

Pelloth, Daniel; Hoffmann, Patrick

Working Paper

Letting Down the Team? Social Effects of Team Incentives - Reproduction Report of Babcock et al. (2015)

I4R Discussion Paper Series, No. 188

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Pelloth, Daniel; Hoffmann, Patrick (2024) : Letting Down the Team? Social Effects of Team Incentives - Reproduction Report of Babcock et al. (2015), I4R Discussion Paper Series, No. 188, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/307159>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

No. 188

I4R DISCUSSION PAPER SERIES

Letting Down the Team? Social Effects of Team Incentives – Reproduction Report of Babcock et al. (2015)

Daniel Pelloth

Patrick Hoffmann

December 2024

I4R DISCUSSION PAPER SERIES

I4R DP No. 188

Letting Down the Team? Social Effects of Team Incentives – Reproduction Report of Babcock et al. (2015)

Daniel Pelloth¹, Patrick Hoffmann^{1,2}

¹University of Augsburg/Germany

²Universitat Autònoma de Barcelona/Spain

DECEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

Letting Down the Team? Social Effects of Team Incentives - Reproduction Report of Babcock et al. (2015)

Daniel Pelloth¹ and Patrick Hoffmann^{1,2}

¹Faculty of Business and Economics, University of Augsburg

²ICTA, Universitat Autònoma de Barcelona

August 29, 2024

The experimental study "Letting Down the Team? Social Effects of Team Incentives" by Philip Babcock and colleagues (2015) proposes that team incentives significantly enhance individual performance through social pressure and peer effects. The findings suggest that individuals are motivated by a desire to avoid disappointing their teammates, indicating that social dynamics, such as guilt and social pressure, play a crucial role in shaping behavior in team settings. In this report, we computationally reproduce the results from the original paper and perform several robustness checks. Overall, we ascertain the good reproducibility of the study and find that the results hold across the performed robustness checks.

1 Introduction

Babcock et al. (2015) investigate the impact of team-based incentives on individual performance. The authors conducted two separate field experiments comparing individual and team-incentive structures and find that team incentives significantly increase effort (measured by study hall or gym attendance), with participants showing a 9%-17% increase in effort compared to those receiving individual incentives. The authors argue that social pressure and the fear of letting down teammates are key motivators, suggesting that individuals are driven by a sense of responsibility to their peers rather than solely by financial gain. Despite individual incentives offering higher private returns, the team-based approach proves more effective in eliciting effort. This report conducts a computational reproduction of all tables and figures using the provided reproduction package and several robustness checks. For the latter, we first excluded missing variables and certain (outlier) observations. Subsequently, we computed heterogeneity checks to illustrate how the effects differ for various subgroups. Finally, the average marginal effect for each class was calculated for individual and team treatments. The robustness checks only refer to the main results of the paper (Table 4 in our report and Table

2 in the original study). Overall, we can successfully reproduce all main tables and figures from the original study, and our robustness analysis lends further support to the direction of the findings.

2 Study overview

The study consists of two separate but similar experiments. In the first experiment, participants were incentivized to visit a 24-hour study room in a university library over a two-week period. Subjects were randomly assigned either to a control group, which received \$2 per visit (up to four visits), or to one of four treatment groups. Participants of the individual treatment condition got an additional \$25 bonus for attending the library at least four times. Subjects of the team treatment condition got the additional bonus of \$25 only if both team members accumulated at least four visits. The same counts for the anonymous team treatment group, with the difference that their team partner was unknown. Lastly, in the choice treatment, the participants could decide which treatment they wanted to be in. In the second experiment, students were incentivized for exercising in the campus gym following a similar design. The major difference is that there were no anonymous team or choice treatments, and the control group did not receive any payment. We reproduced the results from both experiments, but focus our robustness checks only on the main results from the first experiments.

3 Reproducibility

The reproducibility assessment is based on the provided Stata replication package, which includes a dataset and a do.file. Unfortunately, the package does not contain a "read-me" file or any other additional information to support the comprehension of the analytical procedure and the code in the do.file. Although this slightly hampers a neasy understanding of the procedure, it does not hinder the overall reproduction. In our understanding, the available dataset contains the raw data from the experiments. The Stata code runs without errors and the outputs largely correspond to the results reported in the paper. Overall, it was possible to reproduce all tables and figures of the original paper, except for Table 4. Table 4 shows descriptive data from a post-experimental survey which is not included in the reproduction package. All reproduced Tables and Figures are listed in the Appendix.

Although we successfully reproduce the general results of the analysis, there are some remarks to make regarding the cleanliness of the code, the formatting, and minor deviations when comparing the replications with the published results. For the second part of Table 4, the code included a calculation error resulting in a calculated cost-per-visit for the team treatment of \$1.85, while the table in the paper shows \$4.87. The mistake in the code arises from the omission of the number of team treatment bonuses in the cost calculation. Correcting this results in the correct value of \$4.87. We also observe a formatting error in the original

Table 2 as standard errors of "Year is 2013" and the constant are listed in the wrong columns. On the one hand, Stata doesn't give standard errors or a coefficient for the year 2013 because of collinearity, on the other hand, the standard error values match those for columns (1)-(3). This is a minor editing error which should nevertheless be mentioned to avoid confusion. Running the code also results in small differences for standard errors (2 cases), coefficients (2), and a p-value (1). For example, the two different SEs are found for the choice treatment in Table 4. Our reproduction yields a SE of 0.234, whereas in the published version it is 0.167. This is difficult to explain. As differences in SE could theoretically arise from random estimation processes differences in software version or configurations, such differences should be noticeable across columns and not only singularly, as in our replication. The differences for the coefficients (Table 7 panel B and Table 10 panel D), as well as the p-value in Table 3 (column 5), could be due to rounding or copying mistakes from Stata to the tables since they are minor (e.g. 1.876 vs. 1.877).

