COMMENT ON "GIVING ACCORDING TO GARP: AN EXPERIMENTAL TEST OF THE CONSISTENCY OF PREFERENCES FOR ALTRUISM"[1]

PHILIP D. GRECH[a], JESSICA HALTEN[b], SVEN HOEPPNER[c] AND HEINRICH H. NAX[a,d]

We present clean experimental evidence that a methodological confound was introduced by Andreoni and Miller (2002) that leads to diametrically opposed conclusions regarding comparisons of preferences between categories of fellow human beings distinguished by gender or age. Our study is a warning not to run the Andreoni-Miller tests of the consistency of altruistic preferences on the basis of their 'interactive' experimental protocol, and in particular not to perform preference estimations with the purpose of said comparisons. It is more interpretable and controllable to run the traditional, standard dictator game.

1. INTRODUCTION

James Andreoni and John Miller ('Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism' *Econometrica* 70(2) [2002]) adapt classical revealed preference theory to the context of altruism to check, rationalize and estimate prosocial and distributional preferences of experimental subjects. This is done on the basis of choice data generated by subjects who face various modified dictator games with different budgets and prices of redistribution. It is a foundational paper in the social preference literature, as it embeds altruism and distributional motives in a standard utility maximization framework.

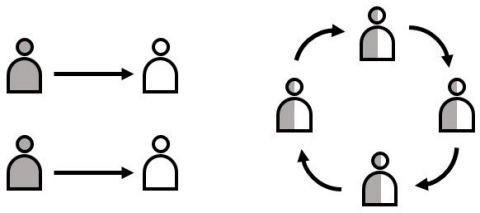
Unfortunately, besides all the conceptual advancement made in the paper, Andreoni and Miller (2002) also added an experimental twist to the implementation of dictator games away from the dichotomy of dictators and receivers, as pointed out in Grech and Nax (2020). In doing so they quite fundamentally depart from the original 'non-interactive' implementation of dictator games as single-player decision experiments (as introduced by Kahneman, Knetsch, and Thaler 1986, Forsythe, Horowitz, Savin, and Sefton 1994) in favor of a novel 'interactive' protocol. In the standard 'non-interactive' dictator game each experimental subject is strictly either only 'Dictator' or 'Recipient' but never both. By contrast, in the novel, 'interactive' protocol subjects are placed in a chain of people giving to

[1]July 4, 2020. We thank Stefan Wehrli for lab assistance.

[a]Behavioral Game Theory, ETH Zurich, Zurich.
[b]School of Management, Yale University, New Haven.
[c]Charles University, Prague.
[d]University of Zurich, Zurich. Correspondence: heinrich.nax@uzh.ch

each other. See Figure I. Naturally, the use of this protocol invites all kinds of additional complexities and confounding factors compared to the standard protocol such as beliefs regarding others' giving behavior, cognitive hierarchy, etc., because everybody is both someone's dictator and someone else's recipient at the same time, thus rendering the decision environment interactive and strategic. In fact, the appropriate solution concept would appear to be Bayesian Nash equilibrium rather than single-player decision theory. Using this protocol is cheaper, obviously, as every subjects generates data, but potentially produces unknown biases.

FIGURE 1.— DICTATOR GAME PROTOCOLS WITH 4 PARTICIPANTS. *Left:* Standard 'non-interactive' as in Kahneman, Knetsch, and Thaler (1986), Forsythe, Horowitz, Savin, and Sefton 1994, etc. – half of the subjects are dictators (shaded). *Right:* Non-standard 'interactive' as in Andreoni and Miller (2002) – all subjects give and receive at the same time.



The use of this protocol has become one of the standard ways to run dictator games. It was first applied for a gender comparison in the U.S. in Andreoni and Vesterlund (2001), with which Andreoni and Miller (2002) shares large parts of its data.[1] There are many other examples listed in Grech and Nax (2020), including a comparative study of elite versus non-elite behavior by Fisman, Jakiela, Kariv, and Markovits (2015), which is where we stumbled across the interactive protocol in the first place. Cameron, Erkal, Gangadharan, and Meng (2013), for example, use it to compare young and old in China. The problem of that protocol is that it invites possibly unmanageable confounds, as the resulting interactive game is hardly tractable in terms of Bayesian Nash predictions. But even if we stick with the single-player view of behavior in the interactive protocol for the analysis, it remains to be studied how the use of this protocol affects the choice data, for example concerning said gender and age differences.

