

# Group size effects and critical mass in public goods games

María Pereda<sup>1,2</sup>, Valerio Capraro<sup>3,\*</sup>, and Angel Sánchez<sup>2,4,5,6</sup>

<sup>1</sup>Universidad Politécnica de Madrid. Departamento Ingeniería de Organización, Administración de empresas y Estadística, Madrid, Spain

<sup>2</sup>Unidad Mixta Interdisciplinar de Comportamiento y Complejidad Social (UMICCS), UC3M-UV-UZ, Spain

<sup>3</sup>Economics Department, Middlesex University London, Business School, The Burroughs, London NW4 4BT, United Kingdom

<sup>4</sup>Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain

<sup>5</sup>Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

<sup>6</sup>Institute UC3M-BS for Financial Big Data (IFiBiD), Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain

\*Corresponding author: v.capraro@mdx.ac.uk

## ABSTRACT

Understanding whether the size of the interacting group has an effect on cooperative behavior has been a major topic of debate since the seminal works on cooperation in the 1960s. Half a century later, scholars have yet to reach a consensus, with some arguing that cooperation is harder in larger groups, while others that cooperation is easier in larger groups, and yet others that cooperation attains its maximum in intermediate size groups. Here we add to this field of work by reporting a two-treatment empirical study where subjects play a Public Goods Game with a Critical Mass, such that the return for full cooperation increases linearly for early contributions and then stabilizes after a critical mass is reached (the two treatments differ only on the critical mass). We choose this game for two reasons: it has been argued that it approximates real-life social dilemmas; previous work suggests that, in this case, group size might have an inverted-U effect on cooperation, where the pick of cooperation is reached around the critical mass. Our main innovation with respect to previous experiments is that we implement a within-subject design, such that the same subject plays in groups of different size (from 5 to 40 subjects). Groups are formed at random at every round and there is no feedback. This allows us to explore if and how subjects change their choice as a function of the size of the group. We report three main results, which partially contrast what has been suggested by previous work: in our setting (i) the critical mass has no effect on cooperation; (ii) group size has a positive effect on cooperation; (iii) the most chosen option (played by about 50% of the subjects) is All Defection, followed by All Cooperation (about 10% of the subjects), whereas the rest have a slight trend to switch preferentially from defection to cooperation as the group size increases.

## Introduction

In the last decades, a variety of disciplines including anthropology, biology, economics, physics and psychology have addressed the issue of the inordinate scale of human cooperation.<sup>1-8</sup> Even in the simplest societies, individuals cooperate in groups of genealogically distant individuals.<sup>9,10</sup> In the perpetually connected society<sup>11</sup> we live in today, instances requiring cooperation arise among groups of people of largely different sizes, and it is therefore of crucial importance to understand if and how group size affects cooperation. This question has been extensively debated in the literature since the seminal works on cooperation in the 1960s, leading to two different viewpoints. On the one hand, some scholars suggest that, in so far as the benefits of cooperation decrease with an increasing number of participants, cooperation should decrease with group size.<sup>12,13</sup> On the other hand, other researchers have argued that benefits do not necessarily decrease as the group size increases and, in that case, they reason that cooperation might even increase with the size of the group.<sup>14,15</sup> These two viewpoints have been mirrored by empirical, theoretical, and numerical works, with some finding a negative effect<sup>16-22</sup>, others finding a positive effect<sup>22-31</sup>, and yet others finding an ambiguous or non-significant effect.<sup>32-34</sup>

A commonly used, stylized setup to probe into the issue of cooperation is the Public Good Game<sup>35</sup> (PGG). A PGG is a cartoon of the problem of provision of resources with three main characteristics: They are jointly provided, they are non-excludable (it is not possible to prevent those who have not paid for them from having access to them), and they are also non-rivalrous (resource consumption by one consumer does not prevent simultaneous consumption by others). Economic

theory presents this as a game in which agents have to decide how much of their resources to contribute to the creation of a public good and how much to spend on private goods. The agents' contribution is multiplied by some factor (linear PGG) and then distributed equally among everybody, irrespective of their own contribution. Even in such a simplified situation, the jury is still out on the issue of size dependence, with some researchers reporting an increase of contribution with the number of players<sup>22-24</sup>, and others finding no effects or negative effects<sup>21,36</sup>; a meta-analysis of 27 experiments<sup>37</sup> concluded that there is mild positive effect of group size.