As improvements, we have eliminated the errors and optimized the do.file. By labeling variables and using more detailed notes and headers, we have ensured better traceability and understanding.

4 Robustness Checks

We check for robustness of the main results (Table 4), using the regression model from Table 4, columns (4)-(6), as the base model. Besides the four treatment groups, this includes the control variables age, male, and library days, as well as the missing values of age, gender, pre-treatment library, and class fixed effects. For simplification and a better overview, these control variables are not listed in the following tables.

Data manipulation checks

As a first manipulation check, we exclude missing values and certain observations from the analysis. While the authors included missing values of library days, age, and male as control variables, we have removed observations with missing data (14). Next, aiming to rule out the effect of outliers and to homogenize the sample, we removed participants older than 26 (13) and those with more than 6 visits to the study room (25). Considering overlaps, we removed 46 observations in total. Participants with a high number of visits could distort the effect, as they might be more intrinsically driven to study and less motivated by the bonus payment which is reached already by 4 visits. The manipulation checks are displayed in Table 1. Columns (1)-(3) illustrate the results for eliminating the missing values. Columns (4)-(6) additionally removed older subjects and those with over six visits. For both specifications, the results are very similar in comparison to each other and in comparison to the main results.

Heterogeneity checks

The authors themselves did not conduct a heterogeneity analysis to check if the results differ for various subgroups. In Table 2, the effect on the number of visits is split according to age

Table 1: Data manipulation checks

	Number of visits (1)	≥ 1 Visit (2)	≥ 4 Visit (3)	Number of visits (4)	≥ 1 Visit (5)	≥ 4 Visit (6)
Any Treatment	0.815*** (0.213)	0.162*** (0.051)	0.236*** (0.045)	0.845*** (0.192)	0.167*** (0.052)	0.250*** (0.045)
Team treatment	0.456** (0.194)	0.117*** (0.044)	0.080* (0.044)	0.404** (0.178)	0.116*** (0.045)	0.073 (0.045)
Anonymous team treatment	-0.598** (0.234)	-0.098* (0.055)	-0.143*** (0.052)	-0.543** (0.218)	-0.092 (0.056)	-0.147*** (0.053)
Choice treatment	0.554** (0.233)	0.124** (0.055)	0.096* (0.054)	0.550** (0.224)	0.125** (0.056)	0.095* (0.055)
R-squared	0.12	0.11	0.11	0.13	0.11	0.12

Notes: Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses. Columns (1)-(3) exclude missing age, sex and library days. Columns (4)-(6) exclude additionally people older than 26 and those more than 6 visits.

*** Significant at 1%; ** significant at 5%; * significant at 10%

(columns 1-2), gender (columns 3-4), and the subsample of participants that visited the study room at least once (column 5). Age was divided into the subgroups ≤ 20 and ≥ 21 to ensure a relatively equal division of the sample. Here, column 5 should not be confused with the original analysis, which accounts for the probability of at least one study room visit.

Our heterogeneity checks show a decrease in the significance of the results. A part from the individual treatment, only three coefficients are significant at the 5% level. One reason for this is the reduced sample size per regression, which makes the detection of a statistically significant correlation less likely. Columns (1)-(2) show a discrepancy between men and women. The individual treatment leads to 0.34 more visits for men compared to women, while the team treatment has a slightly better effect for women (0.17). The most striking result is that the negative effect of the anonymous team treatment is entirely attributed to the male participants. Thus, the negative coefficient of the control variable male in Table 4 can be explained by the anonymous team treatment. Concerning age, a surprising difference of 0.35 visits can also be recognized for the team treatment. Column 5 shows that the team treatment did not result in more visits compared to the individual treatment. The majority of the effect from Table 4 can therefore be explained by more people in the team treatment visiting the study room at least once.

Overall, Table 2 confirms the results of Table 4. The team treatment did lead to more visits to the study room than the individual treatment for male, female, and both age groups. Furthermore, it illustrates that the anonymous team treatment does not induce increased effort for any subgroup. Finally, it appears that the different treatments could lead to different

Table 2: Heterogeneity check for number of visits

	Male (1)	Female (2)	≤ 20 (3)	≥ 21 (4)	Participants with at least one visit (5)
Any Treatment	0.977*** (0.311)	0.639** (0.277)	1.009*** (0.273)	0.833** (0.343)	0.261 (0.273)
Team treatment	0.300 (0.266)	0.472* (0.268)	0.184 (0.246)	0.535* (0.297)	-0.000 (0.149)
Anonymous team treatment	-1.085*** (0.317)	0.108 (0.380)	-0.190 (0.376)	-0.901*** (0.336)	-0.259 (0.207)
Choice treatment	0.352 (0.348)	0.647** (0.320)	0.510* (0.297)	0.551 (0.371)	0.176 (0.168)
R-squared	0.13	0.21	0.15	0.18	0.11

Notes: The table displays the effects for various subgroups as indicated by the column header. Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses.

