Measuring Experimenter Demand in an Effort Experiment: Pre-analysis Plan[*]

Jonathan de Quidt[†]   Johannes Haushofer[‡]   Chris Roth[§]

August 26, 2016

1   Introduction

Experimenter demand effects pose an important challenge to understand and interpret results from laboratory and field experiments (Rosenthal, 1966; Zizzo, 2010). For example, experimenter demand effects are an important obstacle to interpreting effect sizes and they constitute a threat to the external validity of experiments. Experimenter demand effects also affect the optimal design of experiments; however, little or no empirical evidence exists that studies how design features affect the relevance of experimenter demand effects. Moreover, experimental economists have a very limited understanding of the relative importance of experimenter demand effects for different outcome measures or experimental paradigms.

In this experiment we manipulate subjects' beliefs about the intentions and desires of the experimenters. In particular, we are interested in the extent to which people's behavior in an effort task is elastic to experimenter demand. Our estimates in turn provide us with a bound for the importance of experimenter demand effects for experimental economics. In our experiment, we manipulate our subjects' beliefs in a real effort task.

Subjects are randomly assigned to one of three main treatment conditions: in the first one, we induce "positive demand", by telling them that "you will do us a favor if you work harder than

[†]Institute for International Economic Studies (IIES); e-mail: jonathan.dequidt@iies.su.se.

[‡]Department of Psychology and Public Affairs, Princeton University; e-mail: haushofer@princeton.edu.

[§]Department of Economics, University of Oxford and CSAE; e-mail: christopher.roth@economics.ox.ac.uk.

you normally would". In the second treatment condition, we induce "negative demand" by telling our participants that "you will do us a favor if you work less hard than you normally would". In the third condition, we do not induce any demand. Subjects are randomly assigned to complete the real-effort task without any incentives for receiving more scores in the real effort task or with monetary incentives, i.e. they are given 1 cent for every 100 points they achieve.

This document proceeds as follows: first, we outline the experimental design. Then in section 3, we describe the setting as well as the sample size. Subsequently, we describe the hypotheses. In section 5, we describe the analysis we would like to conduct. Finally, in section 6, we define all of our outcome variables of interest.

2 Experimental Design

2.1 Effort Game

We use the same real effort task as in the seminal work by DellaVigna and Pope (2016). In this task subjects have to press two buttons on the keyboard alternately. In particular they receive the following instructions:

"*On the next page you will play a simple button-pressing task. The object of this task is to alternately press the a and b buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you alternate button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points. Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or the task will not be approved. Feel free to score as many points as you can.*"

2.2 Treatment Arms

We also use two of the treatment arms employed by DellaVigna and Pope (2016):

- **No monetary incentives**: "Your score will not affect your payment."

- **Monetary incentives**: "You will be paid one extra cent for every 100 points."

2.3 Demand Treatment Arms

On top of the the two main treatment arms, we cross-randomize three main "demand treatment arms":

- **Positive Demand:** Our subjects receive the following instructions: *You will do us a favor if you work harder than you normally would.*

- **Negative Demand:** Our participants get the following instructions: *You will do us a favor if you work less hard than you normally would.*

- **No Demand:** We do not induce any demand.

In addition to those main treatment arms, 250 additional individuals will be randomly assigned to complete the "no demand condition" under a different incentive scheme. In particular, we pay them 4 cents for every 100 points they score. This additional treatment arm in turn allows us to estimate the cost of effort function. For more details, see DellaVigna and Pope (2016).

2.4 Updated Beliefs

Thereafter, our respondents complete a set of questions on their beliefs about the experimenters' intentions and their hypotheses. In particular, we ask the following two questions:

- What do you think is the result that the researchers of this study want to find?

  - They want to find that people who were shown my instructions on average will work harder.

  - They want to find that people who were shown my instructions on average will work less hard.

- What do you think was the hypothesis of this research study?

  - The experimenters hypothesized that on average participants who were shown my instructions would work harder.

  - The experimenters hypothesized that on average participants who were shown my instructions would work less hard.

2.5 Demographics

Finally, we ask participants to complete a questionnaire on demographics, which includes variables on gender, age, education, income etc.

3 Setting, Sample Size and Power

We will run our experiment on Amazon Mechanical Turk, an online platform which is widely used to conduct experiments. We will only recruit participants who currently live in the United States. Moreover, workers must have completed at least 500 HITs, they must have an overall rating of more than 95 percent, and they must not have participated in one of our prior experiments on demand effects. As in DellaVigna and Pope (2016) we pay participants a show-up fee of \$1. We apply the same rules for the inclusion into the experiment as in DellaVigna and Pope (2016). In particular, we exclude participants that

- do not complete the MTurk task within 30 minutes of starting

- exit and then re-enter the task as a new subject (as these individuals might see multiple treatments)

- score 4000 or more points (it is physically impossible to score more than 3,500 points; so individuals scoring more than 3,500 points may be using bots)

- do not have valid mTurk ids and therefore did not get their HIT approved.