One way to make sense of these apparently contradictory results is by assuming that the effect of the group size on cooperation is not domain-general but depends on specific properties of the interaction. As suggested above, one potential discriminator is the way the benefit of full cooperation increases as a function of the group size. The intuition is simple: Since it is known that cooperation is positively related to the benefit of full cooperation<sup>38,39</sup>, if the benefit for full cooperation increases as the size of the group increases, then larger groups might get more tempted to cooperate; if, instead, the benefit for full cooperation decreases as the function of the group size, then larger groups might be less tempted to cooperate. Once again, from the theoretical viewpoint the issue is far from resolved. Thus, Peña considered evolutionary models, finding that the outcome of general nonlinear public goods games depends not only on the average group size but also on the variance of the group-size distribution in case groups are heterogeneous<sup>40</sup>; also, he showed that larger group sizes can have negative effects (by reducing the amount of cooperation in some cases) and positive effects (by enlarging the basin of attraction of more cooperative outcomes) on the evolution of cooperation<sup>41</sup>. Other works addressing multiplayer games from a more general but still evolutionary viewpoint point to the complexity of this problem<sup>42-44</sup>.

In order to gain insight into this problem, Capraro & Barcelo<sup>45</sup> considered a class of  $N$ -person general public goods game parameterized by a function  $\beta(\Gamma, N)$ , representing the marginal return for cooperation when  $\Gamma$  people out of  $N$  cooperate. In the presence of  $\Gamma$  cooperators, the payoff of a cooperator is defined as  $\beta(\Gamma, N) - c$  ( $c > 0$  represents the cost of cooperation), while the payoff of a defector is defined as  $\beta(\Gamma, N)$ . In order to have a social dilemma, two properties are required:

- Full cooperation gives a larger payoff than full defection:  $\beta(N, N) - c > \beta(0, 0)$ ;
- Defecting is individually optimal, regardless of the number of cooperators: for all  $\Gamma \leq N$ , one has  $\beta(\Gamma - 1, N) > \beta(\Gamma, N) - c$ .

Barcelo & Capraro<sup>22,45</sup> argue that, when  $\beta(\Gamma, N)$  is constant or even decreasing in  $\Gamma$ , then group size should have a negative effect on cooperation, whereas, when  $\beta(\Gamma, N)$  is increasing in  $\Gamma$ , then group size should have a positive effect on cooperation. They show that these two predictions are satisfied in two experiments and can be formalized by assuming that agents play according to the cooperative equilibrium model introduced by Capraro.<sup>46</sup> Putting these results together, it follows that, if we consider a piecewise linear-then-constant function  $\beta(\Gamma, N)$ , then group size should have a curvilinear, inverted-U effect on cooperation.

Note that public goods games with piecewise linear-then-constant return for cooperation are interesting because they approximate those real-life social dilemmas for which the public good has natural limits such that, from some point on, any new cooperator brings no additional benefit. For example, suppose that a hospital needs 50 litres of blood for the 50 victims of a terror attack. Of course, every litre of blood is initially important because it can save one life. However, after reaching the 50 litres, collecting more blood brings no additional benefit. Another example regards academic collaborations. Suppose that a team of researchers wants to start a multidisciplinary project that lies between three disciplines, for instance, psychology, anthropology and physics; and suppose that the team is looking for psychologists, anthropologists and physicists to take part to the project. Clearly, once each of these three disciplines is sufficiently covered, getting on board one more specialist would bring no additional benefit. The practical importance of this type of public goods games raises a crucial question: What is the group size effect on public goods games with a piecewise linear-then constant return for cooperation? Previous work using field experiments on real-life public goods games have repeatedly shown that medium-size groups tend to cooperate more.<sup>47-51</sup> In particular, connecting to one of our examples above, numerous studies have confirmed that, in academic research, the research quality of a research group is maximized by intermediate-size groups.<sup>52-54</sup> These findings are, thus, broadly consistent with the view that public goods game with a piecewise linear-then-constant returns for cooperation should give rise to a inverted-U effect of group size on cooperation.

Following this theoretical prediction, Capraro & Barcelo<sup>45</sup> carried out one experiment to probe into the existence of this curvilinear dependence of the group size. As predicted, they found evidence of an inverted-U dependence of the number of cooperators with the group size. However, while their prediction was that the rate of cooperation should start decreasing at  $N = 10$ , it turned out that it reached its maximum in groups of size  $N = 15$ . This immediately poses the question as to how the optimal group size for cooperation depends on the size at which  $\beta(N, N)$  becomes constant. This is an important question in light of possible applications of the curvilinear effect: knowing at which group size cooperation attains its maximum has obvious applications to team formation. In addition, a second question that was not answered by previous work regards the behavior of people at the individual level and its dependence on the group size. All the empirical studies on group size effect on

cooperation that we are aware of uses a between-subject design, such that different subjects interact in groups of different sizes. A major innovation of our work is that we implement a within-subject design, such that the same subject interacts in groups of different sizes, which will allow us to explore heterogeneity in people's behavior as a function of the group size.

