Is the acknowledgement of earned entitlement effect robust across experimental modes and populations?

Forthcoming at Sociological Methods and Research

Abigail Barr^a, Luis Miller^{b,1}, Paloma Ubeda^c

^aSchool of Economics, University of Nottingham, Room B44, Sir Clive Granger Building, University Park Nottingham, NG7 2RD, UK; ^bSpanish National Research Council (IPP-CSIC), C/ Albasanz 26, 28037 Madrid, Spain;

^cDepartamento de Economía Aplicada y Estadística, UNED, C/Paseo de la Senda del Rey 11, 28040 Madrid, Spain.

Acknowledgments: A. Barr acknowledges support from the Economic and Social Research Council via the Network for Integrated Behavioural Sciences (Award No. ES/K002201/1). L. Miller acknowledges support from the Spanish Ministry of Economy and Competitiveness (grant ECO2015-67105-R). P. Ubeda acknowledges support from the Spanish Ministry of Science, Innovation and Universities (grant PGC2018-097875-A-100)

Abstract

We present a set of studies the objective of which was to test the robustness of the acknowledgment of earned entitlement effect across different experimental modes and populations. We present three sets of results. The first is derived from a between-subject analysis of two independent, but comparable samples of non-student adults. One sample participated in a standard, behavioral laboratory experiment and the other participated in a survey experiment. The two methods returned similar treatment effects. The second set of results relates to a sample of students drawn from a behavioral laboratory's pool of registered subjects. They participated in both the behavioral lab and survey experiments. We perform a between-subject comparison of the two treatment-elicitation methods but, this time, focusing on the same sample of subjects. Again, the treatment effects are very similar. Finally, we establish that within-subjects there is some consistency between decisions made under the two methods.

Keywords: survey experiments | laboratory experiments | non-student samples | cross-validation | distributive preferences

¹ To whom correspondence should be addressed: E-mail: luis.miller@csic.es

1. Introduction

With reference to experiments in the social sciences, Deaton and Cartwright (2018:12) recently wrote that ‘Establishing causality does nothing in and of itself to guarantee that the causal relation will hold in some new case, let alone in general.’ Bardsley et al. (2010:51) share this opinion, writing that, ‘strictly, all that happens in a particular laboratory experiment is what happens in it’. Using the conventional terminology in social science research, a major concern is that experiments may lack *external validity* (Campbell 1957; Campbell and Stanley 1963; Cook and Campbell 1979). This is disquieting, given that the main motivation for virtually any experiment is to say something of relevance *beyond* the local conditions of a particular experimental trial. A common response to this tension is to argue that what one generalizes is not a particular experimental result, but a theoretical construct: ‘Even simple theory can do much to interpret, to extend, and to use RCT results’ (Deaton and Cartwright 2018:15). Under this interpretation, ‘an experimental setting is externally valid if it is an instance of the theory it tests and neither adds nor subtracts anything from the theory’ (Zelditch 2014:96). Therefore, an experiment is *just* one instantiation of a theoretical construct and, if a new experiment offers further evidence supporting the theoretical construct, we can say that the external validity of the theoretical construct has increased.²

Evidence of external validity must, by definition, be empirical (Morton and Williams 2010:266), and most discussions on the generalizability of experimental results recommend *scientific replication* as the cornerstone of scientific advancement and knowledge accumulation (Camerer 2015; Falk and Heckman 2009). But, there are two very different ways of replicating an *inference* or *knowledge claim*: (1) by reproducing the exact experimental settings of the original

² Our investigation is related to the concepts of *construct* and *external* validity in the canonical taxonomy of Shadish et al. (2002:20-21). The former concerns the generalization from an experimental instance to a higher order theoretical construct; the latter the generalization to other instances.

trial;³ (2) by testing the same knowledge claim using a different experimental mode.⁴ Focusing on this second definition of replication, two methods emerge as the ideal candidates for increasing the generalizability of a theoretical construct: *theory-guided meta-analysis* and *cross-validation* or *empirical cross-checking*. The former uses data from previously conducted experiments to study whether methods, populations, and experimental settings affect findings. The latter compares experiments involving different methods to test the same theoretical construct. The most common example of empirical cross-validation or cross-checking in experimental social science involves the comparison of laboratory and field results on the same theoretical construct (Guala 2005: chapter 9; Bardsley et al. 2010: chapter 5). Correspondence between the lab and the field is typically interpreted as evidence of external validity. Here we extend this idea of empirical cross-validation to the correspondence between a behavioral and a survey experiment.

