Economics Bulletin

Volume 31, Issue 4

# Good news for experimenters: Subjects are hard to influence by instructors′ cues

Ivo Bischoff
*University of Kassel*

Björn Frank
*University of Kassel*

## Abstract

An important concern of experimenters is that instructors′ nonverbal cues might change subject behavior. We let a professional actor try to produce this bias on purpose, finding only weak evidence for an "instructor demand effect", and only for female subjects.

We would like to thank actor Peter Cieslinski and cameraman Martin Neumeyer, who did a great job. Furthermore, we are indebted to (in alphabetical order) Christian Bergholz, Michael Betz, Frédéric Blaeschke, Rasums Bode, Lara Bühren, Christoph Bühren, Heike Krönung, Sha Li, Simon Peschges, Marco Pleßner and Michel Schrader for support during the experiment. We also thank Julia Bischoff-Weiss, Adriana Kramer and Beverley Locke for helpful hints and comments.
**Citation:** Ivo Bischoff and Björn Frank, (2011) "Good news for experimenters: Subjects are hard to influence by instructors′ cues", *Economics Bulletin*, Vol. 31 No. 4 pp. 3221-3225.
**Contact:** Ivo Bischoff - bischoff@wirtschaft.uni-kassel.de, Björn Frank - frank@uni-kassel.de.
**Submitted:** October 29, 2011. **Published:** November 23, 2011.

"Today's artifact may be tomorrow's independent variable."
William J. McGuire (1969, p.13)

## 1. Introduction

Given that experimental environments are always artificial to a certain degree, subjects are generally uncertain as to what behavior is appropriate and are thus susceptive to any cue that helps. Consequently even small, seemingly innocent changes in instructions or in the environment in which the experiment takes place can lead to significant change in subjects' behavior. This raises the concern that instructors might spoil the experiment even though they typically have no influence on the wording of the instructions that they read to the subjects. Ortmann (2005) argues that the instructors' expectations concerning the outcome of the experiment might unconsciously affect their accent, tone of voice and gestures when presenting the experiment. If interpreted as cues for appropriate behavior, this may drive subjects' behavior and mitigate the experiment's external and internal validity. Unfortunately, "…effects (such as reactions to idiosyncratic facial expressions, or expressive movements), … are difficult to control." (Ortmann, 2005, p.57).

We call these effects on subjects' behavior "social instructor demand effect" (SIDE). The SIDE's closest relative is the instructor demand effect (Zizzo, 2010), with "social experimenter demand effects" referring to cues (such as wording of the instructions) given by *the experimenter* in his or her role as the person who makes up the design. The role of the *instructor* has so far only been studied in psychology, cf. the Rosenthal Effect (Rosenthal and Fode, 1963). Although there is a large body of (economic) literature showing that the instructor has a significant effect on student performance in learning tasks (e.g., Bosshardt and Watts, 1990), these effects have generally not been tested in experiments in which the instructors are restricted to presenting a predefined text.

In constructing a test for a SIDE, we started with the following hypothesis: Assume that an instructor can – possibly without any intention to do so – induce certain behavioral patterns among subjects by the way he presents the predefined instructions. If this is true, we can expect that a professional actor is able to evoke this SIDE when asked to do so because actors are explicitly trained to use accent, tone of voice and gestures to a certain end. If, on the other hand, a professional actor cannot induce changes in behavior, this indicates that a SIDE is not easily produced.

## 2. Experimental set-up and main hypotheses

To test for a possible SIDE, we play the solidarity game (Selten and Ockenfels, 1998). The solidarity game has a simple set-up, yet asks subjects for decisions in a field where hints about appropriate behavior are likely to have an impact on their actions. At the beginning of the experiment, every subject is randomly assigned to a three-person group, and does not know the identity of the others. Each subject has to throw a die once. If the number 1, 2, 3, or 4 appears, she wins €5 (approximately equal to 10 German marks used in the original experiment by Selten and Ockenfels, 1998). If the number 5 or 6 appears, she loses.

Before the die is thrown, every participant has to decide on the amount of money that she is willing to give to other members in her group in case she wins €5 while other group members lose. Specifically, she has to make two decisions. First: How much (if any) of her €5 will she give to a losing party, in case she wins and *exactly one* other group member loses? Second: How much of her €5 will she give to each losing member in case she is the only winner in her group? The stated amounts of money are called X1 and X2 respectively.