## Results

### Experimental setup

We conducted an online experiment with 200 subjects, that took place in eight sessions, one every two weeks between October 2017 and February 2018. The participants were Spanish volunteers from the IBSEN pool of subjects<sup>55</sup>. In each session, participants played one round of a  $N$ -person general public goods game with Critical Mass (i.e.,  $\beta(\Gamma, N)$  piecewise linear-then-costant). Sessions differed in the size of the interacting group, from  $N = 40$  in the first session to  $N = 5$  in the last one. Groups were randomly formed at the end of every session to compute the payoff. However, to avoid iterated game effects, participants received no feedback about the choices made by others or their own payoff until the last session was finished. To explore the influence on cooperation of the value where  $\beta(N, N)$  stabilizes (hereafter,  $N_C$ ), we carried out two different treatments, namely  $N_C = 10$  and  $N_C = 20$ , each treatment with 100 participants initially assigned to it. Of these 200 volunteers we recruited, only 163 showed up the first day, having as a result a first day dropout of 18.5%. The experiment ended up with 107 volunteers that made all the decisions (average drop out rate of 5% per session). The experiment was implemented in IBSEN-oTree<sup>55</sup>. Participants played online through a web browser in a computer, tablet or mobile phone; and could made their decisions during eight hours each session day. Email reminders were sent every two weeks (i.e., before every session) to remind people to participate. Further details can be found in the **Methods** section. Only participants that completed the experiment were paid, as it was stated in the instructions. Participants were paid using PayPal. The average earning was 8 Euros. The 25% participants who scored highest in each treatment got a chance to enter a lottery of 50 Euros, and correspondingly two participants earned an additional price of 50 Euros.

### Data analysis: Cooperation as a function of the group size

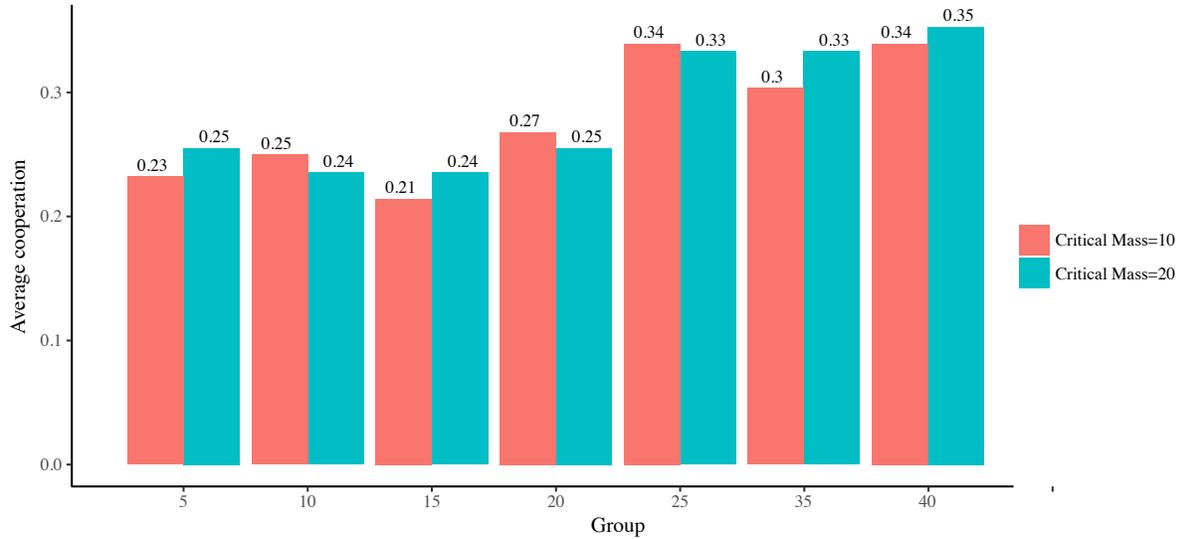
As already indicated above, a total of 93 subjects (47.3% females) dropped out over the course of the experiment. Moreover, because of an internet connection problem, the condition  $N = 30$  was interrupted before data completion and, as a consequence, only 98 subjects participated in this condition. In view of these circumstances, the analysis is structured as follows: We first analyze subjects for whom we have all decisions and excluding the condition  $N = 30$ . Subsequently, as robustness checks, we include also subjects who dropped out during the experiment and the condition  $N = 30$ . As we will see below, the overall pattern of results is robust across these exclusions.

