European Economic Review 85 (2016) 272–287

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Group size and the (in)efficiency of pure public good provision

Johannes Diederich, Timo Goeschl, Israel Waichman *

Department of Economics, Heidelberg University, Bergheimer Str. 20, 69115 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 21 July 2015
Accepted 2 March 2016
Available online 18 March 2016

JEL classification:
C93
D74
H41

Keywords:
Public good game
Large group
Extra-laboratory experiment
Free riding
Belief elicitation

ABSTRACT

Are larger groups better at cooperation than smaller groups? This paper investigates, under controlled conditions, the presence and direction of a possible group size effect in pure public good provision by large heterogeneous groups. Employing subjects drawn from the general population and introducing Internet-based procedures to study this question, we collected experimental evidence from 1110 subjects playing a linear public goods game in groups of 10, 40, and 100 members. We find a positive and significant group size effect: Increasing group size by a factor of 10 (4) increased efficiency by 10 (6) percent. The effect arose at the intensive margin and with repetition. Those who contributed contributed more in larger groups. Larger and smaller groups had similar initial contribution levels, but cooperation rates declined more slowly in the larger groups. Free-riding was invariant to group size, despite subjects' persistent beliefs of a negative group size effect at the extensive margin. Further econometric examination of the data supports these findings and provides starting points for future theoretical and experimental research on the group size effect.

1. Introduction

Many social dilemmas of policy relevance involve a large number of heterogeneous participants that differ with respect to important characteristics such as gender, income, age, and social preferences. The theoretical limit case for such large-scale social dilemmas is the private provision of pure public goods (PG) in a heterogeneous population of infinitely many agents (Andreoni, 1988). One frequently mentioned real-world counterpart is climate change (Nordhaus, 1993), but many other effectively pure public goods such as knowledge production, environmental quality, and national defense share this feature (e.g., Buckley and Croson, 2006).

When pure PG are provided by large groups of heterogeneous individuals, the scalability of cooperation is a question of obvious relevance. Do cooperation levels for pure PG vary with the size of the group and, if yes, how? Despite the significance of this question, there is—to our knowledge—little evidence derived under controlled conditions that directly speaks to this question.[1] A distinguished experimental literature has examined the issue of group size and pure PG, but only

* Corresponding author.
E-mail addresses: diederich@eco.uni-heidelberg.de (J. Diederich), goeschl@eco.uni-heidelberg.de (T. Goeschl), waichman@eco.uni-heidelberg.de (I. Waichman).

[1] There is a small literature that examines this question using non-experimental data. Zhang and Zhu (2011) study a subsample of the general population, voluntary contributors to the Chinese-language version of Wikipedia, and find that negative exogenous variations in its potential readership are associated with decreased contribution rates. Other contributions that study field evidence on the provision of public good by large groups are, e.g., Goetze et al. (1993) and Haan and Kooreman (2002). To what extent the field evidence can be directly compared with experimental results from specific game forms is not immediately clear. The degree to which the specific public goods investigated are strictly pure, i.e., exhibit a constant MPCR, or are more comparable to congestible public goods is a matter of discussion.

http://dx.doi.org/10.1016/j.euroecorev.2016.03.001
0014-2921/

in the context of relatively homogenous populations and mostly for small group sizes.[2] In the typical experiment, up to 10 student subjects interact in the context of the voluntary contribution mechanism (VCM) (e.g., Isaac et al., 1984; Isaac and Walker, 1988; Goeree et al., 2002; Carpenter, 2007; Nosenzo et al., 2015). The majority of such laboratory experiments with small groups do not find evidence for a group size effect in the VCM, with the exception of Isaac and Walker (1988) and Nosenzo et al. (2015) who report a positive group size effect for a marginal per-capita return (MPCR) of 0.3, but not for MPCRs of 0.75 or above. Evidence on large groups of students comes from only two sources: one is the seminal on-campus VCM experiment with groups of 4, 10, 40, and 100 subjects by Isaac et al. (1994) in which the researchers find a positive group size effect for an MPCR of 0.3, but not for an MPCR of 0.75. The other, Weimann et al. (2014), is a laboratory VCM experiment with groups of 30 and 40 members with MPCRs between 0.04 and 0.12, groups of 60 members with MPCRs between 0.02 and 0.06 and groups of 100 members with MPCRs between 0.02 and 0.04 that investigates the interplay of group size and MPCR rather than specifically testing for pure group size effects.