A few recent papers across the social sciences study the correspondence between behavioral and survey results measuring similar theoretical constructs. Becker et al. (2012) find weak correlations between economic preferences and psychological personality measures. Falk et al. (2018) present a dataset containing experimentally validated survey measures of five individual preferences and behavioral tendencies much studied by experimental social scientists. Chen and Tam (2019) correlate three measures of inequality preferences: behavioral games, vignette studies and conventional attitudinal survey items. Eifler (2010), Eifler and Petzold (2019) and Petzold and Wolbring (2019) discuss the validity of survey experiments more generally by looking at the correspondence between behavioral intentions reported in survey experiments and actual behavior

³ As in the recent replicability debate (Gilbert et al. 2016; Camerer et al. 2018).

⁴ Duch et al. (2019) show that ‘multi-mode replication has as good, if not better, payoffs than single-mode replication.’ By mode, they mean ‘how the experimental treatments are delivered to subjects.’

in the field. They find similarities between the two experimental modes, especially regarding treatment effects, but also quantitative deviations between the two.⁵

In this paper, we present the results of a cross-validation exercise to test the generalizability of a widely studied theoretical prediction: *earned entitlement is acknowledged*. First proposed by Homans (1961), this principle implies that justice exists when the received benefit of a group member is proportional to her investments. Adams (1965) materialized this idea in the well-known ‘equity formula’. Contrary to the Rawlsian egalitarian ideal (Haidt 2012; Rawls 1971), Homans’ and Adams’ proportionality principle offers a normative justification for some inequalities to be tolerated, justified, or accepted. In the last sixty years, a number of experimental studies have been conducted on the empirical validity of the equity or proportionality principle in sociology, economics, psychology, political science, and philosophy (e.g., Cook and Hegtvædt 1983; Konow 2003; Scott et al. 2001; Wagstaff, Huggins and Perfect 1996). One of the main conclusions of these studies was anticipated by Homans (1961: 246) when he claimed that people ‘differ in their ideas of what legitimately constitutes investment, reward, and cost, and how these things are to be ranked’. Here, we build on more recent definitions of justice as proportionality (Barr et al. 2015; Barr, Miller and Ubeda 2016; Cappelen et al. 2007; Demel et al. 2019; Haidt 2012; Konow 1996; Konow 2000; Konow 2003; Winter et al. 2012), and propose a testable principle—the acknowledgement of earned entitlement—that will be defined and operationalized in the Methods section below.

Our cross-validation involves survey and laboratory experiments with heterogeneous samples. At the core of this strategy is an experimental paradigm that has been extensively used in

⁵ From a theoretical point of view, our comparison speaks to the distinction between attitudes or intentions and behavior. The correspondence between attitudes and behaviors have been addressed by psychological approaches like the *Theory of Planned Behavior*. In a review of the empirical validity of this theory, Ajzen (1991) reports strong correlations between attitudes and behaviors.

several social sciences and has been previously called a *distributive justice experiment* (Barr et al. 2016). A distributive justice experiment typically entails two consecutive phases. In the production phase, people undertake an individual or group task. This phase might be hypothetical (a vignette or story created by the researchers) or real (actual participants in the experiment undertake the task). In the distribution phase, participants are asked to distribute the surplus created in the production phase. Again, this can be done hypothetically, that is, participants subjectively rate the fairness or social appropriateness of various proposed distributions or it can be real, participants actually (re)distribute monetary or non-monetary allocations. The main experimental manipulation involves informing participants in the distribution phase about the source of the surplus that is to be distributed. In *Earned* treatments,⁶ at the start of the distribution phase, participants are informed about the contribution or performance of individual subjects in the production phase. In *Random* or *Luck* treatments, at the start of the distribution phase, participants are informed that the initial distribution has been generated randomly, by chance, or by another arbitrary mechanism. The main theoretical prediction is that, on average, across methods, settings, and populations, a difference in productivity (*Earned* treatment) instead of luck (*Random* treatment) as the source of inequality causes an increase in the acceptability of that inequality.

We present three sets of results. The first is derived from a between-subject analysis of two independent, but comparable samples of non-student adults (referred to collectively as the adult sample below). One sample participated in a standard, incentivized behavioral experiment and was recruited following a number of strategies, including making use of mailing lists of public institutions, employment centers and local companies.⁷ The other sample was part of an ongoing

⁶ Sometimes also referred as *Merit*, *Effort*, or *Contribution* treatments.