- did not enter a single button press (these workers may have experienced technical difficulties).

In total we recruit 1750 workers. We will have 750 workers in the "no incentive" treatment arm and 750 workers in the incentive treatment arm with an incentive of 1 cent per 100 points. Out of those 1500 workers in the "no incentive" and "1-cent inventive" arm, one third will additionally receive the positive demand instructions, one third will additionally receive the negative demand instructions and the remaining third will not receive any additional instructions (i.e. they receive the "no demand" treatment). In addition, we will have 250 participants in the "4-cent-incentive"-arm, i.e. participants receive a monetary payoff of 4 cents for every 100 points they score. They will not receive any positive or negative demand treatment. For our main test of interest, i.e. comparing the behavior of people in the positive and negative demand condition, we have a

power of .8 to detect standardized effect sizes of 0.177 at $\alpha = 0.05$. To test for heterogeneous demand treatment effects by monetary incentives and gender we have a power of .8 to detect effect sizes of 0.251 at $\alpha = 0.05$ respectively.

4 Main Hypotheses

Hypothesis 1 (Positive Demand): We hypothesize that individuals in the "positive demand condition" will exert higher levels of effort.

Hypothesis 2 (Negative Demand): We hypothesize that individuals in the "negative demand condition" will exert lower levels of effort.

5 Analysis

5.1 Baseline Balance

We will test for baseline balance for the following variables:

- gender

- age

- log income (income is the midpoint of the interval specified by the respondent)

- household size

- ethnicity (dummies for White, Black, Hispanic, and Asian)

- employment status (dummies for unemployed, part-time employed, and employed full-time)

- education (dummy for person with at least bachelor degree)

- experience on MTurk (number of HITs completed on MTurk)

- political orientation (taking value one for Republicans and zero otherwise)

We will regress each of these variables on a treatment indicator to see if there are imbalances. We will account for multiple hypothesis testing by regressing the treatment indicator on all of

the variables, and we will conduct a joint F-test, to see if the coefficients are jointly different from zero.

5.2 Main Specifications

5.2.1 Summary of main results

We present our main results by providing the means for each of our treatment arms separately.[1] In total we have six main treatment arms:

- No monetary incentives and no demand

- No monetary incentives and positive demand

- No monetary incentives and negative demand

- Monetary incentives and no demand

- Monetary incentives and positive demand

- Monetary incentives and negative demand

Thus in total, we present 6 means and standard deviations for our treatment arms of interest. We will also present these results graphically by the means of bar charts and by providing conventional 95 percent confidence intervals around the mean.

5.2.2 Test 1: does demand affect effort?

First, we simply compare the behavior of people in the positive demand group and in the negative demand group with that of people in the control group. We z-score our outcome variables at the (incentive) treatment arm-level, using the mean and s.d. for the no-demand group (Kling et al., 2007). We regress our z-scored outcome variable, ZY_i, on a treatment indicator, POS_i, which takes value one for people who receive the positive demand treatment, and value zero for all the other participants and on a dummy, NEG_i taking value one for people who receive the negative demand treatment, and value zero for all the other participants. Specifically, the equation that we estimate is:

$$ZY_i = \beta_0 + \beta_1 POS_i + \beta_2 NEG_i + \varepsilon_i$$

---

[1]For the following analysis we will not make any use of the data from the additional incentive arm in which people get 4 cents per 100 points they score.

We conduct three main tests:

- Do people increase their effort when exposed to positive demand? $\beta_1 > 0$

- Do people decrease their effort when exposed to negative demand? $\beta_2 < 0$

- Is the overall response to demand non-zero? $\beta_1 - \beta_2 = 0$

We correct for multiple hypothesis testing in these three tests by using the "sharpened q-value approach" (Anderson, 2008; Benjamini et al., 2006). In particular, we will adjust our p-values for a false discovery rate of .05.

5.2.3 Test 2: does demand differ for the treatment arms?

We also test whether the elasticity of behavior to experimenter demand differs in the different treatment arms. To do so, we interact our demand indicators with a dummy, $Treatment_i$, taking value one for people who complete the task for monetary incentives, while it takes value zero for people who complete the task without any monetary incentives. More precisely, we estimate the following equation:

$$ZY_i = \beta_0 + \beta_1 POS_i + \beta_2 NEG_i + \beta_3 Treatment_i \times POS_i + \beta_4 Treatment_i \times NEG_i + \varepsilon_i$$

To test for differences in response to demand by the incentive treatment arms we conduct the following test:

$$\beta_3 - \beta_4 \geq 0$$

which tests whether the size of the demand effect (difference in behavior under positive and negative demand) differs in the different incentive treatment arms.