*** Significant at 1%; ** significant at 5%; * significant at 10%

effects depending on gender and age. Additional research could address this gender and age gap and clarify the extent to which both factors play a role.

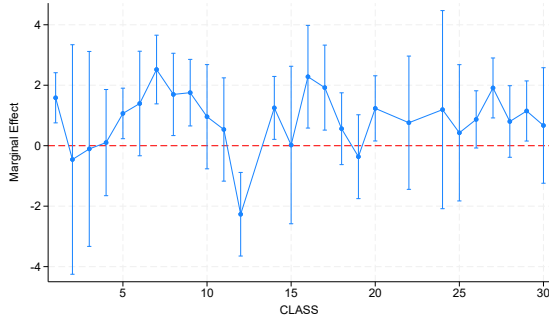
Average marginal treatment effect for each class

In this section, we investigate whether the team treatment effect changes depending on the different classes. For this purpose, we calculate the average marginal effect (AME) for each class. The regression is based on Table 4 columns (1)-(3), with the four treatment conditions and standard errors clustered at the group level. It is not possible to estimate the AME with the "absorb(CLASS)" command since it absorbs the effect of "class". To avoid creating a new regression model, the model without fixed effects is used, as this led to very similar results.

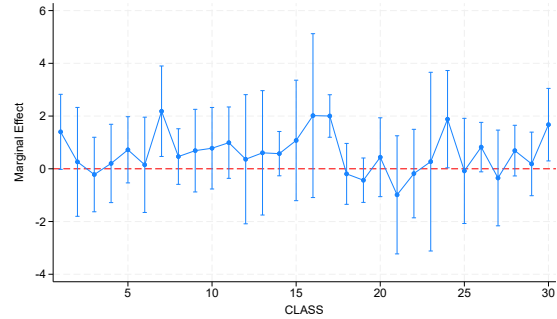
In total, the participants were selected from 30 different courses. The course and study degree could potentially influence the number of study room visits, as certain classes are better suited for studying in the library. Although the sampled students' classes are only from the economics and law faculties, there could be a difference between attendance, as some courses are rather math-based and some require memorization. The results are illustrated by a margins plot (Figure 1). Column A shows the AME for the individual treatment. Column B illustrates the additional effect of the team treatment. It is particularly important to determine whether a systematic pattern is recognizable or if certain classes have a higher effect on several outcomes and if one can see the same effects for both treatment groups. Certain courses could be beneficial for both treatments or only for one of the two.

Overall, we see that the AME differs greatly between columns A and B. For the majority

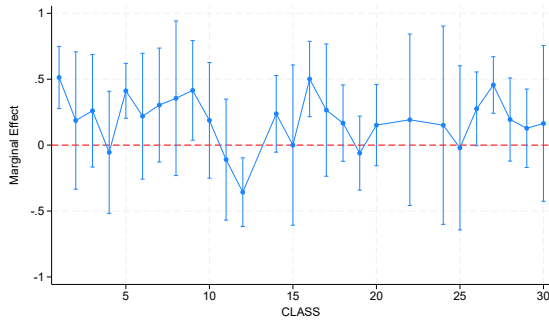
Column A. Any treatment
Panel A. Number of visits



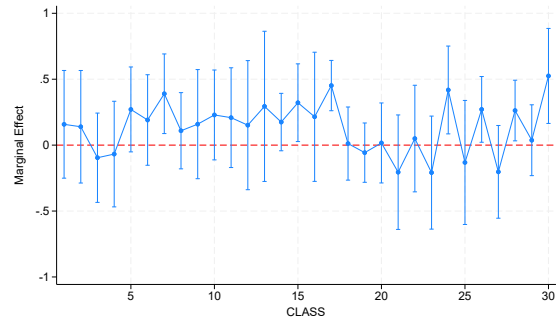
Column B. Team treatment
Panel D. Number of visits