⁷ For more details see Barr et al. (2016).

program evaluation and participated in a survey experiment during the first day of the program. Despite the many differences between the behavioral and survey experimental designs (see Methods section for details), they returned similar treatment effects, consistent with acknowledgement of earned entitlement in both cases. The second set of results relates to a sample of students drawn from a behavioral laboratory's pool of registered subjects (referred to below as the student sample). They participated in both the behavioral and the survey experiment. We perform a between-subject comparison of the two treatment-elicitation methods but, this time, focusing on the same sample of subjects. Again, the treatment effects are very similar. Finally, using the data from the latter experiment, we establish that, within-subjects there is some consistency between the decisions and responses made under the two modes.

Our study relates closely to the recent paper of Bader et al. (2018) on the sensitivity of laboratory findings to changes in experimental samples and settings. In the case of prosocial behavior, they find that quantitative effects vary significantly across conditions, while qualitative treatment effects are mostly consistent. The two papers are complementary in that they assess the generalizability of laboratory results in two different types of study, those that focus on preferences and behavioral tendencies that map directly onto choices, and those that focus on preferences that map onto differences in choices across contexts, i.e., that map onto treatment effects rather than choices.

2. Methods

In this section, first, we briefly present the design of our behavioral experiment, which is identical to that used by Barr et al. (2015), Barr et al. (2016), and Demel et al. (2019), and set out the analytical approach used to establish whether acknowledgement of earned entitlement manifests

as a treatment effect in the experiment. Second, we describe the design of the survey experiment, explain how it relates to the same theoretical framework, while differing in many ways from the behavioral experiment, and set out the analytical approach used to establish whether acknowledgement of earned entitlement manifests as a treatment effect in the experiment. Finally, we describe all the samples involved in the experiments, as well as the experimental procedures followed in each study.

2.1. Behavioral experiment

This experiment is non-computerized and has two stages. Participants first engage in an individual real-effort task and then play a four-person distributive justice (DJ) game. The real effort task involved sorting yellow and blue gravel into various containers for seven minutes. In the earned treatment, each participant's performance ranking in this real effort task determined his or her initial endowment in the four-person DJ game. In the random treatment, the initial endowments were randomly assigned. In both treatments, within a playing four, one participant received an initial endowment of 16€, one 12€, one 10€, and one 6€.

The DJ game was undertaken using specially designed and manufactured trays. Each participant received a tray. Each tray was divided into four quadrants, each quadrant relating to a participant. The tray-receiving participant's own quadrant was blue and located at the side of the tray closest to the participant when the tray was placed on a desk in front of him or her. Each quadrant contained a number of counters indicating the initial endowment of the corresponding participant. Each counter was worth €1. The participants were invited to rearrange the counters across the quadrants as they saw fit, while being instructed not to remove any of the counters from the tray. Once all the participants had made their allocations, those of one of the four, randomly selected, determined the payoffs.

In addition to their payoffs from the DJ game, each participant received €4. In the random treatment, this €4 was presented as a flat fee for the real effort task. In the earned treatment, the €4 was added to each of the possible earnings levels and then set aside to be collected at the end of the session. Thus, the €4 represented a minimum total final payoff for each experimental participant. There was no additional show-up fee.⁸

To investigate whether a sample of participants in the experiment described above acknowledge earned entitlement, focusing on the allocations made to others by those who did not take everything for themselves in the DJ game,⁹ one estimates linear regression Model 1:

$$x_{ij \neq i} = a_1 y_j + a_2 (E_i * y_j) + u_i + \varepsilon_{ij}$$

where E_i takes the value 1 if i played under the earned treatment and zero if i played under the random treatment; a_1 and a_2 are the coefficients to be estimated; u_i is the allocator fixed effect; and ε_{ij} is an allocation-specific idiosyncratic error. All other things being equal, a_1 is the slope of the relationship between j 's initial endowment and i 's allocation to j under the random treatment, $a_1 + a_2$ is the slope of the relationship between j 's initial endowment and i 's allocation to j under the earned treatment, and acknowledgement of earned entitlement manifests as a positive a_2 , i.e., as a significantly steeper slope in the relationship between j 's initial endowment and i 's allocation to j under the earned as compared to the random treatment.

