Journal of the Economic Science Association (2024) 10:152–164
https://doi.org/10.1007/s40881-023-00151-5

REPLICATION PAPER

The Preference Survey Module: evidence on social preferences from Tehran

Michael Kosfeld[1] · **Zahra Sharafi**[1]

Received: 27 January 2022 / Revised: 30 August 2023 / Accepted: 15 September 2023 /
Published online: 19 October 2023

Abstract

We provide evidence on the extent to which survey items in the Preference Survey Module and the resulting Global Preference Survey measuring social preferences—trust, altruism, positive and negative reciprocity—predict behavior in corresponding experimental games outside the original participant sample of Falk et al. (Manag Sci, 2022. https://doi.org/10.1287/mnsc.2022.4455). Our results, which are based on a replication study with university students in Tehran, Iran, are mixed. While quantitative items considering hypothetical versions of the experimental games correlate significantly and economically meaningfully with individual behavior, none of the qualitative items show significant correlations. The only exception is altruism where results correspond more closely to the original findings.

Keywords Preference Survey Module · Global Preference Survey · Validation · Replication · Social preferences

JEL Classification C81 · C83 · C90 · D01 · D03

1 Introduction

Falk et al. (2022) develop the "Preference Survey Module" (PSM) to measure important economic preferences such as trust, altruism, positive and negative reciprocity, risk preferences, and time discounting. The module identifies survey items that are able to predict individual behavior in incentivized experimental economic games such as the trust game, dictator game, ultimatum game, etc. It serves as an important

✉ Michael Kosfeld
kosfeld@econ.uni-frankfurt.de

Zahra Sharafi
sharafi@econ.uni-frankfurt.de

[1] Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany

basis for the subsequently developed "Global Preference Survey" (GPS) providing a novel, comparable set of representative data on the above-mentioned economic preferences from more than 70 countries across the globe (Falk et al., 2018).[1]

Results in Falk et al. (2018) demonstrate the validity of the utilized preference measures by convincingly documenting—both across and within countries—significant correlations with a range of relevant economic outcomes in line with economic theory. Still, it remains an open question to what extent PSM and GPS measures predict behavior in experimental games in *other participant samples* than the one used in the original validation.[2] Bauer et al. (2020) provide a first replication analysis for a subset of preference measures with participants from low-income households in Kenya. They find that quantitative survey items, which are based on hypothetical, i.e., non-incentivized games, are generally good predictors of participants' choices in the experiments, while qualitative items based on self-assessments in real-world scenarios do not correlate significantly with experimental behavior.

The present paper contributes to these analyses by providing results from a replication study we implemented with student participants in Tehran in April 2018. We chose Tehran for two main reasons: first, we wanted to validate the survey modules in a country that has sufficient contrast to Germany, where the original experiments took place, in terms of culture, language, religion, history, geography, etc. Tehran as a capital city of a country located in the Middle East fits very well to this condition. Second, we had access to a lab at the University of Tehran that resembles the lab at the University of Bonn well regarding its student sample. This helped us besides the new country to remain close to the original study in important other aspects. However, our access to the lab was limited, which led us to restrict the sample size and the range of preferences to be included in the present study. Our data come from laboratory experiments with 102 students at the University of Tehran and we focus on social preferences. In consequence, we only included survey items and experiments from Falk et al. (2022) that aim at measuring these preferences.

Overall, our results are mixed and, by and large, complementary to the findings of Bauer et al. (2020). While for altruism we find that PSM items are significantly correlated with behavior, for the other social preference dimensions we find that correlations between survey items and behavior in the corresponding games are often low and insignificant. In general and very similar to Bauer et al. (2020), quantitative items that elicit behavior in hypothetical games perform better than qualitative items asking for self-assessments or behavior in real-world scenarios. When applying a comparable item selection procedure as in Falk et al. (2022) to identify the survey items that best predict individual behavior in the experimental games in Tehran, we never identify any of the qualitative items included in the PSM but almost always identify the quantitative item.

Our replication study differs from Bauer et al. (2020) in several dimensions. Firstly, our participant pool is different: university students in Iran v. low-income

[1] See also https://www.briq-institute.org/global-preferences/home.