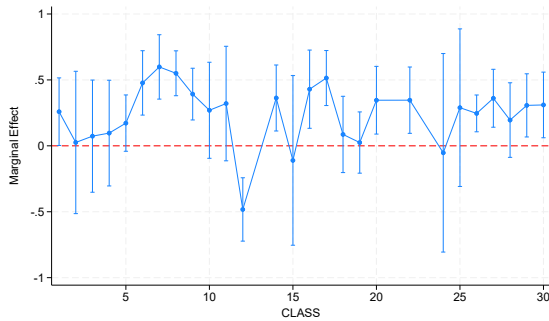
Panel B. ≥ 1 Visit



Panel E. ≥ 1 Visit



Panel C. ≥ 4 Visit



Panel F. ≥ 4 Visit

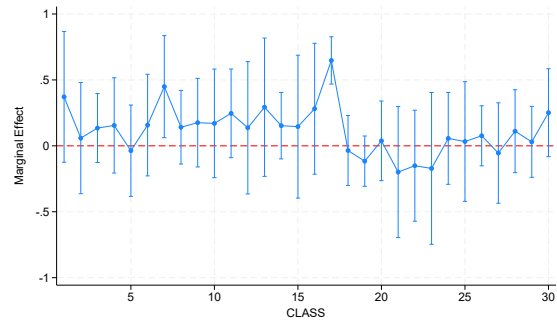


Figure 1: Average marginal effect for each class

Notes: The figure displays the average marginal effect for the number of study room visits, ≥ 1 visit, and ≥ 4 visits, along with 95 percent confidence bands. The regression is based on Table 4 columns (1)-(3), the four treatment conditions with standard errors are clustered at the group level.

of courses, no pattern can be detected. However, there are exceptions, such as class 7, 16, and 17, in which the AME is significantly higher in both groups (≈ 2). Unfortunately, the data does not reveal which courses they represent, which hinders drawing specific conclusions. Besides this, there are some courses where the outcome is only higher for column A or B.

This shows that there is a possibility that the academic degree and the courses taken can influence one or both treatment effects. However, the results must be taken with caution, as the sample size of most classes is very small (for 50 percent under 30). It is therefore not surprising that the majority of the coefficients are not significant and that there is a large fluctuation. Further analyses with larger samples would be necessary to gain a more precise understanding into these tendencies.

5 Conclusion

In this reproduction report, we evaluated the reproducibility and robustness of the study "Letting Down the Team? Social Effects of Team Incentives" by Philip Babcock and colleagues (2015). Apart from one table, we were able to fully reproduce the findings from the paper. The outputs from our reproduction match the results of original paper to a large extent, and only single values deviate slightly.

We conducted one data manipulation check by removing missing values, as well as participants over 26 and participants with more than 6 visits to the study room. In both specifications, the results are very similar to the published results. A heterogeneity test for gender and age further confirms the results as the team treatment did lead to more study room visits than the individual treatment for all four subgroups. Moreover, by looking at the subsample of participants who have visited the study room at least once, we could show that team treatment effect from Table 4 can be explained by more participants visiting the study room at least once, compared to the individual treatment. Finally, the average marginal treatment effect for each class illustrates that it might be possible that certain classes are better suited for incentivized studying. However, the majority of the coefficients are not significant and there are large fluctuations, which can be attributed to the reduced sample size as classes were rather small. Therefore, this final check must be taken with caution, and replications with larger samples would be necessary to gain a better understanding. Overall, even after the additional checks, the results of the paper remain intact.

References

- [1] Babcock, Philip; Bedard, Kelly; Charness, Gary, Hartman, John; Royer and Heather Royer (2015): Letting Down the Team? Social Effects of Team Incentives. *Journal of the European Economic Association*, 13(5), 841–870.

A Appendix

Table 3: Pre-treatment descriptive statistics for study room rounds

	Control group mean (1)	Individual treatment (2)	Team treatment (3)	Anonymous treatment (4)	Choice treatment (5)
Male	0.55 (0.50)	-0.09 [0.09]	-0.03 [0.52]	-0.05 [0.47]	-0.05 [0.44]
Age	20.57 (2.62)	-0.19 [0.34]	-0.25 [0.14]	0.08 [0.81]	-0.35 [0.17]
Library days	1.88 (0.05)	0.30 [0.16]	0.18 [0.33]	0.19 [0.46]	0.16 [0.56]
Sample size	160	204	398	162	169

Notes: Differences in columns (2)-(5) are from OLS regressions that include class-year fixed effects. Regression samples in columns (2)-(5) include the control group and the group listed at the top of each column. Standard deviations for the control group are in parentheses; p-values for differences are in square brackets. Sample size refers to the group listed in the column heading. Samples sizes vary slightly across variables due to nonreporting.

Source: own illustration (replication of Table 1).

Table 4: Study room visits by treatment status

	Number of visits (1)	≥ 1 Visit (2)	≥ 4 Visit (3)	Number of visits (4)	≥ 1 Visit (5)	≥ 4 Visit (6)
Any treatment	0.900*** (0.211)	0.176*** (0.051)	0.243*** (0.045)	0.842*** (0.212)	0.170*** (0.051)	0.242*** (0.045)
Team treatment	0.397** (0.197)	0.109** (0.044)	0.069 (0.044)	0.428** (0.193)	0.110** (0.044)	0.072 (0.044)
Anonymous team treatment	-0.579** (0.238)	-0.088 (0.055)	-0.133** (0.053)	-0.612*** (0.234)	-0.102* (0.055)	-0.147*** (0.052)
Choice treatment	0.542** (0.234)	0.127** (0.054)	0.103* (0.053)	0.550** (0.233)	0.121** (0.055)	0.095* (0.054)
Male				-0.116 (0.138)	-0.026 (0.032)	-0.022 (0.031)
Age 20				-0.122 (0.194)	-0.040 (0.047)	-0.037 (0.045)
Age 21				-0.133 (0.209)	-0.018 (0.050)	-0.032 (0.050)
Age 22				-0.326 (0.269)	-0.047 (0.062)	-0.132** (0.058)
Age 23+				-0.045 (0.274)	-0.030 (0.062)	-0.025 (0.064)
Library days				0.162*** (0.036)	0.033*** (0.008)	0.020** (0.008)
Year is 2013	-0.637*** (0.167)	-0.129*** (0.037)	-0.114*** (0.037)	(.)	(.)	(.)
Constant	1.432*** (0.176)	0.406*** (0.043)	0.217*** (0.036)			
Includes year-specific class fixed effects	No	No	No	Yes	Yes	Yes
R-squared	0.06	0.05	0.06	0.12	0.11	0.11