To carry out the analysis, we use the following variables:  $C$  is equal to 1 if a player cooperates, and 0 if a player defects;  $N_C$  is equal to 10 if a player participates in the condition where  $\beta(N, N)$  stabilizes at  $N = 10$ , and equal to 20 if a player participates in the condition where  $\beta(N, N)$  stabilizes at  $N = 20$ ;  $N$  is equal to the size of the group. Since  $C$  is a binary variable, the analysis will be conducted using a logistic model estimating the probability of cooperation  $p(C)$  as a function of the dependent variables  $x_1, \dots, x_n$  (which will be specified later):

$$p(C) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 x_1 + \dots + \alpha_n x_n)}} \quad (1)$$

We begin by considering subjects for whom we have all the observations and excluding the condition  $N = 30$ . In Figure 1 we plot the percentage of people who cooperated. As a first step of the analysis, we observe that logistic regression predicting  $C$  as a function of  $N_C$  finds no statistically significant effect ( $p = 0.818$ ). Thus we can collapse data across the  $N_C$  treatments. Subsequently, logistic regression predicting  $C$  as a function of  $N$  finds a statistically significant positive effect, such that larger groups tend to cooperate more (coeff= 0.017,  $z = 2.45$ ,  $p = 0.014$ ). The aforementioned fact that  $N_C$  does not affect  $C$  is reflected in the fact that the positive effect of  $N$  on  $C$  is similar in the two treatments ( $N_C = 10$ : coeff= 0.016,  $z = 1.70$ ,  $p = 0.089$ ;  $N_C = 20$ : coeff= 0.017,  $z = 1.86$ ,  $p = 0.078$ ).

We now turn to the robustness checks. As a first check, we repeat the analysis by including  $N = 30$  in the data (19.44% cooperators in  $N_C = 10$  and 21.62% cooperators in  $N_C = 20$ ). As before, logistic regression predicting  $C$  as a function of  $N_C$  finds no statistically significant effect ( $p = 0.926$ ), and hence we collapse again data across the  $N_C$  treatments. In line with the previous results, we find a statistically significant positive effect of  $N$  on  $C$  (coeff= 0.014,  $z = 2.18$ ,  $p = 0.029$ ), which is similar in the two treatments ( $N_C = 10$ : coeff= 0.015,  $z = 1.54$ ,  $p = 0.123$ ;  $N_C = 20$ : coeff= 0.015,  $z = 1.54$ ,  $p = 0.123$ ). As a second robustness check, we include subjects who dropped out during the experiment, while excluding subjects who participated in  $N = 30$ . Once again, we find that  $N_C$  has no statistically significant effect on  $C$  ( $p = 0.424$ ) and, after collapsing data across the  $N_C$  treatments, we find a statistically significant positive effect of  $N$  on  $C$  (coeff= 0.016,  $z = 2.66$ ,  $p = 0.008$ ). Splitting the analysis by  $N_C$  reveals that the group size effect is significant only for  $N_C = 20$  ( $N_C = 10$ : coeff= 0.013,  $z = 1.53$ ,