[2] The experiments in Falk et al. (2022) are conducted with student participants at the University of Bonn, Germany.

households in Kenya. Secondly, we include the full set of PSM/GPS items and experiments on social preferences, while Bauer et al. include only the two GPS items for each preference but consider also time discounting and risk preferences. Thirdly, our protocol stays as close as possible to Falk et al. (2022), while Bauer et al. adjust their implementation in several ways. Still, both our replication studies contain the same message: quantitative survey items predict well, while qualitative items do only poorly. We can only speculate at this point why this is the case. One possible explanation is that qualitative items are simply noisier than quantitative items. Neither our study nor Bauer et al. (2020) can rule this out definitely, as both are based on relatively small sample sizes of about 100 participants each. Future studies should therefore focus on replications that involve larger *N*. An alternative explanation is that qualitative items are more country-/culture-dependent than quantitative items, because different real life experiences feed into participants' answers. Below, we provide some evidence for this explanation in the case of positive reciprocity. Still, more research is warranted.

2 Design

Our experimental design closely follows Falk et al. (2022). In their study, about 400 students from the University of Bonn each participated in two laboratory sessions scheduled one week apart. In both sessions, subjects answered a non-incentivized survey and made decisions in incentivized experiments. Survey and experiment(s) for the same preference were never run in the same session. Half of the subjects participated in the first session in a survey related to risk and time preferences and in experiments relating to social preferences, and in the second session, it was the other way round. For the other half, the order was reversed.

We only consider social preferences in this study. 102 students from the University of Tehran participated in two laboratory sessions scheduled one week apart in April 2018. All students answered the social preference survey in the first session and participated in the corresponding experiments in the second session. Whenever available we used the Persian translation from Falk et al. (2018) for questions in the survey. We translated the remaining questions as well as the experimental instructions ourselves. The English wording of all survey items and experimental instructions are provided in the Appendix B.

Based on Falk et al. (2022), social preferences elicited in the experiments consider trust, altruism, as well as positive and negative reciprocity. Trust is measured by first-mover behavior in two different versions of the investment game (Berg et al., 1995), where the amount sent by the first mover is either doubled or tripled. Altruism is measured by a dictator game with a charitable organization as a receiver. Positive reciprocity is measured by second-mover behavior in the investment game (Berg et al., 1995) using the strategy method. Finally, the measure of negative reciprocity is based on the second-mover's minimum acceptable offer in an ultimatum game (Güth et al., 1982) as well as investment into punishment after the unilateral defection of the opponent in a prisoner's dilemma game (Falk et al., 2005). Subjects are randomly matched with each other in these games based on a perfect stranger

protocol. The order of games is fixed as follows: investment game, dictator game, ultimatum game, and prisoner's dilemma game.

All experiments in our study were programmed in oTree (Chen et al., 2016). For the survey we used Limesurvey. Earnings were calculated in points during the experiment and exchanged into Iranian Rial at the end based on an exchange rate of 100 points/10,000 Iranian Rial. The exchange rate is comparable to Falk et al. (2022) in terms of average earnings as a student assistant. On average, a session lasted 65 min and subjects earned 359,710 Iranian Rial in the experiments.

3 Results

Table A1 in the Appendix A1 provides a summary and comparison of average behavior in the social preference experiments in Falk et al. (2022) and in our sample. Overall, Tehran subjects reveal a higher level of pro-sociality with differences in positive reciprocity being statistically significant in particular. While these differences are informative and consistent with differences based on GPS measures from Falk et al. (2018), they are not our main focus in this paper. In the following, we explore the extent to which answers in the survey correlate with behavior in the experiments and whether items selected for the PSM and GPS, based on the original data, predict behavior in the corresponding games in Tehran.

To do so, we proceed as follows. For each preference, we first check whether the PSM and GPS items are correlated with behavior in the corresponding games. This answers the question, to what extent the two modules are valid instruments to predict social preferences from the incentivized experiments conducted in Tehran. Subsequently, we apply the same procedure as in Falk et al. (2022) to select the two survey items that best predict subjects' behavior in our data. The results, details of which are reported in the Appendix A2, produce a (potentially different) battery of survey questions to which the PSM and GPS can be compared.

3.1 Trust

Table 1 presents our results for trust. Panel A shows the correlations between the selected items for each module and the behavioral measure in the original study, while Panel B shows the same results for our replication study. In each panel, we report Spearman correlation coefficients and OLS coefficients obtained from a multivariate regression of the standardized behavioral measure on standardized survey items. Recall that trust is behaviorally measured by the average amount a subject sends as a first mover in two different versions of the investment game, where the amount to be sent is either doubled or tripled.