Cost per visit		Average visit per participant
Control group	\$1.77	1.09
Individual treatment	\$6.86	1.95
Team treatment	\$1.85	2.37
Anonymous team	\$2.90	1.75
Choice treatment	\$6.65	2.24
Total	\$3.67	1.99

Notes: Sample size is 1093. Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses. Columns (4)-(6) also include indicators for missing age and sex.

*** Significant at 1%; ** significant at 5%; * significant at 10%

Source: own illustration (replication of Table 2)

Table 5: Original Table 2

	Number of visits (1)	≥1 Visit (2)	≥4 Visits (3)	Number of visits (4)	≥1 Visit (5)	≥4 Visits (6)
Any treatment	0.900** (0.211)	0.176** (0.051)	0.243** (0.045)	0.842** (0.212)	0.170** (0.051)	0.242** (0.045)
Team treatment	0.397** (0.197)	0.109** (0.044)	0.069 (0.044)	0.428** (0.193)	0.110** (0.044)	0.072 (0.044)
Anonymous team treatment	-0.579** (0.238)	-0.088 (0.055)	-0.133** (0.053)	-0.612** (0.234)	-0.102* (0.055)	-0.147** (0.052)
Choice treatment	0.542** (0.167)	0.127** (0.054)	0.103* (0.053)	0.550** (0.237)	0.121** (0.055)	0.095* (0.054)
Male				-0.116 (0.138)	-0.026 (0.032)	-0.022 (0.031)
Age 20				-0.122 (0.194)	-0.040 (0.047)	-0.037 (0.045)
Age 21				-0.133 (0.209)	-0.018 (0.050)	-0.032 (0.050)
Age 22				-0.326 (0.269)	-0.047 (0.062)	-0.132 (0.058)
Age 23+				-0.045 (0.274)	-0.030 (0.062)	-0.025 (0.064)
Library days				0.162** (0.036)	0.033** (0.008)	0.020** (0.008)
Year is 2013	-0.637**	-0.129**	-0.114**	- (0.167)	- (0.037)	- (0.037)
Constant	1.432**	0.406**	0.217**	- (0.176)	- (0.043)	- (0.036)
Includes year-specific class fixed effects	No	No	No	Yes	Yes	Yes
R-squared	0.06	0.05	0.06	0.12	0.11	0.11
Cost per visit						
	Control group	Individual treatment	Team treatment	Anonymous team	Choice treatment	
Average visits per participant	1.09	1.95	2.37	1.75	2.23	
Cost per visit	\$1.77	\$6.86	\$4.87	\$2.90	\$6.65	

Notes: Sample size is 1093. Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses. Columns (4)–(6) also include indicators for missing age and sex.

**Significant at 5%; *significant at 10%.

Table 6: Study room visits by treatment status 2011

	Number of visits (1)	≥ 1 Visit (2)	≥ 4 Visit (3)	Number of visits (4)	≥ 1 Visit (5)	≥ 4 Visit (6)
Any treatment	0.886** (0.347)	0.173** (0.079)	0.216*** (0.071)	0.768** (0.352)	0.149* (0.081)	0.199*** (0.073)
Team treatment	0.606* (0.317)	0.125* (0.067)	0.167** (0.068)	0.740** (0.333)	0.148** (0.072)	0.186*** (0.071)
Anonymous team treatment	-0.466 (0.295)	-0.079 (0.065)	-0.075 (0.064)	-0.444 (0.299)	-0.084 (0.070)	-0.079 (0.065)
Male				-0.153 (0.204)	-0.069 (0.049)	-0.044 (0.047)
Age 20				-0.382 (0.397)	-0.123 (0.092)	-0.029 (0.096)
Age 21				-0.451 (0.411)	-0.142 (0.097)	-0.017 (0.098)
Age 22				-0.792 (0.503)	-0.154 (0.115)	-0.173 (0.112)
Age 23+				-0.394 (0.494)	-0.111 (0.110)	-0.030 (0.117)
Library days				0.164** (0.065)	0.032** (0.013)	0.012 (0.014)
Constant	1.333*** (0.235)	0.400*** (0.057)	0.187*** (0.045)			
R-squared	0.06	0.05	0.08	0.12	0.11	0.13

Notes: Sample size is 491. Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses.

*** Significant at 1%; ** significant at 5%; * significant at 10%

Source: own illustration (replication of Table.A4).