5.2.4 Test 3: does demand differ between men and women?

Subsequently, we test whether the elasticity of behavior to experimenter demand differs for men and women. To do so, we interact our demand indicators with a dummy, $Male_i$, taking value one for males. More precisely, we estimate the following equation:

$$ZY_i = \beta_0 + \beta_1 POS_i + \beta_2 NEG_i + \beta_3 MALE_i \times POS_i + \beta_4 MALE_i \times NEG_i + \beta_5 MALE_i + \varepsilon_i$$

To test for gender differences in response to demand we conduct the following test:

$$\beta_3 - \beta_4 \geq 0$$

which tests whether the size of the demand effect (difference in behavior under positive and negative demand) differs by gender.

5.2.5 Test 4: Beliefs about experimental hypotheses

To check whether the demand treatments affected people's beliefs about what the researchers want to find and hypothesized, we estimate the following equation:

$$\text{Belief}_i = \alpha_0 + \alpha_1 POS_i + \alpha_2 NEG_i + \varepsilon_i$$

As before, we conduct three main tests:

- Are people more likely to believe we wanted/hypothesized high effort when exposed to positive demand? $\alpha_1 > 0$

- Are people more likely to believe we wanted/hypothesized low effort when exposed to positive demand? $\alpha_2 < 0$

- Is there an overall effect effect of demand on beliefs? $\alpha_1 - \alpha_2 = 0$

As above, we account for multiple hypothesis adjustment by adjusting the p-values for these six tests (three per beliefs question, two questions) for a false discovery rate of .05 (Anderson, 2008).

5.3 Dealing with Imbalances

If there are imbalances between the treatment group and the control group, we will also estimate the above equations with a set of controls. We will choose as controls all of the variables for which there are imbalances.

5.4 Differential Attrition

We will test whether attrition is related to the treatment by estimating the following equation:

$$A_i = \pi_0 + \pi_1 Treatment_i + \Pi^T X_i + \varepsilon_i$$

where A_i indicates if a participant did finish our experiment, $Treatment_i$, and where X_i is a vector of pre-determined characteristics. We will use the same set of pre-determined characteristics as for the baseline balance test.

If the coefficient π_1 on the treatment indicator is significant at the 5 percent level, we will use Lee bounds for the statistical analysis. This will allow us to bound our estimates. If the coefficient π_1 is not significant at the 5 percent level, we will conduct the statistical analysis without adjusting for attrition.

References

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484).

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika*, 93(3):491–507.

DellaVigna, S. and Pope, D. (2016). What motivates effort? evidence and expert forecasts. *National Bureau of Economic Research*.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Rosenthal, R. (1966). Experimenter Effects in Behavioral Research.

Zizzo, D. J. (2010). Experimenter Demand Effects in Economic Experiments. *Experimental Economics*, 13(1):75–98.

A   Experimental Instructions

Consent Form

You are being invited to take part in a this research study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please take the time to read the following information carefully. Please ask the researcher if there is anything that is not clear or if you need more information.

Purpose of the research:

You are asked to participate in a research study conducted by Johannes Haushofer, PhD.

Your participation in this research is completely VOLUNTARY. If you choose to participate you may subsequently withdraw from the study at any time without penalty or consequences of any kind. If you choose not to participate, that will not be penalized in any way or incur any losses.

The aim of the study is to examine social, and economic behavior. In particular, the study will test your performance and preferences on simple behavioral tasks. These tasks will be explained to you separately.

To be eligible to participate you have to be 18 years and above.

Study Procedures:

If you volunteer to participate in this study, you may be asked to make simple economic decisions, about which you will have been precisely instructed beforehand. Depending on your decisions, you may earn additional money in the experiment.

The experiment will last approximately 12 minutes.

Benefit and Risk:

There are no direct benefits to you from participating in this research study, other than the specified compensation. However, if you wish, you can send an email message to Johannes Haushofer (haushofer@princeton.edu) and we will send you a copy of any manuscripts based on the research (or summaries of our results). There are no foreseeable risks to participating in this

study. If you feel uncomfortable, you may stop participating at any time.

Confidentiality:

Your participation in this study will remain confidential, and your responses will be kept private, and no identifying information such as your name, phone number etc. will be collected at any point during

this research. Please note that research data will remain confidential, accessible only to the principal

investigator and the research staff in charge of this project. To preserve anonymity, any link between

your survey response and your mTurk account will be deleted upon payment.

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity.

Compensation:

You will be compensated for participating in the study. You will receive a minimum payment $1 for the 12-minute study. You may also receive additional payment depending on your performance in the task.

