Journal of the Economic Science Association (2022) 8:34–44
https://doi.org/10.1007/s40881-022-00120-4

ORIGINAL PAPER

Incentivization matters: a meta-perspective on dictator games

Philip D. Grech[1] · **Heinrich H. Nax**[2,3] · **Adrian Soos**[4]

Received: 10 August 2020 / Revised: 13 September 2022 / Accepted: 18 September 2022 /
Published online: 11 October 2022
© The Author(s) 2022

Abstract

Dictator game experiments come in three flavors: plain vanilla with strictly dichotomous separation of dictator and recipient roles, an interactive alternative whereby every subject acts in both roles, and a variant thereof with role uncertainty. We add information regarding which of these three protocols was used to data from the leading meta-study by Engel (Exp Econ 14(4):583–610, 2011) and investigate how these variations matter. Our meta-regressions suggest that interactive protocols with role duality compared with standard protocols, in addition to being relevant as a control for other effects, render subjects' giving less generous but more efficiency-oriented. Our results help organize existing findings in the field and indicate sources of confounds.

Keywords Dictator game · Experiments · Meta Study · Methodology

JEL Classification C24 · C91 · D03

While some of the content we present here corrects results reported in Engel (2011), we are first and foremost thankful to Christoph Engel for his original study, for supplying us with his meta-study dataset and analysis files, and for helping us to make data and analyses publicly available. We are also grateful for helpful comments from the editors, Maria Bigoni and Dirk Engelmann, and from several anonymous referees. The replication material for the study is available on the Open Science Framework at https://osf.io/xc73h/. Nax acknowledges support through the SNF Eccellenza Grant "Markets and Norms".

✉ Heinrich H. Nax
heinrich.nax@uzh.ch

Philip D. Grech
pgrech@ethz.ch

[1] Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland

[2] Behavioral Game Theory, ETH Zurich, Zurich, Switzerland

[3] Institute of Sociology, University of Zurich, Zurich, Switzerland

[4] Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

1 Motivation

The economic study on altruistic giving has relied heavily on lab experiments labelled as 'dictator games', where one subject is given an endowment (money, tokens, etc.) to be shared with someone else. What has received little attention so far, however, is that different experimental protocols are being used to implement this game (see Fig. 1), namely

'standard' protocols (introduced by Kahneman et al. (1986) and Forsythe et al. (1994)) where only half of the subjects acting as the givers make decisions and the other half is entirely passive,

'interactive' protocols (introduced by Andreoni and Miller (2002)) where every subject acts both in the role of dictator and recipient, and thus eventually earns two payments from keeping and receiving at the same time, and

'role uncertainty' protocols, a hybrid of the standard and interactive protocols, whereby all subjects submit dictator decisions ex ante but only half of them are randomly picked eventually and paid out with a randomly paired subject whose decisions do not matter for final payments.[1]

Until recently, differences between these protocols have tended to remain unacknowledged by experimentalists. Corroborating this point is the fact that in the meta-study on dictator games by Engel (2011), which remains one of the standard references on the subject, there are no variables to control for such protocol differences. To gain a better understanding of whether and how protocol differences matter, we take a fresh look at Engel's meta-regression by adding additional controls

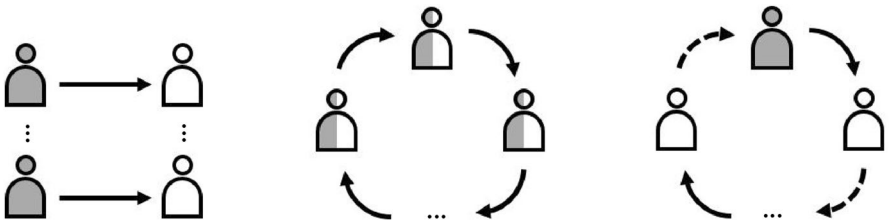


Fig. 1 Three flavors of experimental dictator game implementations. Left: standard—half of the subjects are dictators (shaded). Middle: interactive—all subjects give and receive at the same time. Right: role uncertainty—while all subjects make dictator decisions, only for half of those, these decisions are carried out in terms of actual cash payments

[1] Similar experiments are also run with Social Value Orientation (SVO) measures; see Murphy et al. (2011) for an implementation.

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

for the three different protocols that have been used under the same 'dictator games' umbrella term. We identify for each treatment included in Engel's meta-study which of the three protocols was used. Our meta-regressions suggest that giving decisions depend on the exact protocol that is being used in a way that has likely confounded earlier studies in the literature that do not account for protocol differences.