**Figure 1.** Average cooperation as a function of group size.

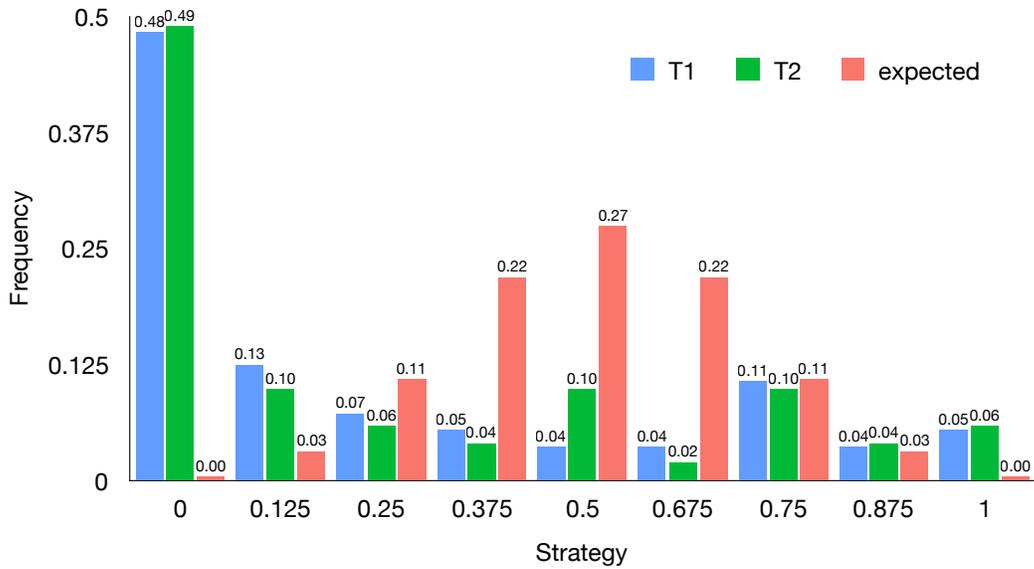
$p = 0.125$ ;  $N_C = 20$ :  $\text{coeff} = 0.019$ ,  $z = 2.21$ ,  $p = 0.027$ ). This suggests that the effect of group size might be stronger for  $N_C = 20$ . However, the difference between the two effects is not statistically significant, as revealed by a logistic regression including the  $N_C \times N$  interaction term ( $p = 0.897$ ). Therefore, also in this case we find that the group size effect does not depend on the critical mass. Finally, in our third and last robustness check, we analyze all data, that is, we include both subjects who dropped out during the experiment and those who participated in  $N = 30$ . The results are qualitatively the same:  $N_C$  has no statistically significant effect on  $C$  ( $p = 0.398$ ), and when we collapse data across treatments  $N$  has a statistically significant positive effect on  $C$  ( $\text{coeff} = 0.015$ ,  $z = 2.45$ ,  $p = 0.014$ ), which is similar in the two treatments ( $N_C = 10$ :  $\text{coeff} = 0.012$ ,  $z = 1.38$ ,  $p = 0.169$ ;  $N_C = 20$ :  $\text{coeff} = 0.018$ ,  $z = 2.07$ ,  $p = 0.039$ ), with the  $N_C \times N$  interaction term turning out to be not statistically significant ( $p = 0.742$ ). Therefore, our analysis combined with the three robustness checks we have considered, allows us to draw two main conclusions: First, the size of the group has a positive effect on cooperation, and second, the positive effect of group size on cooperation does not depend on the critical mass. We have also conducted the same analyses by adding a quadratic term for the group size effect, in order to test for the inverted-U effect. Results remain qualitatively the same, and the quadratic term is never significant.

One additional step we have taken to assess the reliability of the above findings is to study separately the average cooperation among subjects who dropped out during the experiments (44 in the treatment  $N_C = 10$  and 49 in the treatment  $N_C = 20$ ), by comparing it with the average cooperation among subjects who did not drop out. The idea behind this check is that, as the experiment proceeded, it could be the case that individuals more prone to cooperation were overrepresented among the dropouts, thus leading to a decrease in average cooperation as the experiment proceeded from the largest to the smallest size. To test for this possibility, we built a dummy variable named *Incomplete*, which takes value 1 if a subject dropped out during the experiment, and 0 if a subject completed the experiment. We then conducted a set of logistic regressions predicting  $C$  as a function of *Incomplete*, for each value of  $N$ . In doing so, we find all the p-values to be larger than 0.2. This suggests that the average cooperation among those who dropped out during the experiment is not statistically different than the average cooperation among those who completed the experiment, which in turn supports the conclusion that our results are unlikely to be driven by the relatively high rate of dropping out.

### Within-subject analysis

In the previous subsection, we have looked at our experiment from the viewpoint of the global results on cooperation and its dependence on the group size. Now we move on to studying heterogeneity in people's behavior. We begin by observing that, since each participant has to take eight independent decisions about whether to cooperate or not, there are  $2^8 = 256$  possible combinations of cooperation and defection. However, as summarized in Table 1, for each  $N_C$  treatment, only 25 of them, less than 10%, are actually being played.

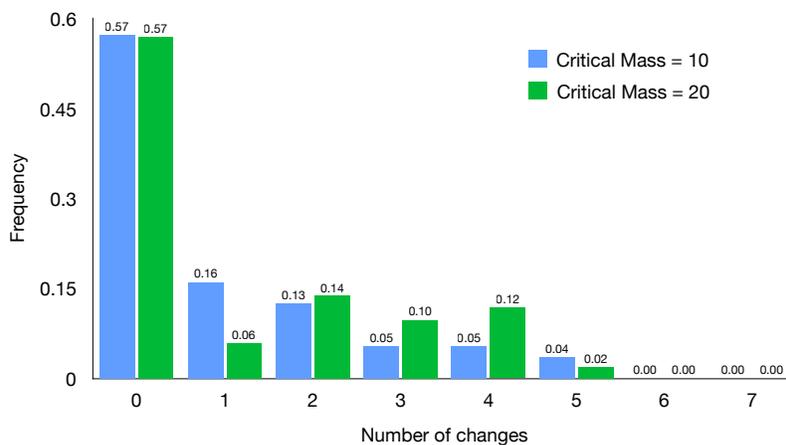
Not surprisingly, in both treatments  $N_C = 10$  and  $N_C = 20$ , the most played strategy is AllD (All Defection), making a total of about 48% of subjects in the treatment  $N_C = 10$  (counting also subjects who did not participate in condition  $N = 30$ ), and 49% of subjects in the treatment  $N_C = 20$ . Perhaps more surprisingly, in both treatments, the second most played strategy, albeit at a large distance from the first, is AllC (All Cooperation), which makes about 9% of the observations in the treatment