Table 7: Study room structural estimates

	Coefficient	95 percentile lower bound	95 percentile upper bound
Panel A: Structural parameters			
Male	-0.001	-0.201	0.199
Age 20	-0.178	-0.464	0.107
Age 21	0.019	-0.260	0.298
Age 22	-0.287	-0.656	0.081
Age 23+	0.083	-0.305	0.471
Library days	0.085	0.033	0.136
Predicted partner bonus status	1.852	1.327	2.376
Individual treatment indicator	0.681	0.401	0.962
Year is 2013	-0.169	-0.390	0.052
Constant	-0.967	-1.314	-0.621
θ	1.717	1.092	2.756
Panel B: θ under different probability of completion assumption			
θ : unconditional team treatment mean	1.701	1.169	2.603
θ : perfect foresight	1.567	0.701	4.399
Panel C: Anonymous treatment, θ under different probability of completion assumption			
θ unconditional anonymous treatment mean	0.995	-0.072	2.208
θ unconditional team treatment mean	0.382	-0.357	1.223

Notes: 1000 bootstrap replications. The sample in Panels A and B excludes anonymous and choice treatment group members and individuals with missing data for themselves or their partners. Panels A and B sample sizes are 745. The sample in Panel C excludes team and choice treatment group members and individuals with missing data for themselves or their partners. Panel C sample size is 517.

Source: own illustration (replication of Table 3)

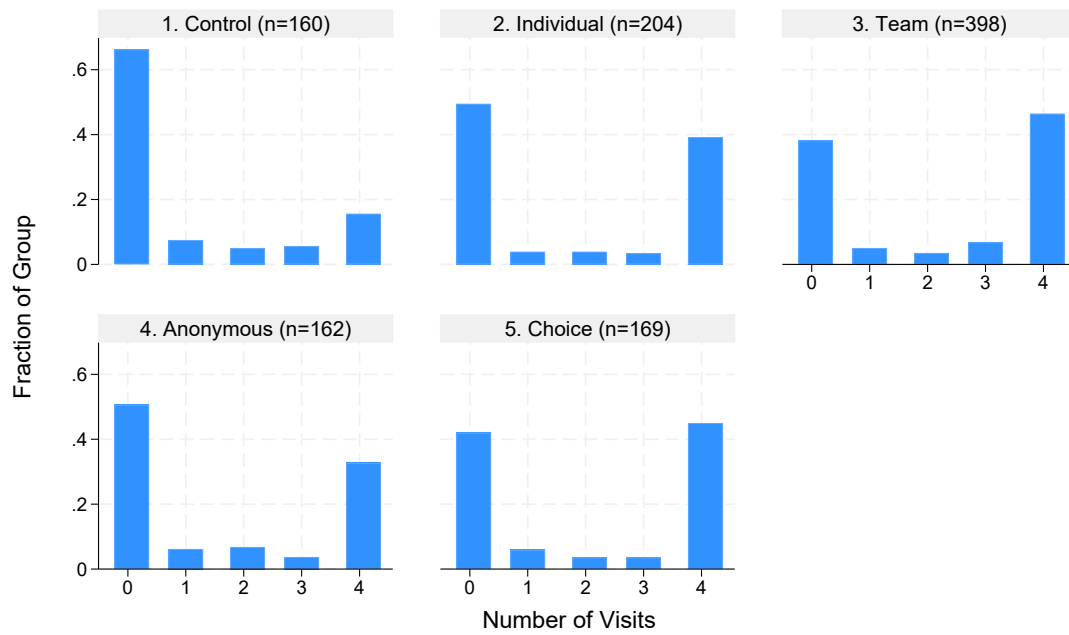


Figure 2: Distribution of study room visits

Notes: The bars labeled 4 visits include all participants with four or more visits.
Source: own illustration (replication of Figure 1)