As is evident from Panel B in Table 1, the quantitative item (T24) from the PSM, which is a hypothetical version of the investment game, significantly correlates with behavior in the incentivized investment game, though coefficients are smaller compared to the original study. In contrast, the qualitative item "*People have only the*

Table 1 Correlation of PSM and GPS items with the behavioral measure of trust in the original study of Falk et al. (2022) (Panel A) and our replication study in Tehran (Panel B)

Item description	Panel A: Falk et al. (2022)		Panel B: Replication study	
	Corr	OLS	Corr	OLS
PSM				
Hypothetical first-mover behavior in the investment game (T24)	0.620***	0.629*** (0.043)	0.513***	0.339*** (0.093)
As long as I am not convinced otherwise, I assume that people have only the best intentions. (T16)	0.283***	0.133*** (0.038)	0.082	0.102 (0.093)
N		382		102
Adjusted R^2		0.452		0.112
F		158.4		7.36
GPS				
As long as I am not convinced otherwise, I assume that people have only the best intentions. (T16)	0.283***	Not reported	0.082	0.123 (0.099)
N		Not reported		102
Adjusted R^2		Not reported		0.005
F		Not reported		1.53

Item numbers are based on the numbering as summarized in the Appendix B2. The first column in each panel is the Spearman correlation between the survey item and the behavioral measure. The second column shows the OLS coefficients from a multivariate regression of the standardized behavioral measure on the standardized module items. Standard errors are in parentheses. *** and ** denote significance at the 1- and 5-percent level, respectively

Table 2 Correlation of PSM and GPS items with the behavioral measure of altruism in the original study of Falk et al. (2022) (Panel A) and our replication study in Tehran (Panel B)

Item description	Panel A: Falk et al. (2022)		Panel B: Replication study	
	Corr	OLS	Corr	OLS
PSM and GPS				
Hypothetical dictator game (A11)	0.391***	0.184*** (0.049)	0.306***	0.189* (0.097)
How do you assess your willingness to share with others without expecting anything in return in the following contexts: Charity (A10)[a]	0.384***	0.321*** (0.044)	0.192*	0.182* (0.097)
Observations		382		102
Adjusted R^2		0.175		0.058
F		41.41		4.13

Item numbers are based on the numbering as summarized in the Appendix B2. The first column in each panel is the Spearman correlation between the survey item and the behavioral measure. The second column shows the OLS coefficients from a multivariate regression of the standardized behavioral measure on the standardized module items. Standard errors are in parentheses. *** and * denote significance at the 1- and 10-percent level, respectively.

[a]In the GPS, the question is re-phrased to the willingness to give to good causes

best intentions" (T16), which is also the only trust item included in the GPS, shows no significant correlation, while the coefficient is positive but small.

Table A2 (Model 1) in the Appendix A2 presents our results for trust when applying the same item selection procedure as in Falk et al. (2022). The two survey items that explain behavior in the investment game best in the Tehran sample are the quantitative item (T24), also selected by the PSM, together with the qualitative item "*Most people would be fair to you"* (T21). Interestingly, replacing the latter item with the standard trust question from the World Value Survey "*In general, one can trust other people"* (T17), works also considerably well. Both do better than the qualitative GPS item "*People have only the best intentions"* (T16).

3.2 Altruism

Unlike the other preference dimensions, in the case of altruism, the PSM and GPS both include a quantitative and a qualitative item. Table 2 shows correlations of them with the behavioral measure of altruism based on the charitable dictator game. Again, Panel A considers the original study of Falk et al. (2022), Panel B the replication study in Tehran. Both items are significantly correlated with behavior in the replication study, at least on the 10 percent level. Similar to trust, correlations of the quantitative item (A11), which is a hypothetical version of the dictator game, are again stronger compared to the qualitative item "*Willingness to share with others in the context of charity"* (A10). In addition, coefficients of the quantitative item (A11) are remarkably close to the results of the original study (0.306 vs. 0.391, 0.189 vs.

0.184), while the coefficients of the qualitative item A11 are only about half the size of the coefficients (0.192 vs. 0.384, 0.182 vs. 0.321).