**Figure 2.** The  $x$  axis reports the frequency with which cooperation appears in a given sequence of choices (e.g,  $x = 0$  corresponds to All Defection, whereas  $x = 1$  corresponds to All Cooperation). The  $y$  axis reports the frequency of people playing strategies with a given proportion of cooperative strategies, for treatments T1 and T2, versus the expected frequency of each strategy is they were chosen at random.

$N_C = 10$  and 8% of the observations in the treatment  $N_C = 20$  (again counting also subjects who did not participated in the session with  $N = 30$ ). All other strategies have been played by less than 4% of the participants. For comparison, Figure 2 shows the distribution of sequences we would expect if the decisions were taken at random. Each “strategy” groups the set of sequences with the same rate of cooperation; for example, cooperating once and defecting 7 times is “strategy 0.125”. The plot makes it evident that participants’ choices are largely intentional: People play AllD far more than expected, and the most expected strategies arising from random choices are barely played. Only highly cooperative strategies (0.75, 0.875) appear to a similar extent to what would be expected by chance, except for AllC, which appears to be more frequent.

Therefore, about half of the participants played the Nash equilibrium of the game, choosing to defect in all sessions, irrespective of the value of the critical mass  $N_C$ . On the other hand, as a consequence of the fact that the majority of the subjects do not change strategy, the average number of times a player changes strategy over the course of the eight sessions is relatively low (1.08 out of maximum of 7). In fact, only 43% of the participants changed their decision at least once: As shown in Figure 3, there are subjects who change strategy even five times, but no one changes strategy more than that.



**Figure 3.** Frequency of people as a function of the number of changes in their sequence of decisions, for treatments  $N_C= 10$  and  $N_C= 20$ .

**Table 1.** Frequency (and number of subjects) playing each strategy, for treatments  $N_C = 10$  and  $N_C = 20$ . D stands for Defection, C stands for Cooperation.

Critical Mass = 10			Critical mass = 20		
Strategy	# Subjects	Frequency	Strategy	# Subjects	Frequency
DDDDDDDD	20	0.3571	DDDDDDDD	20	0.3922
DD_DDDDD	7	0.1250	DD_DDDDD	5	0.0980
CCCCCCCC	3	0.0536	CCCCCCCC	3	0.0588
DC_CCCCC	2	0.0357	DCDDDDDD	2	0.0392
CDDDDDDD	2	0.0357	DDDDDDCD	1	0.0196
CD_CCCCC	2	0.0357	DDCDDDDD	1	0.0196
CC_CCCCC	2	0.0357	DCDCDDDD	1	0.0196
DDDDDDDC	1	0.0179	DCDCCDD	1	0.0196
DDDCDDDD	1	0.0179	DCCDCCDC	1	0.0196
DD_DCDDD	1	0.0179	DC_CCCCC	1	0.0196
DCDDCDDD	1	0.0179	CDDDDDDD	1	0.0196
DCDCCCC	1	0.0179	CDDDCDDD	1	0.0196
DCCCDDCD	1	0.0179	CDCCDCCC	1	0.0196
DC_DDDDD	1	0.0179	CD_DCDDC	1	0.0196
DC_DCCCC	1	0.0179	CD_CDDDD	1	0.0196
CDDDDDCD	1	0.0179	CD_CCDDD	1	0.0196
CDDCDDDD	1	0.0179	CD_CCCCC	1	0.0196
CDCCDCC	1	0.0179	CCDCDDDC	1	0.0196
CD_DDDDD	1	0.0179	CCDCCDCC	1	0.0196
CD_CDCDD	1	0.0179	CCCCDCCC	1	0.0196
CCDDDDDD	1	0.0179	CCCCCDD	1	0.0196
CCCCDDDD	1	0.0179	CC_DDDCC	1	0.0196
CCCCDDD	1	0.0179	CC_CDDCD	1	0.0196
CC_DDDCD	1	0.0179	CC_CDCDD	1	0.0196
CC_CCDDC	1	0.0179	CC_CCCCC	1	0.0196