⁸ Instructions and experimental materials can be found in the Online Appendix and in Barr et al. (2015), Barr et al. (2016) and Demel et al. (2019).

⁹ The decisions of those who take everything for themselves are excluded because they reveal only that the decision-maker is selfish and nothing about acknowledgement of earned entitlement. Only 12 (6.1%) and 7 (5.6%) subjects in the adult and student samples, respectively, behaved in this way. Including these subjects' decisions does not change the findings reported in the paper.

2.2. Survey instrument

The survey instrument presents participants with a scenario or vignette within which a decision-maker has to choose between two possible final resource distributions across three other people and asks the participants to state how fair they think each of the decision-makers' possible choices is on a seven-point scale, ranging from "very unfair" to "very fair" (-3 and +3 respectively in the results section below).

The scenario or vignette describes the second phase of a behavioral experiment involving groups of four subjects. Within this experiment, initial resource allocations were unequal. One of the group members received 16€, one 11€, and one 6€. The fourth group member received no initial resource allocation, but was to decide to either leave the (unequal) distribution across the other three as it was or equalize by allocating 11€ to each.

On the first page of the survey instrument, the distribution of initial allocations is represented graphically as counters or tokens on a circular "tray" divided into three sectors, each a different color (Figure 1). Then, on a second page (Figure 2), the distribution of initial allocations is shown again, using the same graphic image, followed by the two final distributions between which the decision-maker in the scenario had to choose. In one of the possible final distributions (Distribution 1), the final allocations equal the initial allocations. In the other (Distribution 2), the allocations are equalized. Below each of the two possible final distributions, appears the seven-point scale that the participants are asked to use to express their evaluations.

Figure 1: Survey experiment, page 1: Vignette

Across treatments in the survey experiment the information given to participants about the origin of the initial allocation in the scenario or vignettes varied. Under the earned treatment, participants were told that initial allocations were related to individual productivity. Under the random treatment, they were told that they were owing to luck. Each participant took part in both the earned and the random treatment. The order was randomized.

Figure 2: Survey experiment, page 2: Evaluation sheet

Suppose Individual A received a tray that looked like this:

on which each of the three segments related to one of the three other people involved in the task and the number of tokens in each segment depended on how productive the person to whom the quadrant relates was in the sorting task.

Then, Individual A had to choose 1 of the following 2 final token distributions:

Distribution 1

Distribution 2

Please indicate how fair you think each of these choices is by ticking one of the red boxes below under each choice

	Distribution 1	Distribution 2
Very fair	☐	☐
Somewhat fair	☐	☐
Somewhat unfair	☐	☐
Very unfair	☐	☐

There are several noteworthy differences between the behavioral and the survey experiments. First, in the behavioral experiment, the subject is the allocator, receives an initial endowment and can reallocate to or from him or herself, whereas in the survey experiment, the subject observes and evaluates an allocator's decisions and the allocator is a third-party who receives no initial endowment and cannot reallocate either to or from him or herself. These differences reflect the fact that some features of the behavioral experiment, while of value within the context of that experiment, were redundant and a potential distraction to the participants within the context of the survey experiment. Specifically, while in a behavioral experiment, by making subjects' payoffs dependent on their own decisions, one expects to enhance the validity of those

decisions as indicators of the subjects' preferences, in a survey experiment, this is not an option. And, once this option is removed, one is free to refocus the design on the primary objective which, in this case, is to investigate whether individuals consider inequality to be more appropriate or fair when it is earned rather than owing to luck. So, in the survey experiment subjects are ideally placed in an evaluative role and it is useful to exclude the possibility that their evaluations are affected by how selfish they think a decision-maker is by rendering that decision-maker unable to affect his or her own payoff, i.e., by placing the decision-maker in a third-party rather than a stakeholder role.

Prior to running the survey experiment described above, we conducted a pilot study involving 79 students drawn from the same laboratory pool as the actual student experiments. In this pilot, we cross-cut the earned versus random initial endowment treatments with whether the decision-maker in the vignette was a third-party or a stakeholder. An earned endowment effect was observed in both decision-maker-type variants. However, within both the earned and random treatment arms, the evaluations varied less under the third-party compared to the stakeholder treatment. The lower variance under the former is consistent with selfishness being excluded as a possible motivation for the decision-maker.¹⁰