Who to contact with questions:

PRINCIPAL INVESTIGATOR:

Johannes Haushofer

Woodrow Wilson School of Public Policy and Department of Psychology

Princeton University

Phone: (617) 360-1605

Email: haushofer@princeton.edu

2. If you have questions regarding your rights as a research subject, or if problems arise which you do not feel you can discuss with the Investigator, please contact the Institutional Review Board at:

Office of Research Integrity and Assurance

Human Research Protection Program

Assistant Director

Phone: (609) 258-0865

Email: irb@princeton.edu

I understand the information that was presented and that:

A. My participation is voluntary, and I may withdraw my consent and discontinue participation in the project at any time. My refusal to participate will not result in any penalty.

B. By signing this agreement, I do not waive any legal rights or release Princeton University, its agents, or you from liability for negligence.

By clicking "Yes, I agree" below, I give my consent to participate in this study:

- Yes, I agree to participate in this survey.

- No, I do not agree to participate in this survey.

Demographics

Please indicate the highest level of education completed.

- Less than High School

- High School or equivalent

- Vocational/Technical School (2 year)

- Some College

- College Graduate (4 year)

- Master's Degree (MS)

- Doctoral Degree (PhD)

- Professional Degree (MD, JD, etc.)

- Other

What is your sex?

- Male

- Female

How old are you?

- Under 18

- 18 - 24

- 25 - 30

- 31 - 40

- 41 - 50

- 51 - 64

- 65 or over

No incentives: no demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

Your score will not affect your payment in any way.

Below is an example of how the game will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual game will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0

Your score will not affect your payment in any way.

No incentives: positive demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

Your score will not affect your payment in any way.

Below is an example of how the game will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual game will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0

Your score will not affect your payment in any way.

You will do us a favor if you work harder than you normally would.

No incentives: negative demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

Your score will not affect your payment in any way.

Below is an example of how the game will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual game will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0

Your score will not affect your payment in any way.

You will do us a favor if you work less hard than you normally would.

1 cent bonus: no demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

As a bonus, you will be paid an extra 1 cent for every 100 points that you score. This bonus will be paid to your account within 24 hours.

Below is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual task will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0
Bonus Payout: $ 0

You will be paid an extra 1 cent for every 100 points that you score

1 cent bonus: positive demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

As a bonus, you will be paid an extra 1 cent for every 100 points that you score. This bonus will be paid to your account within 24 hours.

Below is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual task will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0
Bonus Payout: $ 0

You will be paid an extra 1 cent for every 100 points that you score.

You will do us a favor if you work harder than you normally would.

1 cent bonus: negative demand

Instructions for effort task

You have 5 minutes maximum to read this page. If you finish early, you may proceed to the next page at your discretion.

On the next page you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible for 10 minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate** button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points.

Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or task will not be approved.

Feel free to score as many points as you can.

As a bonus, you will be paid an extra 1 cent for every 100 points that you score. This bonus will be paid to your account within 24 hours.

Below is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 5 as this is just practice, but the actual task will not have a limit.

Press 'a' then 'b'...

Points: 0

Effort Task

Press 'a' then 'b'...

Points: 0
Bonus Payout: $ 0

You will be paid an extra 1 cent for every 100 points that you score.

You will do us a favor if you work less hard than you normally would.

Beliefs

Thereafter our respondents complete a set of questions on their beliefs about the experimenters' intentions and their hypotheses. In particular, we ask the following two questions:

- What do you think is the result that the researchers of this study want to find?

  - They want to find that people who were shown my instructions on average will work harder.

  - They want to find that people who were shown my instructions on average will work less hard.

- What do you think was the hypothesis of this research study?

  - The experimenters hypothesized that on average participants who were shown my instructions would work harder.

  - The experimenters hypothesized that on average participants who were shown my instructions would work less hard.

Other demographics

The main part of the survey is now over. We will now just ask you some general questions about yourself.

What is the highest level of education you have completed? [12th grade or less; Graduated high school or equivalent; Some college, no degree; Associate degree; Bachelor's degree; Post-graduate degree]

What is your religion? [Christianity, Judaism, Islam, Hinduism, None, Other]

What is your ethnicity? [White, Black, Hispanic, Asian, Other]
What category would best describe your political orientation? [Democrat; Republican; Other]
Which of these describes your current situation most accurately? [Employed full-time, Employed part-time, Unemployed and looking for a job, Unemployed but not looking for a job, Retired, Other]

In which region do you currently reside? [Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA), Midwest (IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD), South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX), West (AZ, CO, ID, NM, MT, UT,

NV, WY, AK, CA, HI, OR, WA)]

What was your family's household income in 2015? [Less than $15,000, Between $15,000$ and $25,000, Between $25,000$ and $50,000, Between $50,000$ and $75,000, Between $75,000$ and $100,000, Between $100,000$ and $150,000, Between $150,000$ and $200,000, More than $200,000$, Prefer not to say]