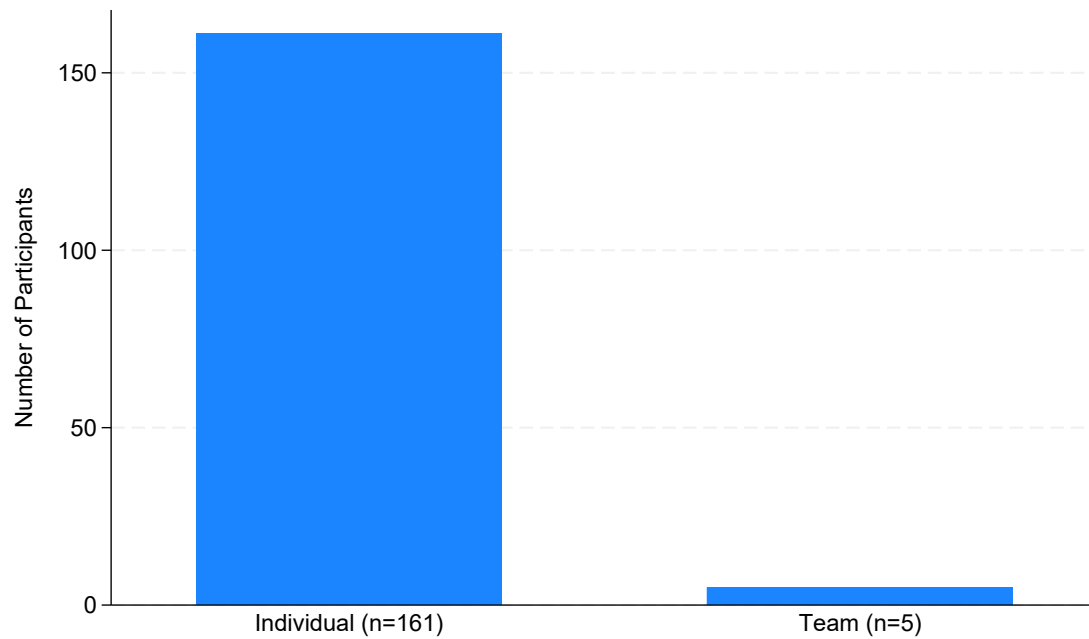


Figure 3: Distribution of choice treatment options

Notes: Includes 166 participants randomized into choice treatment; excludes the three participants moved from a control group to serve as partners. The odd numbers are due to a matching error; in one instance two choice subjects opting for the team treatment were accidentally paired. All results are the same if they are excluded from the sample.

Source: own illustration (replication of Figure 2)

Table 8: Pre-treatment descriptive statistics for Rec Center round

	Control group mean (1)	Individual treatment (2)	Team treatment (3)
Male	0.55 (0.50)	0.01 [0.86]	0.14 [0.02]
Age	21.14 (2.43)	0.22 [0.52]	0.04 [0.88]
Self-reported days of exercise/week	3.98 (2.29)	0.13 [0.73]	0.21 [0.51]
Gym visits in the previous week	1.17 (1.65)	-0.02 [0.93]	-0.13 [0.49]
Sample size	87	87	190

Notes: Differences are from OLS regressions that include class fixed effects. Regression samples in columns (2) and (3) include the control group and the group listed at the top of each column. Standard deviations are in parentheses; p-values for differences in square brackets. Sample size refers to the group listed in the column heading.

Source: own illustration (replication of Table 5).

Table 9: Rec Center visits for individual and team treatments

	Number of visits (1)	≥ 1 Visit (2)	≥ 4 Visit (3)	Number of visits (4)	≥ 1 Visit (5)	≥ 4 Visit (6)
Any treatment	1.690*** (0.471)	0.161** (0.074)	0.379*** (0.067)	1.806*** (0.365)	0.178*** (0.064)	0.395*** (0.059)
Team treatment	0.347 (0.412)	0.143** (0.059)	0.011 (0.067)	0.288 (0.340)	0.135** (0.054)	-0.014 (0.061)
Male				0.319 (0.259)	0.056 (0.044)	0.081* (0.048)
Age 20				-0.724 (0.440)	-0.121* (0.062)	-0.111 (0.076)
Age 21				-0.556 (0.490)	-0.077 (0.069)	-0.124 (0.083)
Age 22				-1.208** (0.611)	-0.210** (0.082)	-0.224** (0.097)
Age 23+				-1.704*** (0.528)	-0.279*** (0.083)	-0.317*** (0.092)
Pre-period Rec Center visits				0.938*** (0.097)	0.102*** (0.012)	0.091*** (0.015)
Constant	2.126*** (0.320)	0.517*** (0.054)	0.172*** (0.041)			
Includes class fixed effects	No	No	No	Yes	Yes	Yes
R-squared	0.08	0.08	0.11	0.39	0.30	0.27
Cost per visit						
	Cost per visit	Average visit per participant				
Control Group	3.56	2.13				
Individual Treatment	5.24	3.82				
Team Treatment	3.89	4.16				
Total	4.13	3.59				

Notes: Sample size is 364. Standard errors are clustered at the group level (i.e., in case of team treatment, the pair level) and reported in parentheses. Columns (4)-(6) also include indicators for missing age.

*** Significant at 1%; ** significant at 5%; * significant at 10%

Source: own illustration (replication of Table 6)

Table 10: Rec Center visits for individual and team treatments: restricted samples.