We asked a professional actor (for theater and film) to present the instructions for the solidarity game in two different ways. The first presentation (treatment LOW) was meant to induce a low level of solidarity, while the second presentation should induce a high level of solidarity (treatment HIGH). Both presentations were recorded on film so that we can use them repeatedly and ensure transparency in experimental methods. In the experimental sessions, the films were used. The actor was paid the customary fee. In addition, we promised him a case of high-quality apple juice (a regional delicacy) if we were to observe significant differences results from the two treatments. The actor used visual as well as vocal cues, but by no means overacted.[1]

Experimental sessions were performed on June 5, 2009, and May 26, 2011, at the University of Kassel, Germany. A total of 182 subjects were recruited during two summer festivities of the Business and Economics Faculty. Expecting the actor/instructor to be able to produce a SIDE, we hypothesize that subjects' contributions are higher in the HIGH treatment than in the LOW treatment. Furthermore, for each treatment, the actor was filmed twice: one time dressed rather formally (in shirt, jacket and tie), one time more casually (in a polo shirt). We use this treatment variable (CASUAL) as an additional device to explore the possible impact of the instructor's dress on subjects' solidarity.[2]

## 3. Results

As shown in table I, we found a weakly significant difference between the X2 across treatments LOW and HIGH in treatment "CASUAL" (t-test, p = 0.1), but no significant differences for X1. Under treatment "NON-CASUAL", differences are found neither for X1 nor for X2.[3] The amounts for X1 and X2 are not found to differ significantly between the dress treatments. In a two-way ANOVA, we find a weakly significant difference in X2 between the treatments LOW and HIGH (F-test, p = 0.1) but not between dress treatments. For X1, the ANOVA produces no differences at all.

*Table I: Average X1 and X2 by treatments and date of experiment [€]*

| | | | X1 | | X2 | |
|---|---|---|---|---|---|---|
| | Date | N | LOW | HIGH | LOW | HIGH |
| CASUAL | June 2009 | 51 | 1.09[n] | | 0.73 | |
| | | 53 | | 1.25 | | 0.92 |
| NON-CASUAL | May 2011 | 40 | 1.06[n] | | 0.79 | |
| | | 41 | | 1.17[n] | | 0.84[n] |

[n] The normality assumption is rejected (p = 0.05)

[1] No subject we talked to afterwards guessed that the experiment was about what the experimenter did. Sample films can be seen here: http://univideo.uni-kassel.de/video/714 (LOW/CASUAL); http://univideo.uni-kassel.de/video/715 (HIGH/CASUAL).

[2] We also performed the experiment in tutorials for compulsory courses in economics held by the authors of this paper. However, here we have social ties between subjects and experimenter that are untypical, that are hard to control for, but that might interfere with SIDE. Hence we felt we cannot pool the observations, and report the results in footnote 3 below.

[3] Similarly, with subjects recruited from our own courses, we did not find a significant difference between LOW and HIGH treatment, neither for X1 (€ 1.14 versus €1.21) nor for X2 (€0.83 versus 0.88).

To make sure that the results above do not represent spurious correlations, we go on to regress X1 and X2 on additional information about subjects from a post-experimental questionnaire. We have to account for the fact that 39 (36) subjects stated X1 = 0 (X2 = 0). Table II presents the results of a two-part regression (Cameron and Trivedi, 2009: 538-544)[4]. In part 1, the endogenous variables are binary: dX1 =1 if X1 >0 and dX1 = 0 if X1 = 0; mutatis mutandis the same holds for dX2. A probit approach is used to analyse the impact of our exogenous variables, especially HIGH and CASUAL, on subjects' decision to give a positive amount. In part 2, we run OLS regressions to explain the level of X1 and X2 (restricted to all cases where X1 > 0 and X2 > 0, respectively).

