 465$):

```
> #Power
>
> #beta 0.5 SD
> IVsize(iv1.fit, power = .9, beta = 17.31, type = "TSLS")
[1] 589
> IVsize(iv2.fit, power = .9, beta = 17.31, type = "TSLS")
[1] 465
```

Main Experiment

For the main experiment, we also use the Applied Econometrics with R (AER) package (Kleiber & Zeileis, 2008). AER is based on the same formulas as ivmodel.

AER:

First, we enter our instrumental variable (i.e., the leadership manipulation) and the control variables (i.e., agreeableness, gender, age, and education) in the first-stage regression to predict the perceptions of combined stewardship and authenticity.

```
> #------------------
> # Instrumental variable regression
> #------------------
> library(AER)
>
> #First-stage IV regression
> step1 <- lm(SL ~ condition + AGRE + sex_f + sex_d + age  + edu_dummy3 + edu_dummy4
 + edu_dummy5 + edu_dummy6, data=data_complete)
> summary(step1)

Call:
lm(formula = SL ~ condition + AGRE + sex_f + sex_d + age + edu_dummy3 +
    edu_dummy4 + edu_dummy5 + edu_dummy6, data = data_complete)

Residuals:
     Min       1Q   Median       3Q      Max
-3.12330 -0.52439  0.05016  0.58260  1.89249

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.288353   0.252810  13.007  < 2e-16 ***
condition    0.525775   0.067724   7.764 3.71e-14 ***
AGRE         0.294233   0.043204   6.810 2.42e-11 ***
sex_f       -0.099234   0.070928  -1.399 0.162320
sex_d       -0.206284   0.481540  -0.428 0.668529
age         -0.002812   0.002971  -0.946 0.344314
edu_dummy3  -0.081299   0.100771  -0.807 0.420126
edu_dummy4  -0.312742   0.096219  -3.250 0.001219 **
edu_dummy5  -0.438860   0.117722  -3.728 0.000212 ***
edu_dummy6  -0.455179   0.188345  -2.417 0.015966 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8214 on 585 degrees of freedom
Multiple R-squared:  0.1874,    Adjusted R-squared:  0.1749
F-statistic: 14.99 on 9 and 585 DF,  p-value: < 2.2e-16
```

The estimated effect of the manipulation on combined stewardship and authenticity perceptions is coef. = 0.526, *SE* = 0.068, *p* = .000 → see the yellow frame.

The *F*-statistic for the excluded instrument has to be calculated in a separate step to be completely consistent with the ivmodel package:

```
> #F-test by hand (Cragg-Donald Wald F statistic)
> Ftest <- waldtest(step1,.~.-condition)
> print(Ftest)
Wald test

Model 1: SL ~ condition + AGRE + sex_f + sex_d + age + edu_dummy3 + edu_dummy4 +
    edu_dummy5 + edu_dummy6
Model 2: SL ~ AGRE + sex_f + sex_d + age + edu_dummy3 + edu_dummy4 + edu_dummy5 +
    edu_dummy6
  Res.Df Df      F     Pr(>F)
1    585
2    586 -1 60.273 3.709e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

With *F*(1, 585) = 60.27, the Wald test shows that the first-stage *F*-statistic of the excluded instrument exceeds the strictest critical value of 16.38 (allowing for a 10% bias) by J. H. Stock and Yogo (2005), meaning that the relevance condition is fulfilled.

We then proceed with the second-stage regression to test Hypothesis 2 and regress performance on the instrumented perceptions and control variables.