Results on the item selection procedure in Table A3 (Appendix A2) reveal that for altruism, again, the quantitative item (A11) is selected, together with the qualitative item "*Other people regard me as an unselfish person"* (A21) or alternatively, the qualitative item "*I am willing to help others even if I expect that I will never meet them again"* (A14). Thus, while the qualitative PSM/GPS item (A10) is significantly correlated with behavior in the dictator game, it does not come out as the winner from the item selection procedure.

3.3 Positive reciprocity

A similar picture emerges for positive reciprocity; see Table 3. Recall that positive reciprocity is measured by the second-mover behavior in the incentivized investment game. A hypothetical version of this game, which is what the quantitative item (PR11) captures, is again highly significantly correlated with this measure in our replication study (PSM, Panel B) although coefficients are smaller compared to the original study (Panel A). In contrast, the qualitative item "*Thank-you present in a hypothetical helping scenario"* (PR9) selected for the PSM, as well as the qualitative items "*Willingness to return a favor"* (PR12), which is additionally included in the GPS, show no significant correlation and coefficients are only very small (Panel B).

Our results from the item selection procedure for positive reciprocity reveal an intriguing finding (see Table A4 in the Appendix A2 for details). Besides the quantitative measure (PR11), which is again selected similar to what we have seen for trust and altruism, the qualitative item "*Hypothetical scenario (need medical treatment): willingness to pay for a thank-you gift*" (PR10) is found to best predict positive reciprocity in the investment game. This item is almost identical to the qualitative item in PSM (PR9) by asking the respondent to decide how much to spend on a thank-you present in return for the help received from a stranger. The only difference is in the situation the respondent is asked to imagine. In PR9, the respondent is lost in an unfamiliar city. A stranger provides help in taking the respondent by car to the desired destination. In PR10, the respondent needs medical treatment in a foreign country but does not have any cash to pay the doctor. A stranger helps by giving the corresponding amount as a gift. In both scenarios, the respondent is then asked to specify how much he or she is willing to spend on a present as a thank you to the stranger. While in the original study with student participants in Bonn responses to the lost-in-an-unfamiliar-city scenario best predict reciprocal behavior in the investment game, the same item has no predictive power with student participants in Tehran. Instead, the medical-help scenario serves as the best predictor here.

We believe a possible and quite intuitive explanation for this discrepancy is the difference in social norms between Germany and Iran with respect to the described situations, leading to different perceptions of the level of kindness that the help received by the stranger signals. Whereas in Germany driving a lost stranger to his or her desired destination is rather unusual and therefore likely contains a strong

Table 3 Correlation of PSM and GPS items with the behavioral measure of positive reciprocity in the original study of Falk et al. (2022) (Panel A) and our replication study in Tehran (Panel B)

Item description	Panel A: Falk et al. (2022)		Panel B: Replication study	
	Corr	OLS	Corr	OLS
PSM				
Hypothetical second-mover behavior in the investment game (PR11)	0.556***	0.486*** (0.049)	0.255***	0.286*** (0.097)
Thank-you gift in the lost-in-an-unfamiliar-city scenario (PR9)[a]	0.353***	0.164*** (0.049)	0.136	0.084 (0.097)
N		360		102
Adjusted R^2		0.329		0.078
F		89.15		5.29
GPS				
Willingness to return a favor (PR12)	0.297	Not reported	0.010	0.034 (0.100)
Thank-you gift in the lost-in-an-unfamiliar-city scenario (PR9)	0.353***	Not reported	0.136	0.131 (0.100)
N		Not reported		102
Adjusted R^2		Not reported		− 0.002
F		Not reported		0.92

Item numbers are based on the numbering as summarized in the Appendix B2. The first column in each panel is the Spearman correlation between the survey item and the behavioral measure. The second column shows the OLS coefficients from a multivariate regression of the standardized behavioral measure on the standardized module items. Standard errors are in parentheses. *** denote significance at the 1-percent level.

[a]Because bottles of wine are a very uncommon gift in Muslim societies, we replaced it by a more neutral and common gift in form of (different sizes of) a chocolate box. Falk et al. (2018) proceed in a similar way for the GPS

signal of kindness, the same behavior is not uncommon and almost expected in Iran. Consequently, the signal of kindness is much weaker, although the described behavior is exactly the same. This does not apply to the medical-help scenario, where in both countries receiving money as a gift from a stranger can be regarded as quite unexpected and therefore a clear signal of kindness.