Our effort to add protocol information to Engel's meta-study is motivated by our own prior investigations in Grech and Nax (2020) regarding the different strategic incentives that different protocols create, in particular by the discovery that standard versus interactive protocols induce starkly different rational-choice benchmarks (game-theoretic predictions). Furthermore, we have evidence that these protocol effects matter behaviorally from several dedicated experiments that we conducted. In Grech and Nax (2020), for an online convenience sample recruited in that study, the interactive protocol resulted in different distributions of giving decisions and higher giving overall. In that experiment, giving multipliers (that determine how much every unit of giving is worth in the hands of the recipient) of the dictator games that were run ranged from 0.1 to 2, which allows some investigation of subjects' efficiency preferences. We found that giving under the interactive protocol was more sensitive to the giving multiplier in a way that made giving more efficient (i.e. giving more when it is cheap and giving less when it is expensive). Subsequently, in Nax et al. (2020) and Grech et al. (2020) we ran additional experiments with different populations and with a wider range of giving multipliers. In Nax et al. (2020), we find that subjects of a social elite in the United States give less than subjects in a non-elite sample in the interactive protocol when the giving multiplier is less than 2, but not when the giving multiplier is higher and also not in the standard protocol. In Grech et al. (2020), for a representative population sample for the United States, we find that the protocol effect interacts with gender, with women being more generous in the interactive dictator game, and men being more generous in the standard dictator game. Our own findings add to a growing experimental literature that has investigated through dedicated experiments the effects of protocol differences on giving in dictator games and related contexts; see Iriberri and Rey-Biel (2011), Rigdon and Levine (2018), Greiff et al. (2018); Eckel et al. (2020) and Andreozzi et al. (2021). In sum, all prior studies reject the hypothesis of protocol equivalence. The image that emerges instead is that dictator games run outside the standard protocol render giving behavior more efficiency-oriented (i.e. more positively correlated with the multiplier) and more sensitive to beliefs about others' giving decisions. In terms of overall levels of giving, effect size and direction depend both on protocol and on the underlying population sample (e.g. Grech and Nax (2020) and Nax et al. (2020) point in different directions for elite and non-elite samples).

In the present paper, we complement the single-study perspective from the aforementioned recent experiments with a meta-regression on past experiments. We therefore set out to check the instructions for each of the 131 papers included in Engel (2011)'s 2011 meta-study and code which protocol was used in each treatment

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

in each of those papers.[2] The meta perspective supports the overall image that protocol effects matter and helps organize the findings emerging from recent experimental investigations further: meta-regressions suggest that, on aggregate, departures from the standard protocol have rendered subjects less generous for low giving multipliers, more generous for higher giving multipliers, and generally more sensitive to the giving multiplier (via a significant positive correlation). Note that these effects are potentially different for different subject pools, as, for example differentiated by gender, age, etc. In this light, the results of several influential papers with 'big' messages concluding that systematic differences among humans as distinguished by gender, age, culture, class, etc. exist deserve some further investigation, because several of them have used non-standard protocols [e.g. interactive in the gender comparison of Andreoni and Vesterlund (2001)] or compared populations across protocols [e.g. elite under interactive compared with non-elite under standard in Fisman et al. (2015)]. Understanding whether and how exactly protocol differences matter is helpful to organize these prior findings, and indicates avenues for targeted replications and reproductions aimed at testing certain conclusions.

2 Data and models

2.1 Meta-analysis in perspective

Our focus here is on protocol differences from a meta-analytical perspective. Choosing such an approach may be considered a timely enterprise as the typical 'single-study' experimental paper is being increasingly criticized for failing to deliver broader messages that are robust to idiosyncrasies of individual experimental implementations. However, we encourage the reader to keep the following caveats pertaining to meta-analyses in mind when we subsequently describe the data set that we shall investigate and the analyses that we shall run.

Perhaps trivially, meta-analyses inherit all systematic shortcomings of the underlying studies on which they are based. Examples include potential omitted variable bias, lack of external validity, etc. Beyond this, meta-studies critically rest on additional assumptions which may be difficult to satisfy. For one, the different underlying experimental implementations need to be largely homogeneous in all relevant aspects not controlled for in the meta-analysis. The dictator game, at first sight, appears to be the ideal candidate for such an endeavor, because the game can be parametrized by very few variables and is very simple—in its basic implementation it is not even a game, strictly speaking—and therefore seems to leave little leeway for researcher degrees of freedom. Upon closer inspection, however, even for dictator games, it is hard to judge whether comparable laboratory procedures were really being respected. Potentially confounding factors range from general lab features

[2] Among the 445 treatments that enter the meta-regression, 383 are standard, 31 interactive, 24 role uncertainty, and 7 were not classifiable [i.e. the studies by Cárdenas et al. (2008) and Ashraf et al. (2006)].