```
> #Second-stage IV regression
> step2 <- ivreg(performance ~ SL + AGRE + sex_f + sex_d + age + edu_dummy3 + edu_du
mmy4 + edu_dummy5 + edu_dummy6
+                  |condition + AGRE + sex_f + sex_d + age + edu_dummy3 + edu_dummy
4 + edu_dummy5 + edu_dummy6, data=data_complete)
> summary(step2, vcov=sandwich, diagnostics=TRUE)

Call:
ivreg(formula = performance ~ SL + AGRE + sex_f + sex_d + age +
    edu_dummy3 + edu_dummy4 + edu_dummy5 + edu_dummy6 | condition +
    AGRE + sex_f + sex_d + age + edu_dummy3 + edu_dummy4 + edu_dummy5 +
    edu_dummy6, data = data_complete)

Residuals:
     Min       1Q   Median       3Q      Max
-175.398  -40.019    3.425   44.156  203.300

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 207.6274    41.3929   5.016 7.01e-07 ***
SL           -5.6447    10.5050  -0.537  0.59124
AGRE         -1.7260     4.9852  -0.346  0.72930
sex_f         8.8535     5.9063   1.499  0.13441
sex_d        42.0995    60.2125   0.699  0.48472
age          -1.2627     0.2392  -5.278 1.84e-07 ***
edu_dummy3    8.4442     7.6310   1.107  0.26894
edu_dummy4   18.5459     8.4455   2.196  0.02849 *
edu_dummy5   13.5734    11.0759   1.225  0.22089
edu_dummy6   43.7451    16.2827   2.687  0.00742 **

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1 585    60.329 3.61e-14 ***
Wu-Hausman         1 584     0.068    0.795
Sargan             0  NA        NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 67.73 on 585 degrees of freedom
Multiple R-Squared: 0.08105,    Adjusted R-squared: 0.06692
Wald test: 6.412 on 9 and 585 DF,  p-value: 1.037e-08
```

The results do not support Hypothesis 2 (coef. = −5.645, *SE* = 10.505, *p* = .591; see yellow frame).

To examine whether the combined stewardship and authenticity perceptions are endogenous, we compare the coefficient from the instrumental variable estimation to that obtained via OLS regression (coef. = −3.057, *SE* = 3.527, *p* = .386; see Table 10 for the OLS results). The absolute value obtained by instrumental variable regression is substantially larger (1.847) than that estimated by OLS regression. The difference between the effect sizes suggests that the effects estimated by OLS regression could have led to biased results.

To test the bias statistically, we perform a Wu-Hausman test for endogeneity (Hausman, 1978; D.-M. Wu, 1973) → see the green frame, we request it with “diagnostics = TRUE” in the summary function. The endogeneity test is not significant, with *F*(1, 584) = 0.068, *p* = 0.795, indicating that the combined stewardship and authenticity perceptions are not biased by endogeneity.

ivmodel:

If we conduct the analyses with the ivmodel package, we obtain the same results (see the yellow frames):

```
> #------------
> # Alternative calculation (instrumental variable regression)
> library(ivmodel)
> #Strength of the instrument
> iv.fit = ivmodelFormula(performance ~ SL + AGRE + sex_f + sex_d + age + edu_dummy3
 + edu_dummy4 + edu_dummy5 + edu_dummy6 |
+                          condition + AGRE + sex_f + sex_d + age + edu_dummy3 + ed
u_dummy4 + edu_dummy5 + edu_dummy6 , data=data_complete,
+                        heteroSE = TRUE)
> summary(iv.fit)

Call:
ivmodel(Y = Y, D = D, Z = Z, X = X, intercept = intercept, beta0 = beta0,
    alpha = alpha, k = k, manyweakSE = manyweakSE, heteroSE = heteroSE,
    clusterID = clusterID, deltarange = deltarange, na.action = na.action)
sample size: 595
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

First Stage Regression Result:

F=60.27278, df1=1, df2=585, p-value is 3.7081e-14
R-squared=0.09340667,   Adjusted R-squared=0.09185694
Residual standard error: 0.8213644 on 586 degrees of freedom
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Coefficients of k-Class Estimators:

             k Estimate Std. Error t value Pr(>|t|)
OLS     0.0000  -3.0571     3.4507  -0.886    0.376
Fuller  0.9983  -5.5981    10.3359  -0.542    0.588
TSLS    1.0000  -5.6447    10.5050  -0.537    0.591
LIML    1.0000  -5.6447    10.5050  -0.537    0.591
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Alternative tests for the treatment effect under H_0: beta=0.