In addition to the differences described above, in the behavioral experiment the decision-maker faces an extensive choice set made up of all the possible distributions of 44 units across four individuals, whereas in the survey experiment the respondents are told that the decision-maker has only two options, leave the initial endowments untouched (Distribution 1) or equalize (Distribution 2). In the behavioral experiment, even though the choice set is extensive, only four allocations need to be made and this can be done in just a few minutes. In contrast, evaluating each and every one of the possible distributions of 44 units across four individuals would have taken

¹⁰ Full results available on request.

hours. So, in the survey experiment, to reduce time costs, we limited the number of final distributions that the subjects had to evaluate to the obvious candidates, Distributions 1 and 2. In the pilot study mentioned above, to explore whether and how extending the list of distributions to be evaluated might affect evaluations, we added another cross-cutting treatment in which the decision-maker's choice set included an inversion of the initial endowment distribution (Distribution 3). Within both the earned and random treatment arms, this distribution was considered to be unfair and its inclusion in the choice set increased the mean fairness evaluation and reduced the variance in evaluations for both Distribution 1 and Distribution 2. The overall effect was greater statistical power when testing the hypotheses of specific interest, but the power gain was insufficient to warrant an increase in the number evaluations made by 50%. We also considered adding a weighted average of Distributions 1 and 2 to the choice set, but decided against it as the weights would have been arbitrary and, again, it would have increased the number of evaluations to be undertaken by 50%. The time saved by asking the survey participants to evaluate only two of the many possible distributions was used to engage each subject in evaluations under both treatments, thereby ensuring perfect balance. The order of the treatments was randomized between subjects.

To establish whether a sample of individuals acknowledges earned entitlement in the survey experiment, one can compare the fairness evaluations of Distribution 1 under the earned and random treatments. If Distribution 1 is considered significantly fairer under the former compared to the latter, it can be taken as evidence that earned entitlement is acknowledged. To corroborate this finding, one can test whether Distribution 2 is considered significantly less fair under the earned compared to the random treatment.

2.3. Samples and Procedures

The participant sample for the behavioral experiment ($N=198$) consists of a group of non-students recruited in the Spanish province of Biscay for the experiments conducted in 2013 and 2014 and reported in Barr et al. (2016) and Demel et al. (2019). Barr et al. (2016) and Demel et al. (2019) describe the procedures followed in the experiment in full. The participant sample for the survey experiment ($N=91$) consists of adults, who were taking part in an intervention conducted by the Government of Biscay. The survey experiment was conducted in 2016. In both samples there are more females (63% and 53% in the survey and incentivized experiment, respectively), most subjects have tertiary education (59% and 84%) and more than half are unemployed (54% and 56%). The sample in the survey experiment is somewhat older (42 vs. 27 years on average). In the statistical analysis, we control for these socio-demographic variables.

Participants in the within-subject experiment were undergraduate students ($N=124$) recruited through the laboratory LABEAN's recruitment site in Bilbao (Biscay). We adopted a 2 (treatments) by 2 (orders) factorial design. Each of these conditions was conducted in a different experimental session. Participants received a participation fee and their earnings from the behavioral experiment. The within-subject experiment was conducted in early 2017.

3. Results

In this section, we present three sets of results. First, we compare the results for the two independent adult samples that participated in the survey experiment and the behavioral experiment (subsection 3.1). Second, we replicate the same analysis but using the data for the student sample that participated in both the survey and the behavioral experiment (subsection 3.2). Finally, focusing on the student sample, we look at the within-individual correlation between behavior in the behavioral experiment and evaluations in the survey experiment (subsection 3.3).

3.1. Between-subject analysis of the adult sample

Figure 3 compares the treatment effects obtained under the two modes. In the left panel, we plot the average evaluation of the proportional tray ($x_i=y_j$) by the adult sample under each treatment in the survey experiment. To exclude learning effects, we include in the analysis only the evaluations made by participants under the treatment (random or earned) they faced first. In the random treatment, the average evaluation of the proportional tray is negative and significantly different from zero. In the earned treatment, the average evaluation of this tray is positive and significantly different from zero. There is a significant treatment effect (Table 1, columns 1 and 2; $p<0.001$) in the evaluation of the proportional tray caused by the experimental treatment. There is also a significant treatment effect in the evaluation of the equal tray. That tray is evaluated more positively in the random than in the earned treatment (Table 1, columns 3 and 4; $p<0.001$).