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

(temperature, seating, color of the wall, etc.) to specific implementation details that are hard to measure (e.g. subtleties in framing). What is more, for the controls that *do* enter the meta-analysis, one must often make simplifying assumptions (e.g. by rendering certain variable categories coarser than in individual studies[3]) in order to compile a data set with manageable complexity. Such information loss, albeit unavoidable for practical reasons, can be difficult to track and may bias results. Further limitations for meta-regressions include issues related to non-uniform reporting of controls[4] and implicit values.[5]

2.2 Engel (2011)'s data set

We build our analysis on Engel's data from his 2011 paper. In addition to coding for every treatment in every study in that paper which protocol was used and integrating this information with Engel's original data (see all analyses and data at our repository on the Open Science Framework)[6], we identify and correct a number of mistakes in Engel's data.[7] Engel's dataset contains 620 treatments, found in 131 dictator game studies published up until 2010. The dependent variable of his meta-regression is the mean fraction of giving, which is reported in 498 of these treatments. Among those, the standard error of the mean is either directly reported in the original study (191), can be reconstructed from it (254), or is missing. This results in $N = 445$ non-missing data points for meta-regression used by Engel. For those, we corrected 286 false standard errors and 10 incorrect means (including 5 with a

[3] For example, Brañas-Garza (2006) finds substantial differences in giving when recipients are poor vs. when they are in need of a medical treatment. Both treatments are pooled in the level 'deserving recipient' in Engel's data set.

[4] This may lead to a Simpson-style paradox: assume that a certain variable x is only observed in one study and suppose that within this study, this variable's effect size is found to be strongly positive. However, if this particular study's reported means are particularly low (compared to other studies), the overall effect of x in the meta-regression may appear negative if no other variable explains the deviation. While most controls in Engel have been considered in at least 15–20 studies, thereby hopefully attenuating the problem, some have only a few times. 'Degree of uncertainty,' for example, is only considered in Andreoni et al. (2009) and Klempt and Pull (2009). In a similar vein, even when sufficiently many observations occur for a given control variable, this may not hold true for individual category levels of that variable. For example, the level 'mandatory' of the variable 'concealment' in Engel's data set only contains two observations. Engel circumvents this problem by treating this variable as continuous (this avoids the introduction of a binary variable for 'mandatory'), which is convenient technically but questionable in terms of interpretability. Similar observations hold for the control 'limited action space'.

[5] For example, the variable 'recipient endowment' is reported as equalling zero in 420 of 445 relevant observations, which seems unlikely given that subjects usually receive a show-up fee when participating in economic experiments.

[6] See https://osf.io/xc73h/.

[7] Obtaining and coding correctly relevant meta-data for such an effort is extremely time-consuming, and we would have never succeeded without Engel's generous sharing of his data and code underlying ((Engel, 2011)). We naturally cannot be certain that we have corrected all errors, especially as only some of the ones we spotted were related to errors in Engel's transcription from underlying studies, while other errors were already present in some of the underlying studies and therefore even harder to spot.

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

suspect value of 0).[8] Appendix A of our Online Supplementary Materials gives an overview of which studies are affected.[9] With these corrections, we recover a pattern according to which studies with a larger sample size have on average smaller standard errors, which was absent in the original data set (see Figure A1, Appendix A, Online SI).

In his regressions, Engel considers 24 control variables which fall broadly into three qualitatively distinct categories; population characteristics (such as age), framing (such as 'give' versus 'take' frame), and incentives (such as the specific strategic setting of the experiment but also whether it was financially incentivized or not).[10] A glossary for all individual variables is given in Appendix B of our Online Supplementary Materials. To Engel's original variables, we add controls for the three different dictator game protocols (interactive, non-interactive and role uncertainty) which fall into the category of incentive variables. One of Engel's incentive controls is 'recipient efficiency,' which is the aforementioned giving multiplier > 0 with which the given amount is multiplied determining what actually gets paid to the recipient. Recipient efficiency is of particular interest for our paper, because our prior theoretical and empirical work (Grech & Nax, 2020) predicted and identified a positive correlation with it for the interactive protocol but not for the non-interactive and role uncertainty protocols.