We now explore in more detail the sequence of actions taken by subjects who changed their decision at least once in the eight sessions. Across the two  $N_C$  treatments and among the 12 subjects who change strategy only once, 8 of them (67%), move from C to D, while only 33% move from D to C, i.e., half of those who changed in the opposite direction. Fourteen subjects change their choice twice, accounting for 28 option changes. 43% of them go from C to D and back, and 57% do the opposite. As for the rest of the changes, a total of 75, 34 of them (45%) are switches from D to C, and the remaining 41 go from C to D. Thus, the only net effect of the changes seem to be in the subjects who changed only one, a behavior that could be a possible reason for the observed increment in cooperation with size. However, because of the small sample size, this should be read as a hint to be confirmed in future work, rather than a well established conclusion.

Finally, we also checked for potential gender effects, especially because previous research suggests that gender may be related to prosocial preferences in a non-trivial way. In particular, Croson & Gneezy<sup>56</sup> argue that women are more altruistic than men in the Dictator game, but that there are no stable gender differences in cooperation games such as the Prisoner's dilemma and the Public Goods game. On the contrary, Molina *et al.*<sup>57</sup> find some evidence supporting larger cooperativeness among female high school students. Two recent meta-analyses<sup>58,59</sup> found that women are more altruistic than men, especially when acting under intuition<sup>59</sup>, and that gender differences in cooperation, if existing, are likely to be very small.<sup>60</sup> In line with this general picture, also in our cooperation problem, we found no gender differences in cooperation. Specifically, we conduct a set of logistic regressions predicting  $C$  as a function of  $Female$ , for each value of  $N$  and also collapsing all group sizes together, finding all p-values to be larger than 0.1.

## Discussion

Here we have contributed to the ongoing debate regarding the effect of group size on cooperative behavior. We have experimentally investigated the possibility of observing nonlinear dependencies of cooperation with group size in public goods game with a critical mass. Our specific research questions related to the possibility, predicted theoretically, of observing a maximum in cooperation as a function of the group size around the critical mass at which additional cooperators do not lead to

increase public good production.

In contrast to this prediction, we observe a monotonous increase of the rate of cooperation as a function of the group size. Furthermore, our results suggest that the critical mass has no influence on the rate of cooperation. In addition, we have also looked at heterogeneity in individual behavior, finding approximately half of the subjects never contributing, another 10% that were unconditional cooperators, whereas the rest show a tendency to change preferentially from defection to cooperation than the converse, as the size of the group increases.

There are several reasons why we may have not observed an inverted-U effect of group size on cooperation in our experiment. One might have been selection bias: it is possible that subjects who dropped out in the middle of the experiment were high cooperators with respect to those who completed the experiment, and this might have killed the pick of cooperation at intermediate-size groups. However, as we have seen in the **Results** section, our analysis shows that subjects who dropped out cooperated at the same rate as those who completed the experiment, thus ruling out this possible explanation. Another possible reason for the discrepancy comes from a closer, data-driven, inspection of the results reported by Capraro and Barcelo<sup>45</sup>. Here, the maximum rate of cooperation was observed for  $N = 15$ . However Figure 1 in Capraro & Barcelo<sup>45</sup> highlights that the case  $N = 15$  gave rise to an unusually high rate of cooperation, suggesting that it might have been an outlier. Treating  $N = 15$  in Capraro & Barcelo<sup>45</sup> as an outlier results in a broad maximum around  $N = 25$ , that could only be noticed by going to groups as large as about 100 people. If the maximum rate of cooperation in public goods game with a critical mass is attained this far from the critical mass for the case  $N_C = 10$ , the situation could be further complicated in the case  $N_C = 20$ , in which the rate of cooperation could reach its maximum for group sizes not considered in our experiment — contributing to mask the effect. It is also possible that the payoff structure of the game affects the results, as has been observed, e.g., by Nosenzo *et al.*<sup>21</sup>, who observed no group size effect for low marginal per capita returns and negative effect of group size for high marginal per capita returns. Another potential explanation is that within-subject experiments give rise to fundamentally different dynamics than between-subjects ones, perhaps for the following reason. In order to have an inverted-U effect of the group size on cooperation in a within subject design, one would need a significant proportion of subjects that change strategy twice: from defection, to cooperation, and then back to defection. This somehow conflicts with the literature suggesting that subjects have preferences for being consistent<sup>61</sup>, especially when it comes to costly prosocial behavior.<sup>62</sup> In line with this view, our analysis of the heterogeneity in people's behavior shows that about 60% of the subjects never change strategy, and about 70% of the subjects change strategy only once. This might have partly contributed to the lack of an inverted-U effect of group size on cooperation. The discrepancy could perhaps be understood in terms of evolutionary game theory, because from that view point the direction and size of the effect could depend on external control parameters such as the mutation rate or the intensity of the selection<sup>5</sup>. Finally, reminding that the existence of an inverted-U effect was deduced by assuming that subjects play according to the cooperative equilibrium model<sup>46</sup>, it is possible that, in our setting, this model does not represent a good approximation of subjects' behavior. Although this might be true in general, it would not explain why we found a positive effect of group size on cooperation, while Capraro & Barcelo<sup>45</sup> found an inverted-U effect. Further experimental work, involving much larger number of subjects and groups, is needed to discriminate among these options and have a more complete understanding of if, when, and how group size impacts cooperative behavior.