To a certain degree, we can of course only speculate whether this explanation is correct. However, if so, i.e., if in Iran the level of kindness perceived in the medical-help scenario is indeed higher than in the lost-in-an-unfamiliar-city scenario, the average response in form of a thank-you present should be higher in the former scenario compared to the latter. This is indeed what we find: 0.76 versus 0.58 (standardized values, Wilcoxon signed rank test, $p < 0.01$). Further, if as a consequence the medical-help scenario is a better measure of positive reciprocity than the lost-in-an-unfamiliar-city scenario, the corresponding distribution of participants' answers in our sample should be closer to typical distributions of positive reciprocity, which are found to be left-skewed (Dohmen et al., 2009; Falk et al., 2018). We find this as well. While the distribution of answers to the lost-in-an-unfamiliar-city scenario is symmetric, the distribution in the medical-help scenario is skewed to the left. See the Appendix A4 for details.

The suggested explanation highlights in our opinion an important challenge for the analysis of cross-country/-cultural differences based on answers to survey items that are framed in real-world scenarios. Participants from different countries and cultural backgrounds may perceive and assess the same question or situation very differently, one important reason being that social norms governing the described scenario differ between the respective countries and cultures. In this respect, questions about behavior in an abstract game—even if non-incentivized—seem to have a powerful advantage.

3.4 Negative reciprocity

We finally come to negative reciprocity. Table 4 contains the results. The behavioral measure for negative reciprocity is the average score obtained from the minimum acceptable offer in an ultimatum game and the amount invested into punishment in a prisoner's dilemma game after unilateral defection of the opponent (cf. Appendix A2 for details). Panel B shows the performance of the proposed survey items in the replication study. Neither the quantitative item (NR9), which is the minimum acceptable offer in a hypothetical ultimatum game nor the qualitative item "*Willingness to punish unfair behavior*" (NR1), are significantly correlated with the behavioral measure and the coefficients are also tiny. Similarly, for the qualitative items that were eventually included in the GPS, we find no significant correlation with negatively reciprocal behavior in the Tehran data. Table 5 shows that results do not change, if we repeat the analysis for both games separately.

Results on the item selection procedure in Table A5 (Appendix A2) show that also for this preference, the survey items that explain behavior in the replication study have a quantitative question "*Hypothetical scenario: willingness to spend on hiring the detective to find and punish an unfair person"* (NR10). The qualitative

Table 4 Correlation of PSM and GPS items with the behavioral measure of negative reciprocity in the original study of Falk et al. (2022) (Panel A) and our replication study in Tehran (Panel B)

Item description	Panel A: Falk et al. (2022)		Panel B: Replication study	
	Corr	OLS	Corr	OLS
PSM				
Minimum acceptable offer in the hypothetical ultimatum game (NR9)	0.342***	0.328*** (0.059)	0.064	0.044 (0.075)
How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly (NR1)	0.161***	0.148*** (0.059)	−0.017	−0.006 (0.075)
N		360		102
Adjusted R^2		0.134		−0.016
F		28.83		0.18
GPS				
If I suffer a serious wrong I will take revenge at the first opportunity. (N11)	0.110	Not reported	0.05	0.036
Willingness to punish unfair behavior towards self	Not reported	Not reported	–	–
Willingness to punish unfair behavior towards others	Not reported	Not reported	–	–
N		Not reported		102
Adjusted R^2		Not reported		−0.007
F		Not reported		0.24

Item numbers are based on the numbering as summarized in the Appendix B2. The first column in each panel is the Spearman correlation between the survey item and the behavioral measure. The second column shows the OLS coefficients from a multivariate regression of the standardized behavioral measure on the standardized module items. Standard errors are in parentheses. *** denote significance at the 1-percent level.