	Visits (1)	≥ 1 Visit (2)	≥ 4 Visit (3)	Visits (4)	≥ 1 Visit (5)	≥ 4 Visit (6)	Sample size
Panel A							364
Any treatment	1.690*** (0.471)	0.161** (0.074)	0.379*** (0.067)	1.806*** (0.365)	0.178*** (0.064)	0.395*** (0.059)	
Team treatment	0.347 (0.412)	0.143** (0.059)	0.011 (0.067)	0.288 (0.340)	0.135** (0.054)	-0.014 (0.061)	
Panel B: Sample restricted to actives							164
Any treatment	1.759** (0.679)	0.023 (0.059)	0.409*** (0.106)	1.862*** (0.625)	0.031 (0.061)	0.443*** (0.099)	
Team treatment	-0.883 (0.539)	-0.022 (0.050)	-0.122 (0.089)	-0.634 (0.502)	-0.020 (0.045)	-0.129 (0.092)	
Panel C: Sample restricted to inactives							200
Any treatment	1.782*** (0.444)	0.286*** (0.091)	0.372*** (0.072)	1.811*** (0.428)	0.308*** (0.090)	0.380*** (0.072)	
Team treatment	1.047** (0.486)	0.240*** (0.086)	0.088 (0.088)	0.711 (0.464)	0.188** (0.084)	0.036 (0.086)	
Panel D: Sample restricted to actives							164
Any treatment	1.759** (0.682)	0.023 (0.059)	0.409*** (0.106)	1.876*** (0.628)	0.030 (0.061)	0.443*** (0.099)	
Team treatment: Inactive partner	-0.887 (0.667)	-0.033 (0.063)	-0.160 (0.110)	-0.400 (0.603)	-0.029 (0.061)	-0.133 (0.110)	
Team treatment: Active partner	-0.881 (0.584)	-0.016 (0.057)	-0.099 (0.102)	-0.797 (0.530)	-0.014 (0.050)	-0.127 (0.106)	
Panel E: Sample restricted to inactives							200
Any treatment	1.782*** (0.445)	0.286*** (0.092)	0.372*** (0.072)	1.840*** (0.427)	0.309*** (0.091)	0.390*** (0.072)	
Team treatment: Inactive partner	0.818 (0.542)	0.252*** (0.094)	0.002 (0.097)	0.450 (0.480)	0.183** (0.089)	-0.055 (0.089)	
Team treatment: Active partner	1.490** (0.612)	0.216** (0.106)	0.255** (0.108)	1.235* (0.651)	0.199* (0.109)	0.217* (0.115)	
Includes control variables	No	No	No	Yes	Yes	Yes	
Includes class fixed effects	No	No	No	Yes	Yes	Yes	
R-squared	0.08	0.08	0.11	0.39	0.30	0.27	

Notes: Sample size is 364. Standard errors are clustered at the group level (i.e., in case of team treatment, the pairlevel) and reported in parentheses. Columns (4)-(6) also include indicators for missing age.

*** Significant at 1%; ** significant at 5%; * significant at 10%

Source: own illustration (replication of Table 7)

Table 11: Rec Center structural estimates

	Coefficient	95 percentile lower bound	95 percentile upper bound
Panel A: Structural parameters			
Male	0.275	-0.051	0.601
Age 20	-0.292	-0.810	0.226
Age 21	-0.357	-0.850	0.136
Age 22	-0.613	-1.269	0.043
Age 23+	-0.913	-1.493	-0.333
Pre-period Rec Center visits	0.293	0.187	0.399
Predicted partner bonus status	2.168	1.483	2.854
Individual treatment indicator	1.291	0.869	1.714
Constant	-1.152	-1.651	-0.653
θ	0.679	0.284	1.209
Panel B: θ under different probability of completion assumptions			
θ : unconditional team treatment mean	0.699	0.376	1.141
θ : perfect foresight (true partner visits)	0.169	-0.216	0.937

Notes: 1,000 bootstrap replications. Sample excludes individuals with missing data for themselves or their partners. The sample size is 362.

Source: own illustration (replication of Table 8)

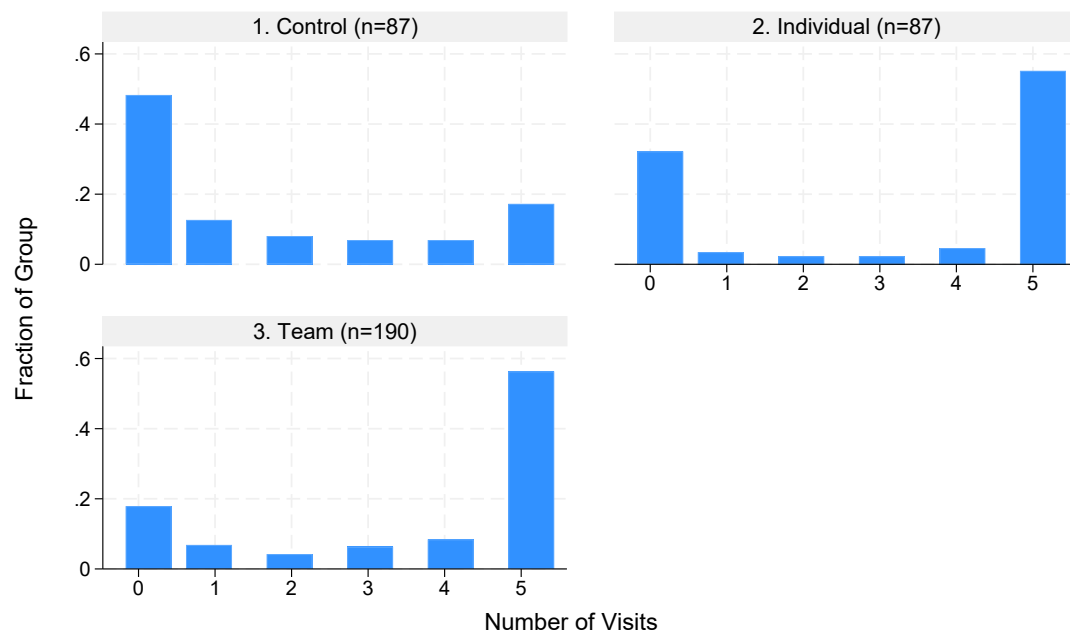


Figure 4: Distribution of Rec Center visits

Notes: The bars labeled 5 visits include all participants with five or more visits.

Source: own illustration (replication of Figure 3)