Figure 3: Between-subject analysis of the adult sample

Note: The figure shows the average fairness evaluations under the two treatments in the survey experiment (left panel) and the average slopes of the initial-endowment-allocation relationships under the two treatments in the behavioral experiment (right panel). The whiskers indicate standard errors. The bars in the right panel (survey) are derived from the regression reported in column 1 of Table 1. The bars in the left panel (behavioral) are derived from the regression reported in column 1 of Table 2.

The right panel of Figure 3 presents the treatment effect in the behavioral experiment.¹¹ The heights of the bars indicate the slope of the relationship between j 's initial endowment and i 's allocation to j in the distribution task. This slope is not significantly different from zero in the random treatment, but positive and significant in the earned treatment, meaning that participants in the earned treatment acknowledge initial endowments when making final allocations, while those in the random treatment do not. The difference between the slopes in the earned and random treatments is positive and statistically significant (Table 2, columns 1 and 2; $p < 0.001$).

Comparing visually the two panels of Figure 3, it becomes clear that both methods return similar treatment effects: both go in the same direction and are statistically significant at the 0.1% significance level.

Table 1: Regression analysis of the acknowledgement of earned entitlement in the survey experiment

Dependent variable = evaluations of the tray

	Adult proportional		Adult equal		Student proportional		Student equal	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Earned	2.750** (0.386)	2.663** (0.419)	-2.604** (0.371)	-2.567** (0.401)	2.519** (0.239)	2.560** (0.242)	-2.519** (0.293)	-2.600** (0.291)
Female		0.707 (0.444)		-0.673 (0.420)		-0.325 (0.245)		0.665* (0.294)
Age		-0.003 (0.023)		-0.023 (0.022)		-0.038 (0.055)		0.077 (0.66)
Unemployed		0.048 (0.447)		0.070 (0.423)				
Tertiary Education		0.276 (0.435)		-0.598 (0.411)				
Constant	-1.058** (0.253)	-1.608** (0.499)	1.736** (0.239)	2.374** (0.476)	-0.644** (0.172)	-0.156 (0.404)	2.300** (0.210)	1.296** (0.488)
Observations	91	83	91	83	123	123	123	123

Note: ** - sig. at 1%, * - sig. at 5%

¹¹ In this experiment, 46% and 30% of participants distribute the money equally across the four participants in the random and earned treatments, respectively, and only 6.1% are completely selfish across treatments. The rest condition distributions on initial endowments to some degree.

Table 2: Regression analysis of the acknowledgement of earned entitlement in the behavioral experiment

Dependent variable = i 's allocation to j (x_{ij})

	Adult		Student	
	(1)	(2)	(3)	(4)
y_j	0.021 (0.038)	0.091 (0.074)	0.096** (0.036)	0.075 (0.087)
y_j * Earned	0.251** (0.047)	0.258** (0.047)	0.160** (0.051)	0.159** (0.053)
y_j * Female		0.036 (0.046)		0.014 (0.053)
y_j * Age		0.009 (0.006)		0.003 (0.011)
y_j * Unemployed		-0.115** (0.045)		
y_j * Tertiary Education		-0.039 (0.065)		
Observations	558	558	351	351
Participants	186	186	117	117

Notes: Samples include allocations to others; participant fixed effects, u_i , included in all models; j 's initial endowment (y_j) = j 's initial endowment expressed as a proportion of the 44 tokens in the game; Earned (E)=1 if i made allocations under the earned treatment, =0 if i made allocations under the random treatment; ** - sig. at 1%.

3.2. Between-subject analysis of the student sample

Figure 4 replicates the analysis reported in Figure 3, but focuses on the student sample that took part in both the survey and behavioral experiment. Again, the evaluation of the proportional tray is significantly higher in the earned compared to the random treatment (Table 1, column 5; $p < 0.001$) and the slope in the incentivized task is significantly higher in the earned compared to the random treatment (Table 2, column 3; $p < 0.001$), although the slope difference is lower than in the adult-sample analysis. Using a student sample, we have replicated the main finding reported above: the survey and behavioral tasks produce significant treatment effects that go in the same direction.¹²

¹² Table 2 reveals one difference between the adult and student samples; in the behavioral experiment, students take random initial endowments into account, but the treatment effect is still significant and with controls the effect of random initial endowments is not significant anymore.

Figure 4: Between-subject analysis of the student sample

Note: The figure shows the average fairness evaluations under the two treatments in the survey experiment (left panel) and the average slopes of the initial-endowment-allocation relationships under the two treatments in the behavioral experiment (right panel). The whiskers indicate standard errors. The bars in the right panel (survey) are derived from the regression reported in column 5 of Table 1. The bars in the left panel (behavioral) are derived from the regression reported in column 3 of Table 2.