3 Results

Engel's main analysis is a multivariate random-effects meta-regression, which weights each study according to measurement accuracy (square of inverse standard error), and uses a restricted maximum likelihood procedure to estimate the between-study variance [ReML; (Patterson & Thompson, 1971)], as well as the Knapp–Hartung modification to estimate confidence intervals (Hartung, 1999; Hartung & Knapp, 2001a, b).[11] To ensure comparability, we have followed the same procedure using the statistical software environment R (v3.6.1).

Table 1 displays Engel's original results and three new meta-regressions that we ran. Regression (1) is the meta-regression as in Engel's paper using Engel's means and standard errors without corrections. Regression (2) is that same meta-regression run on the corrected data set. Regressions (3) and (4) are our meta-regressions (also

[8] For information at the level of individual studies and treatments, we refer the reader to our Open Science Framework repository (cf. Footnote 6).

[9] Our efforts complement prior replications of Engel by Zhang and Ortmann (2012) and Zhang and Ortmann (2014), confirming several of the issues identified therein.

[10] Engel (2011)'s list of variables excludes some prominent variables, such as gender, race or stakes, which is justified by stating that including them "would severely reduce the number of observations" ((Engel, 2011)) (footnote 22). In principle, this argument applies to the inclusion of any variable, so it is not exactly clear from this statement what the inclusion principle actually was. Note that some of our own dedicated single-study experiments are aimed at replicating findings related to some such variables, see Nax et al. (2020) for class status and Grech et al. (2020) for gender.

[11] The regression in Engel is run using Stata's 'metareg' command. A current Stata version (v16) as well as the free software environment R (v3.6.1) produce only slightly different numerical outputs, due to subtle differences in algorithmic implementation.

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

Table 1 Meta-regression

	(1) Engel	(2) Corrected standard errors	(3) + Protocol differences	(4) + Interaction effect
Limited action space	− 0.062^{+}	− 0.028	− 0.024	− 0.024
Degree of uncertainty	− 0.036	− 0.024	− 0.034	− 0.035
Incentive	− 0.010	− 0.018	− 0.034*	− 0.034*
Repeated game	− 0.064*	− 0.064***	− 0.070***	− 0.070***
Group decision	− 0.054^{+}	− 0.086*	− 0.085*	− 0.087*
Identification	0.042	0.050**	0.048*	0.048**
Social cue	0.005	− 0.027	− 0.033	− 0.035
Concealment	− 0.065*	− 0.074**	− 0.074**	− 0.075**
Double blind	− 0.024	0.006	0.005	0.007
Take option	0.067	− 0.060	− 0.065	− 0.065
Deserving recipient	0.086***	0.134***	0.124***	0.124***
Recipient earned	0.128*	0.127***	0.130***	0.129***
Recipient efficiency	0.026^{+}	0.020^{+}	0.018^{+}	0.011
Multiple recipients	0.148***	0.111**	0.115***	0.115***
Recipient endowment	− 0.173***	− 0.127***	− 0.122***	− 0.121**
Dictator earned	− 0.174***	− 0.172***	− 0.172***	− 0.173
Real money	0.025	0.021	0.015	0.014
Degree of social proximity	− 0.053***	− 0.046***	− 0.050***	− 0.049***
Student	− 0.104**	− 0.119***	− 0.116***	− 0.120***
Child	− 0.117**	− 0.103**	− 0.103**	− 0.105**
Middle age	0.001	0.000	0.000	0.002
Old age	0.336***	0.325***	0.339***	0.335***
Developing country	0.015	− 0.012	− 0.005	− 0.005
Indigenous society	− 0.009	0.002	0.003	− 0.001
Interactive			− 0.046*	− 0.138**
Role uncertainty			− 0.048^{+}	− 0.049^{+}
Non-classifiable			− 0.034	− 0.029
Recipient efficiency × interactive				0.083*
Intercept	0.416***	0.428***	0.469***	0.479***
N	445[a]	445	445	445
Adjusted R^2	0.483	0.543	0.548	0.556

Significance levels: 0.1% is coded as *** 1% as ** 5% as * and 10% as $^{+}$. In the text, we refer to coefficients significant at $^{+}$ as 'marginally significant,' and to effects with higher levels of significance as 'significant'

[a]Engel (2011) indicates $N = 603$ in his regression table, but reports the correct value in his article

run on the corrected data set) that include the protocol effects, with and without an interaction effect between recipient efficiency and interactive. Note that in each of the four meta-regressions, we adhered to the same use of categorical vs. continuous

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

variables as Engel (cf. above). For full regressions where all variables are treated as categorical, we refer the reader to Appendix C of our Online Supplementary Materials (qualitative conclusions remain unchanged).