Anderson-Rubin test (under F distribution):
F=0.2824328, df1=1, df2=585, p-value is 0.59531
95 percent confidence interval:
 [-27.4033575324458, 15.7237036154825]

Conditional Likelihood Ratio test (under Normal approximation):
Test Stat=0.2824328, p-value is 0.59531
95 percent confidence interval:
 [-27.4033499307018, 15.7236962840138]
```

We can request not only the coefficients of the instrument perceptions but also of the other variables in the model:

```
> coefother(iv.fit) #all coefficients
$OLS
             Estimate Std. Error    t value     Pr(>|t|)
AGRE        -2.494881  3.8146748 -0.6540219 5.133547e-01
sex_f        9.048219  5.8573972  1.5447508 1.229473e-01
sex_d       41.949124 59.9747636  0.6994463 4.845512e-01
age         -1.253530  0.2376984 -5.2736162 1.885512e-07
edu_dummy3   8.723423  7.5200519  1.1600216 2.465132e-01
edu_dummy4  19.420991  7.7083667  2.5194690 1.201846e-02
edu_dummy5  14.802954  9.9947783  1.4810688 1.391266e-01
edu_dummy6  45.062525 15.5900629  2.8904646 3.989455e-03
intercept  198.360442 23.5154538  8.4353227 2.220446e-16

$TSLS
             Estimate Std. Error    t value     Pr(>|t|)
AGRE        -1.726007  4.9852378 -0.3462237 7.292991e-01
sex_f        8.853484  5.9062613  1.4989997 1.344131e-01
sex_d       42.099530 60.2124780  0.6991828 4.847157e-01
age         -1.262705  0.2392177 -5.2784773 1.838397e-07
edu_dummy3   8.444183  7.6309867  1.1065650 2.689369e-01
edu_dummy4  18.545862  8.4455282  2.1959387 2.848815e-02
edu_dummy5  13.573379 11.0759492  1.2254823 2.208867e-01
edu_dummy6  43.745115 16.2827311  2.6865956 7.423576e-03
intercept  207.627448 41.3928753  5.0160190 7.006147e-07

$LIML
             Estimate Std. Error    t value     Pr(>|t|)
AGRE        -1.726007  4.9852378 -0.3462237 7.292991e-01
sex_f        8.853484  5.9062613  1.4989997 1.344131e-01
sex_d       42.099530 60.2124780  0.6991828 4.847157e-01
age         -1.262705  0.2392177 -5.2784773 1.838397e-07
edu_dummy3   8.444183  7.6309867  1.1065650 2.689369e-01
edu_dummy4  18.545862  8.4455282  2.1959387 2.848815e-02
edu_dummy5  13.573379 11.0759492  1.2254823 2.208867e-01
edu_dummy6  43.745115 16.2827311  2.6865956 7.423576e-03
intercept  207.627448 41.3928753  5.0160190 7.006147e-07

$Fuller
             Estimate Std. Error    t value     Pr(>|t|)
AGRE        -1.739849  4.9508656 -0.3514231 7.253975e-01
sex_f        8.856989  5.9044552  1.5000519 1.341404e-01
sex_d       42.096822 60.2081387  0.6991882 4.847123e-01
age         -1.262540  0.2391436 -5.2794213 1.829380e-07
edu_dummy3   8.449209  7.6276025  1.1077150 2.684402e-01
edu_dummy4  18.561616  8.4207371  2.2042745 2.789365e-02
edu_dummy5  13.595514 11.0393992  1.2315447 2.186140e-01
edu_dummy6  43.768831 16.2563105  2.6924209 7.296719e-03
intercept  207.460624 40.8665077  5.0765440 5.172746e-07
```