3.3. Within-subject analysis of the student sample

Finally, we study the within-subject correlation between behavior and evaluations in the student sample. For each student participant i , first, we compute the slope of the relationship between j 's initial endowment and i 's allocation to j in the behavioral experiment.¹³ Then, we look at whether and how these correlate with the i 's evaluations under the same treatment in the survey experiment. In the left panel of Figure 5, evaluations of the proportional tray are plotted against individual

¹³ We do this by estimating Model 1 for each individual subject separately. We regress the three allocations to other participants in the game on these other participants' initial endowments.

slopes in the behavioral experiment, pooled across treatments. As expected, there is a positive and significant relationship (*Spearman's Rho* = 0.266; $p=0.004$). However, there is also evidence of considerable noise in the relationship. In the right panel of Figure 5, evaluations of the equal tray are plotted against individual slopes in the behavioral experiment. Here, as expected, the correlation is negative and significant (*Spearman's Rho* = -0.310; $p<0.001$) but, once again, noisy. In summary, we find that, on average, those who evaluate the proportional tray more positively in the survey also acknowledge performance in the behavioral task more. In contrast, on average, those who evaluate the equal tray more positively, acknowledge performance in the behavioral task less.

Figure 5: Correlation between behavior and evaluations

Note: The figure shows correlations between evaluations of the proportional (left) and equal (right) trays in the survey task and the slope of the relationship between initial endowments and final allocations in the behavioral experiment.

4. Discussion

Social sciences differ in the methods they use to estimate similar, sometimes the same, treatment effects. For instance, in experimental economics, behavioral experiments prevail, while in experimental psychology, politics and sociology hypothetical survey tasks are more commonplace. An important question is whether these distinct methodological approaches produce different insights. In this paper, we have focused on a particular treatment effect that has received a great deal of attention in sociology, economics, psychology and political science for almost sixty years (Homas 1961; Adams 1965; Konow 2003): the idea that people tend to respect initial inequalities more when those inequalities are earned rather than owing to random assignment. We have compared results from a behavioral experiment with results from a survey experiment, and we have done so using both student and non-student samples. Both methods yield remarkably similar aggregate treatment effects for both samples indicating that the *acknowledgment of earned entitlement* theoretical prediction has, at least, some external validity.

This result is in line with previous papers reporting similar treatment effects in field and survey experiments (Eifler 2010; Petzold and Wolbring 2019). One of the strengths of our validation exercise is that we obtain very similar treatment effects using experimental methods that differ in a number of design and implementation features. This being the case, our cross-validation findings may be viewed as conservative. However, further research is needed to uncover potential heterogeneous effects involving the interaction between the *acknowledgement of earned entitlement* and other operationalizations of the units, treatments, and experimental settings.

The within-subject correlations between behaviors and evaluations add further validity. However, here, the noise in the observed relationships is worthy of note. In future work it would be interesting to isolate the effects played by real incentives, individual selfishness and other

contextual and individual characteristics in driving a wedge between individuals' evaluations of the behavior of others and their own behavior and to think about what this implies of our methodological choices.

References

Adams, J. S. 1965. "Inequity in social exchange," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, vol. 2. New York: Academic Press.

Ajzen, I. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50: 179-211.

Bader, F., B. Baumeister, R. Berger, and M. Keuschnigg. 2019. "On the Transportability of Laboratory Results." *Sociological Methods and Research*.

Bardsley, N., R. Cubitt, G. Loomes, P. Moffat, C. Starmer, and R. Sugden. 2010. *Experimental economics: Rethinking the rules*. Princeton: Princeton University Press.

Barr A., J. Burns, L. Miller, and I. Shaw. 2015. "Economic Status and Acknowledgement of Earned Entitlement." *Journal of Economic Behavior and Organization* 118:55–68.

Barr A., L. Miller, and P. Ubeda. 2016. "Moral Consequences of Becoming Unemployed." *Proceedings of the National Academy of Sciences of the United States of America* 113(17):4676–4681.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4:453-478.