A comparison of regressions (1) and (2) reveals that the above-mentioned data misspecifications do have an impact and change both effect sizes and p values of individual variables. By and large, however, Engel's results remain surprisingly robust except for the effect of identification.

Regression (3) is our baseline specification to analyze protocol effects. In line with theoretical predictions from Grech and Nax (2020), we find that interactive protocols lead to significantly lower mean giving. This confirms the direction of the effect that we found for the social elite and representative samples from the United States in Grech et al. (2020) and Nax et al. (2020), but is opposite to the effect in the experiment in Grech and Nax (2020) where we recruited an online convenience sample via MTurk. We also find a marginally significant effect of lower mean giving for role uncertainty. This is notable since the rational choice benchmark under role uncertainty is the same as that for the ordinary, non-interactive dictator game: players maximizing their expected utility under role uncertainty can only influence their final outcome by their own action in the case where they act as dictators. In this case, however, their decision problem is identical to a non-interactive dictator game. We point out that we find effects from protocol differences despite the relatively low number of studies following interactive or role uncertainty designs, which provides further support to our central claim that different protocols should not be used interchangeably, i.e. without acknowledging their differences.

In regression (4), which is our preferred specification, we further include an interaction effect between recipient efficiency and interactivity, motivated by our theoretical predictions in Grech and Nax (2020) according to which Nash equilibrium play is such that giving depends positively on recipient efficiency (jumping from null to full) for any subject who is at least moderately altruistic. This prediction applies to all such individuals, both efficiency-oriented and equality-oriented ones. By contrast, under the standard protocol and under role uncertainty, a negative relationship is rational for equality-oriented individuals, and a positive relationship is rational for efficiency-oriented individuals. The interaction term has a significant positive effect which suggests that it was really the interactive implementations that were driving the previously marginally significant effects of recipient efficiency. Finally, by comparing Regression (4) [or (3), for that matter] to Engel's original Regression (1) (or its corrected version, Regression (2)) we find that the previously insignificant effect of incentivization becomes significant and substantive when controls related to protocol are added.

4 Conclusion

Our meta-regressions produce results that help organize what several recent single-study papers have found: compared with standard protocols, non-standard protocols render subjects' giving on average less generous and make their giving decisions more efficiency-oriented. Protocols with role uncertainty seem to go in a similar

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

direction as the interactive protocol despite not sharing the same game-theoretic predictions. Notwithstanding the various limitations of meta-analyses—as discussed in Sect. 2.1—these findings corroborate the view that protocol variations may lead to qualitatively different results and therefore deserve more attention. Strictly speaking, the three protocol categories that we separated dictator games into are still simplifications of the actual decision-making contexts in the lab. For instance, the size of the loop in an interactive implementation varies and is often unknown to participants, and it remains to be understood how beliefs and loop sizes affect giving decisions. In some standard implementations, the recipient is the same external entity (e.g. a charity) to which donations can be made by multiple dictators which plausibly leads to a crowding out of donations (as opposed to the case of multiple recipients which the meta-regression suggested has a significant positive effect). Another issue is that participants in a dictating role often take more than one giving decision throughout an experiment, but only one such decision is typically randomly chosen for payment. Therefore, the trichotomy we present and study here ought to be further refined and tested in the future.

One important distinction between the effect types that we made in the present paper, which we would like to stress again here, concerns the three broadly different categories of variables: population effects, effects due to framing, and effects due to different incentives. We believe this distinction is useful to think about other experimental games beyond dictator games too. Organized per these three categories, our results summarize as follows. Some population effects (such as age) were found to have unambiguous and sizeable effects, others not. This implies that it is important to qualify certain findings in light of the constitution of the underlying population which may create some biases, as also our differences in results regarding social elite, representative and online convenience sample indicated. The direction of the expected bias may depend on the protocol that is being used. Similarly, framing effects may have unambiguous and sizeable effects, but not necessarily so. It is important to control for framing effects inherent to an experimental design, as some of these may vary strongly between studies. Finally, variables related to incentive effects are mostly significant and matter. These effects include protocol variations—which were the focus of our efforts here—that need to be carefully controlled for as interactions with framing effects and population effects are likely. Further dedicated single-study experiments are therefore needed to organize some of these effects, particularly as they concern protocol differences and interactions with population differences.