*Table II: Results of the two-part regression*

| Variable | dX1 | dX2 | X1 | X2 |
|---|---|---|---|---|
| | Probit | Probit | OLS | OLS |
| HIGH | -0.0918 | -0.0102 | 0.1564 | 0.0905 |
| | (0.335) | (0.3435) | (0.1657) | (0.1204) |
| CASUAL | -0.0713 | -0.0793 | 0.0759 | -0.0641 |
| | (0.2629) | (0.2707) | (0.1213) | (0.0877) |
| AGE | 0.0289 | 0.0438* | 0.0058 | 0.0068 |
| | (0.0248) | (0.0255) | (0.0115) | (0.0084) |
| STUDENT | 0.1913 | 0.3179 | -0.0755 | 0.0249 |
| | (0.3421) | (0.3405) | (0.173) | (0.1266) |
| FEMALE | -0.3395 | -0.2625 | 0.077 | -0.0569 |
| | (0.3278) | (0.3338) | (0.168) | (0.1229) |
| FEMALExHIGH | 0.8194* | 0.8413* | -0.1404 | -0.0393 |
| | (0.4883) | (0.4987) | (0.2272) | (0.1641) |
| CONSTANT | 0.4475 | -0.0683 | 1.6002*** | 0.9909*** |
| | (0.8288) | (0.8503) | (0.405) | (0.2949) |
| WVS variables[a)] | *included* | *included* | *included* | *included* |
| N | 182 | 182 | 143 | 146 |
| (Pseudo)-$R^2$ | 0.1951 | 0.2188 | 0.1136 | 0.1496 |
| $\chi^2$ | 40.36*** | 41.56*** | | |
| F statistic | | | 1.01 | 1.42 |
| Breusch-Pagan-stat. | | | 0.14 | 0.31 |

* significant at p = 0.1, *** significant at p = 0.01 (standard errors in parentheses)
[a)] Detailed results available upon request

[4] The Tobit approach proved inadequate because residuals are not normally distributed (and heteroscedastic). Moreover, it covers up the fact that our explanatory variables have substantial explanatory power for the decision whether to redistribute at all (i.e. state X1, X2 > 0) but not for the level of X1 and X2. Regressions using a Heckman approach do not support a selection bias. With respect to our treatment variables HIGH and CASUAL, both approaches yield the same results as the two-part regressions reported above. The results are available from the authors upon request.

The $\chi^2$- and F-statistics reveal that our explanatory variables have substantial explanatory power with respect to dX1 and dX2 but not for X1 and X2.
Demographic variables we control for are AGE, STUDENT (a dummy variable taking the value 0 for non-student visitors of the campus day), and FEMALE (1 for women, 0 for men). While FEMALE is never significant, the interaction variable FEMALExHIGH has a weakly positive impact on dX1 and dX2. This is in line with the existing literature suggesting that women's behaviour in experiments is more situation specific than men's, and thus possibly more sensitive to variations and cues in the (experimental) environment (Croson and Gneezy, 2009). Furthermore, we used a number of questions from the *World Values Survey* (WWS) to control for subjects' trust, beliefs and life-satisfaction. Our main result, as already shown in table I, remains intact: Our treatment variables HIGH and CASUAL are never significant (not even at a 10 percent level). The "social instructor demand effect" on subjects' behavior (SIDE) is hard to produce even on purpose.

## 4. Conclusion

We find weak evidence that female subjects prove susceptive to a social instructor demand effect (SIDE). However, this effect is too weak to drive our overall results. In sum, this experiment provides good news. Social instructor demand effects are hard to produce – even when a professional actor intends to do so. Thus, the likelihood that instructors' unintended cues mitigate internal validity in economic experiments is low.

## References


Bosshardt, William and Michael Watts (1990) "Instructor Effects and Their Determinants in Pre-College Economic Education" *Journal of Economic Education* **21**, 265-276.

Cameron, A. Colin and Pravin K. Trivedi (2009) *Microeconometrics Using Stata*, Stata Press: College Station, Texas.

Croson, Rachel and Uri Gneezy (2009) "Gender Differences in Preferences" *Journal of Economic Literature* **47**, 448-474.

McGuire, William J. (1969) "Suspiciousness of Experimenter's Intent" in *Artifact in Behavioral Research* by Robert Rosenthal and Ralph L. Rosnow, Eds., Academic Press: New York, San Francisco, London, 13-57.

Ortmann, Andreas (2005) "Field Experiments in Economics: Some Methodological Caveats" in *Field Experiments in Economics* (Research in Experimental Economics Vol. 10) by Jeffrey P. Carpenter, Glenn W. Harrison and John A. List, Eds., Bingley: Emerald, 51-70.

Rosenthal, Robert and Kermit L. Fode (1963) "The effect of experimenter bias on the performance of the albino rat" *Behavioral Science* **8**, 183-189.

Selten, Reinhard and Axel Ockenfels (1998) "An experimental solidarity game" *Journal of Economic Behavior & Organization* **34**, 517-539.

Zizzo, Daniel John (2010) "Experimenter demand effects in economic experiments" *Experimental Economics* **13**, 75-98.