The existing evidence on student groups represents a valuable, but rather limited base from which to infer the presence and nature of a group size effect in pure PG for large heterogeneous groups. The limitations are not only quantitative, but also qualitative: groups drawn exclusively from one specific demographic sub-population are relatively homogeneous compared to groups that combine subjects with heterogeneous characteristics and may therefore behave differently (e.g., Andersen et al., 2010; Falk et al., 2013). In addition, students represent a very specific subset of the general population in terms of age, education, and income. This matters because in previous small-scale PG experiments comparing student and non-student subjects, there were significant differences in behavior (Gächter et al., 2004; Belot et al., 2015). Group size effects observed among groups of student subjects therefore represent potentially biased estimates of the group size effect present in the more common case of heterogeneous groups.

The present paper bridges the gap between the current evidence base and the question whether size matters when large heterogeneous groups provide pure PG. It does so by drawing on new evidence from VCM experiments with group sizes of 10, 40, and 100 subjects at an MPCR of 0.3. This evidence is novel for two reasons. One is that the observations come from a highly heterogeneous set of subjects with respect to gender, age, education, and place of residence. The composition of the subject pool therefore brings us considerably closer to what can be regarded as a typical degree of heterogeneity in the general population. This subject pool also behaves differently from a sample of student subjects undergoing the same experiment. The second reason is that the evidence is generated through experimental procedures that offer new answers to the inherent challenges of conducting large-scale VCM experiments with a heterogeneous subject pool. Among the challenges, three are most salient. One is the challenge of recruiting a suitable number of subjects from the general population rather than from the more accessible pool of student subjects. A second is the necessity of ensuring that groups of subjects whose size exceeds the typical laboratory capacity interact reliably and repeatedly under controlled conditions. And a third challenge is the resource requirements of large-scale VCM experiments because aggregate payoffs to large groups quickly exhaust experimental funds even when the marginal return from the public good is relatively modest.[3] The joint presence of these challenges explains the relative paucity of evidence on the research question that this paper investigates. It also informs the present design, namely an extra-laboratory procedure that retains—for the sake of comparability—many of the features of Isaac et al. (1994), but that recruits a diverse sample of (mutually) anonymous subjects from the general population while ensuring a high degree of compliance through the use of a simple-to-use Internet-based interface. We demonstrate below that this response to the three challenges both strengthens the internal validity of the results and can offer a template for future large-scale VCM experiments with heterogeneous populations.

Based on this new evidence, the paper delivers the following results. First, larger groups of heterogeneous subjects were on average more efficient in producing the pure PG than smaller groups. These differences in cooperation at different group sizes operated at different margins and emerged over time. At the extensive margin, cooperation was invariant to group size: small groups and large groups had statistically indistinguishable shares of free-riders. At the intensive margin, by contrast, the relationship was clearly positive: those who contributed contributed more in a larger groups. The positive group size effect was mostly driven by different rates of decline in cooperation at the intensive margin. Second, subjects' beliefs were partially inconsistent with the observed outcome: subjects in larger groups reported that they expected significantly more strict free-riding, giving rise to a negative group size effect in beliefs at the extensive margin. At the same time, they correctly predicted the positive intensive margin effect of group size on contributions. An econometric analysis confirms the core results and demonstrates that across all rounds, the group size effect at the intensive margin is robust to including a full range of sociodemographic controls, round effects, and beliefs. Our findings inform both for future theoretical as well as experimental work on the drivers of the group size effect.

We proceed as follows: we explain the experimental design and procedure in Section 2. In Section 3, we present analysis and experimental results, which we discuss in Section 4. Finally, Section 5 concludes.

[2] The question of what constitutes a 'small' and what a 'large' group has no obvious answer. Experiments on PG with 10 subjects or less are abundant while those with more than 20 members are relatively rare. Against this background, we would argue that any group size above 20 may be considered 'large'.

[3] For an illustration, consider a linear PG experiment with 100 group members where endowment per subject (aggregated over all rounds) is $10 and the MPCR is 0.3. If all subjects contribute all their endowments to the public good, the payment to subjects will be $30,000 in order for the researcher to obtain a single independent observation at the group level.