We have been working on understanding these effects for quite some time. First, we

[1]Despite Andreoni and Vesterlund (2001) appearing earlier than Andreoni and Miller (2002) in time, we conclude that Andreoni and Miller (2002) is the original paper, because Andreoni and Vesterlund (2001) cite an earlier working paper version of Andreoni and Miller (2002), but not vice versa. In fact, Andreoni and Miller (2002) thank Lise Vesterlund for help collecting the data in the acknowledgements.

developed some rational-choice foundations for the interactive protocol (Grech and Nax, 2020), and found experimental evidence that behavior in the interactive game differs from that in the non-interactive. It is not yet clear how from that analysis due to limited variation in the subject pool.[2] Second, we went into all of the instructions of all papers mentioned in the meta study of Engel (2011), and checked precisely which protocol was used in every single one (see https://osf.io/xc73h/ where we store the relevant data, and report this ongoing work). At least ten percent of studies use the interactive protocol, tendency rising, with several of the most influential, controversial and best-published ones amongst them. The resulting meta analysis suggests that the use of the interactive protocol results in less giving but a more positive responsiveness to the giving multiplier (i.e. the inverse of the price of redistribution), thus producing more selfish- and efficient-looking data. Third, criticizing the elite versus non-elite conclusions in Fisman, Jakiela, Kariv, and Markovits (2015), we have identified how their kind of protocol potpourri, where one set of subjects plays one protocol and another set of subjects the other, can distort comparability.

Here, we report on a large set of interactive and standard experiments, through which we shall show how the protocol interacts with various categories that are of interest regarding comparisons such as gender, age, and earnings. We conclude that the interactive protocol should be avoided when aiming to pursue a comparative agenda, because it produces a perfect storm: the protocol is virtually invisible at the aggregate, but might produce diametrically opposed conclusions.

2. CONFOUNDS

Our dataset comes from an experimental investigation of protocol confounds on a sample of 1'464 active dictator subjects, of which 633 are Yale students including 220 women, and the rest are from a representative sample of the US population with 831 active dictators of which 407 are female. We rely on categorization as 'female' by stated gender in the panel, and define 'old' as 50 years of age or older. All materials, data, analyses files, detailed outputs, etc. are available at the Open Science Framework at https://osf.io/tgqsb/?viewonly=d09d2801727a408ab3c5eb8764f351fa.

Table I illustrates that the interactive protocol generates data that reproduces the famous finding of Andreoni and Vesterlund (2001) that men give more when it is cheap, while women give more when it is expensive. By contrast, in the standard protocol, the conclusions are–if they are publishable at all given how not in vogue they might seem–that men

[2]See https://osf.io/fsg52/ for all the pre-registration, data and the entire project.

give more in all but one situation. Similar phenomena are present for comparisons of age groups. In the interactive protocol the young appear to give more when it is cheap, and the old when it is expensive. These differences are mostly gone, or reversed in favor of the young, in the standard protocol. In light of data from the US today, we strongly reject the hypothesis of systematic differences in altruism by age or gender as found in earlier studies.

The implications of our findings are that studies aiming to draw conclusions of that fundamental a caliber ought to operate with the best and most established (ideally pre-registered) tools available in a way that lends itself to direct interpretability, comparability and falsification. As Sir Karl says so beautifully–"for nothing is easier than to select statistical evidence so that it is *favourable* to a statistical hypothesis—-if we wish to do so." (Karl Popper (1959) in *The Logic of Scientific Discovery*, p. 418).

REFERENCES

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

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

Cameron, L., N. Erkal, L. Gangadharan, and X. Meng (2013): "Little emperors: behavioral impacts of China's One-Child Policy," *Science*, 339(6122), 953–957.

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

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

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

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

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

Popper, K. (1959): *The Logic of Scientific Discovery*. Routledge.

TABLE I

GENDER CONFOUNDS. We report whose mean giving, differentiated by gender, age, income and education is higher when faced with various budget lines. Key findings regarding gender and age are gone, or reversed.

Dictator Game protocol used	Multi-plier	Endow-ment	Who is fair? ♀ / ♂	Old/Young
Standard	.10	200	♂	Young
	.20	100	♂	Young
	.25	40	♂	Young
	.33	40	♂	Old
	.50	60	♂	Old
	.50	75	♀	Young
	1.00	40	♂*	Young
	1.00	60	♂	Young
	1.00	80	♂	Young
	1.00	100	♂	Young
	2.00	60	♂***	Young
	2.00	75	♂***	Young*
	3.00	40	♂**	Young
	4.00	40	♂***	Young**
	5.00	20	♂***	Young
	10.00	20	♂***	Young
	average		♂***	Young*
Non-standard	.10	200	♀***	Old***
	.20	100	♀***	Old***
	.25	40	♀***	Old***
	.33	40	♀***	Old***
	.50	60	♀***	Old***
	.50	75	♀***	Old***
	1.00	40	♀**	Old***
	1.00	60	♀**	Old***
	1.00	80	♀***	Old**
	1.00	100	♀**	Old**
	2.00	60	♂	Young
	2.00	75	♀	Young
	3.00	40	♂	Young
	4.00	40	♂	Young
	5.00	20	♂	Young
	10.00	20	♂*	Young
	average		♀***	Old***

*/**/***: significant at the ten/five/one-% level based on Mann–Whitney–Wilcoxon tests.