In general, conclusions regarding differences between people need to be drawn carefully, as one should aim not to confound poorly understood protocol effects with such conclusions, and this meta-analysis provides further evidence that there are such effects. At present, interpreting meta-analytical results based on previously conducted studies remains difficult as only very few studies have carefully and transparently made their implementation details and data available. What is more, with the exception of our own ongoing work and Grech and Nax (2020) there has been little to no effort to pre-register dictator game experiments in the past. The trouble with meta-analyses mirrors some issues that are present

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

in experimental economics more broadly, where more transparency is needed to tackle the ongoing credibility and replication crises.[12] With improved frameworks for making data available, the hope is that promising methodologies such as meta-analyses could thrive productively in economics.[13]

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40881-022-00120-4.

Acknowledgements Funding was provided by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (Eccellenza).

Funding Open access funding provided by University of Zurich.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Andreoni, J., Douglas, B., & Bernheim. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77*(5), 1607–1636.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica, 70*(2), 737–753.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics, 116*(1), 293–312.

Andreozzi, L., Faillo, M., & Saral, A. S. (2021). Reciprocity in dictator games: An experimental study. SSRN working paper 3959663.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics, 9*(3), 193–208.

Brañas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior and Organization, 60*(3), 306–320.

Cárdenas, J.-C., Candelo, N., Gaviria, A., Polanía, S., & Sethi, R. (2008). Discrimination in the provision of social services to the poor: A field experimental study. Inter-American Development Bank Research Network working paper 54.

Eckel, C. C., Grossman, P. J., & Zhan, W. (2020).Does how we measure altruism matter? Playing both roles in dictator games. Department of Economics, discussion paper 05/20.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610.

Fisman, R., Jakiela, P., Kariv, S., & Markovits, D. (2015). The distributional preferences of an elite. *Science, 349*(6254), aab0096.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*(3), 347–369.

[12] See, for example, Page et al. (2021) for a discussion of this crisis.

[13] Christoph Engel kindly allowed us to make all the data underlying the project publicly available including his original dataset and our additions–see our Open Science Framework repo at https://osf.io/xc73h/.

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press

Grech, P., Halten, J., Hoeppner, S., & Nax, H. H. (2020). Comment on "Giving according to GARP: An experimental test of the consistency of preferences for altruism. SSRN http://ssrn.com/abstract=3766723. Accessed 1 Aug 2022.

Grech, P. D., & Nax, H. H. (2020). Rational altruism? On preference estimation and dictator game experiments. *Games and Economic Behavior, 119,* 309–338.

Greiff, M., Ackermann, K. A., & Murphy, R. O. (2018). Playing a game or making a decision? Methodological issues in the measurement of distributional preferences. *Games, 9*(4), 80.

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal: Journal of Mathematical Methods in Biosciences, 41*(8), 901–916.

Hartung, J., & Knapp, G. (2001). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine, 20*(12), 1771–1782.

Hartung, J., & Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine, 20*(24), 3875–3889.

Iriberri, N., & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics, 14*(2), 160–180.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business, 59*(4), S285–S300.

Klempt, C., & Pull, K. (2009). Generosity, greed and gambling: What difference does asymmetric information in bargaining make? Jena economic research papers working paper 021.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771–781.

Nax, H. H., Depoorter, B., Grech, P., Halten, J., Hoeppner, S., Newton, J., Pradelski, B., Soos, A., Wehrli, S., Ikica, B., & Jantschgi, S. (2020). Are elites really less fair-minded? SSRN http://ssrn.com/abstract=3766728

Page, L., Noussair, C. N., & Slonim, R. (2021). The replication crisis, the rise of new research practices and what it means for experimental economics. *Journal of the Economic Science Association, 7*(2), 210–225.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58*(3), 545–554.

Rigdon, M. L., & Levine, A. S. (2018). Gender, expectations, and the price of giving. *Review of Behavioral Economics, 5*(1), 39–59.

Zhang, L., & Ortmann, A. (2012). A reproduction and replication of Engel's meta-study of dictator game experiments. UNSW Australian School of Business research paper.

Zhang, L., & Ortmann, A. (2014). The effects of the take-option in dictator-game experiments: A comment on Engel's (2011) meta-study. *Experimental Economics, 17*(3), 414–420.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://doi.org/10.1007/s40881-022-00120-4 Published online by Cambridge University Press