Table 5 Correlation of the PSM items with behavior in the Ultimatum and Prisoners' Dilemma game separately (Replication study only)

Item description	Ultimatum game		Prisoner's Dilemma	
	Corr	OLS	Corr	OLS
Minimum acceptable offer in the hypothetical ultimatum game (NR9)	0.128	0.149 (0.098)	−0.061	−0.061 (0.100)
How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly (NR1)	−0.123	−0.115 (0.099)	0.232**	0.102 (0.100)
N		102		102
Adjusted R^2		0.017		−0.005
F		1.89		0.73

Item numbers are based on the numbering as summarized in the Appendix B2. The first column in each panel is the Spearman correlation between the survey item and the behavioral measure. The second column shows the OLS coefficients from a multivariate regression of the standardized behavioral measure on the standardized module items. Standard errors are in parentheses. ** denote significance at the 5-percent level

question "*If someone behaves unfairly towards me in sports, I will also behave unfairly towards them*" (NR16) has a clear link to negative reciprocity.

Overall, quantitative items seem to perform consistently better than qualitative items in predicting individual behavior in our replication study. The adjusted R^2s are small, indicating a low predictive power of the PSM and GPS modules for the observed behavior in general.

4 Conclusion

We report results from a replication study with university students in Tehran aimed at testing to what degree survey items selected for the Preference Survey Module (PSM) as well as the Global Preference Survey (GPS) predict behavior in corresponding games in a different country than where the PSM was originally developed. Our results show that except for altruism, where both the quantitative and the qualitative survey item are found to significantly predict behavior, at least on the 10-percent level, for the other social preference dimensions only the quantitative item, which consider hypothetical versions of the corresponding experimental game, reveal a statistically significant and economically meaningful correlation. For negative reciprocity, neither the qualitative nor the quantitative items are significantly correlated with behavior. As none of the quantitative items—except for altruism—are included in the GPS, the results cast doubt on a straightforward generalization of correlations between survey items in the PSM and GPS and behavior in corresponding experimental games to other countries.

Our study corroborates findings from Bauer et al. (2020), despite several differences in the experimental implementation between their and our replication (participants, measures, protocol): Quantitative items seem to predict well across countries and cultures; qualitative items do not, or less so. The result is important, because it suggests that quantitative items that are based on questions about behavior in abstract games—even if hypothetical, i.e., non-incentivized—may have a powerful comparative advantage in cross-country analyses.

Obviously, we need more comprehensive evidence before jumping to conclusions. Firstly, the number of participants in both our study and in Bauer et al. (2020) is relatively small.[3] This implies that noise may play a role, in particular with respect to the observed differences between quantitative and qualitative survey items. Secondly, existing replications consider only two countries, Iran and Kenya. Thirdly, our study focuses on social preferences only and hence cannot say much about the predictive power of PSM and GPS items for risk and time preferences. All three issues can be resolved in future research. Then, the aim should be to not only test whether the results from Falk et al. (2022) can be replicated in full in other countries—it would be surprising if they could—but to broaden the research agenda towards a comprehensive validation and development of truly global measures of economic preferences.

[3] Except for negative reciprocity, however, a power analysis suggests that the sample size in our study is sufficient to replicate the original findings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40881-023-00151-5.

Acknowledgements We are grateful to Anke Becker, Thomas Dohmen, Andreas Grunewald, Ferdinand von Siemens as well as seminar participants at the FLEX Grüneburgseminar and the MM Brown Bag seminar at Goethe University Frankfurt for helpful comments. We also thank two anonymous reviewers for their very useful feedback. We are particularly grateful to Rosa Hendijani for her support in using the University of Tehran lab. IRB approval was received from the Ethics Committee of the Faculty of Economics and Business at Goethe University Frankfurt.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The replication material for the study is available at https://doi.org/10.17605/OSF.IO/459KW.

Declarations

Conflict of interest The authors declare no competing financial or non-financial interests.

References

Bauer, M., Chytilová, J., & Miguel, E. (2020). Using survey questions to measure preferences: Lessons from an experimental validation in Kenya. *European Economic Review, 127*, 103493. https://doi.org/10.1016/j.euroecorev.2020.103493

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Chen, D. L., Schonger, M., & Wickens, C. (2016). OTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2009). Homo reciprocans: Survey evidence on behavioural outcomes. *The Economic Journal, 119*(536), 592–612. https://doi.org/10.1111/j.1468-0297.2008.02242.x

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences*. *The Quarterly Journal of Economics, 133*(4), 1645–1692. https://doi.org/10.1093/qje/qjy013

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2022). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*. https://doi.org/10.1287/mnsc.2022.4455

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica, 73*(6), 2017–2030. https://doi.org/10.1111/j.1468-0262.2005.00644.x

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7