- Camerer, C. F. 2015. “The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List,” in G. Frechette (ed.), *Handbook of Experimental Economic Methodology*. Oxford: Oxford University Press.
- Camerer, C. F., A. Dreber, F. Holzmeister, T. H. Ho, J. Huber, M. Johannesson, ... and A. Altmeld. 2018. “Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015.” *Nature Human Behavior* 2(9):637.
- Campbell, D. T. 1957. “Factors relevant to the validity of experiments in social settings.” *Psychological bulletin* 54(4):297.
- Campbell, D. T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Rand McNally & Company.
- Chen, J., and T. Tam. 2019. “Uses of Artificial and Composite Treatments in Experimental Methods: Reconsidering the Problem of Validity and Its Implications for Stratification Research.” *Research on Social Stratification and Mobility*.
- Cook, T.D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.
- Cook, K. S., and K. A. Hegtvedt. 1983. Distributive Justice, Equity, and Equality. *Annual Review of Sociology* 9:217–241.
- Cappelen, A., A. Hole, E. Sorensen, and B. Tungodden. 2007. “The Pluralism of Fairness Ideals: An Experimental Approach.” *American Economic Review* 97(3):818–827.

- Deaton, A., and N. Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials." *Social Science & Medicine* 210:2-21.
- Demel, S., A. Barr, L. Miller, and P. Ubeda. 2019. "Commitment to political ideology is a luxury only students can afford: A distributive justice experiment." *Journal of Experimental Political Science* 6(1):33-42.
- Duch, R., D. Laroze, T. Robinson, and P. Beramendi. 2019. "Multi-Modes for Detecting Experimental Measurement Error." *Political Analysis*.
- Eifler, S. 2010. "Validity of a factorial survey approach to the analysis of criminal behavior." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6:139–146.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133(4):1645–1692.
- Falk, A., and J. J. Heckman. 2009. "Lab experiments are a major source of knowledge in the social sciences." *Science* 326(5952):535-538.
- Gilbert, D. T., G. King, S. Pettigrew, and T. D. Wilson. 2016. Comment on "Estimating the reproducibility of psychological science." *Science* 351(6277):1037-1037.
- Guala, F. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Güth, W. 1988. “On the behavioural approach to distributive justice –A theoretical and experimental investigation.” In S. Maital (ed.), *Applied Behavioral Economics*. Brighton: Wheatsheaf Books.

Güth, W., and R. Tietz. 1985. “Strategic power versus distributive justice: an experimental analysis of ultimatum bargaining.” In H. Brandstatter and E. Kirchler (eds.), *Economic Psychology* Linz: Rudolf Trauner Verlag.

Haidt, J. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon Books.

Hoffman, E., and M. L. Spitzer. 1985. “Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice.” *Journal of Legal Studies* 14: 259–298.

Homans, G. C. 1961. *Social behavior: Its elementary forms*. London: Routledge Kegan Paul.

Konow, J. 1996. “A Positive Theory of Economic Fairness.” *Journal of Economic Behavior & Organization* 31(1):13–35.

Konow, J. 2000. “Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions.” *American Economic Review*, 90(4): 1072–1091.

Konow, J. 2003. “Which is the fairest one of all? A positive analysis of justice theories.” *Journal of Economic Literatur*. 41(4):1188–1239.

Morton, R., and K. Williams. 2010. *Experimental Political Science and the Study of Causality*. Cambridge: Cambridge University Press.

Petzold, K., and T. Wolbring (2019). “What Can We Learn From Factorial Surveys About Human Behavior? A Validation Study Comparing Field and Survey Experiments on Discrimination.” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 15(1):19–30.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Scott, J., R. Matland, P. Michelbach, and B. Bornstein. 2001. Just Deserts: An Experimental Study of Distributive Justice Norms. *American Journal of Political Science* 45(4):749–767.

Selten, R. 1976. “The Equity Principle in Economic Behavior.” In H. Gottinger, and W. Leinfellner (eds.), *Decision Theory and Social Ethics*, Dordrecht: D. Reidel Publishing Company.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company. Wagstaff, G. F., J. Huggins, and T. J. Perfect. 1996. Equality ratio equity, general linear equity and framing effects in judgments of allocation divisions. *European Journal of Social Psychology* 26(1):29–41.

Winter, F., H. Rauhut, and D. Helbing. 2012. “How Norms can Generate Conflict: An Experiment on the Failure of Cooperative Micro-Motives on the Macro-Level.” *Social Forces* 90:919-946.

Zelditch, M. 2014. “The External Validity of Experiments That Test Theories.” In. Webster, M. and J. Sell (eds.), *Laboratory Experiments in the Social Sciences*. Elsevier.