## Methods

All participants in the experiments reported in the manuscript signed an informed consent to participate when enrolling in the IBSEN volunteer pool<sup>55</sup>. In agreement with the Spanish Law for Personal Data Protection, their anonymity was always preserved. This procedure was approved by the Ethics Committee of Universidad Carlos III de Madrid, the institution responsible for funding the experiment, and the experiment was subsequently carried out in accordance with the approved guidelines.

The experimental instructions were prepared following as closely as possible those used in<sup>45</sup>, albeit translated into Spanish. A screenshot of the decision page is shown in Fig. 4. Here we include a translated version of the instructions:

*“You are going to participate in an experiment with other people. You will play in groups of  $N$  people.*

*It will take you about 5 minutes. Each wave, your earnings will be shown in points, and only if you complete the 8 waves of the experiment, these earnings will be converted to money. Additionally, if you complete the eight sessions of the experiment and your total score (of the eight sessions) is among the 25% best, you will participate in a lottery for a single prize of 50 Euros.*

*Each session, you will have to decide to join either Group A or Group B. Your earnings will depend on the group you decide to join and on the size of the two groups, A and B, as follows:*

[Table 2 goes here.]

*Example:*

*If you join A, and the number of people that also decided to join A (size A) is 3 people, you will earn 15 points.*

*If you join B, and the number of people that decided to join A (size A) is 5 people, you will earn 35 points.*

## Su decisión

Tiempo disponible para completar esta página: ⌚ 4:41:09

### ¿A qué grupo quiere unirse?

Su decisión:

A  B

[Siguiente](#)

#### Instrucciones

Usted va a participar en un experimento junto con otras personas. Jugará en un grupo de  $n$  personas, incluido usted.

Estimamos que le llevará 5 minutos de su tiempo. Cada sesión acumulará unas ganancias expresadas en puntos. Al final del experimento, solo si usted completa las 8 sesiones, podrá ver los resultados de cada una de las sesiones y sus ganancias serán convertidas a dinero real. Además, si completa las 8 sesiones del experimento y su puntuación en el total de las ocho sesiones está entre las 25% mejores, participará en un sorteo de un premio único de 50 Euros.

Cada sesión, tendrá que decidir si unirse al grupo A o al grupo B. Sus ganancias dependerán de cuál es el grupo al que decida unirse, así como del número de personas que decidan unirse a uno u otro grupo, según se muestra en la siguiente tabla:

Tamaño del grupo A	Ganancias de cada integrante del grupo A	Ganancias de cada integrante del grupo B
0	0	10
1	5	15
2	10	20
3	15	25
4	20	30
5	25	35
6	30	40
7	35	45
8	40	50
9	45	55
10	50	60
> 10	50	60

#### Ejemplo:

Si decide unirse al grupo A, y el número de personas que deciden unirse a A, incluyendo a usted, (tamaño del grupo A) es 3, sus ganancias son 15 puntos.

Si decide unirse al grupo B, y el número de personas que deciden unirse a A (tamaño del grupo A) es 5, sus ganancias son 35 puntos.

Recibirá el pago de sus ganancias al final del experimento. Sus ganancias serán convertidas a dinero y transferidas a la cuenta Paypal que nos proporcionó al comienzo del experimento, en un plazo máximo de 7 días desde la finalización del experimento. Por favor, asegúrese de que la dirección de la cuenta Paypal que nos proporcionó es correcta, ya que será la dirección donde le haremos el pago. Si desea corregirla, escribanos por email antes de la finalización del experimento. No podremos repetirle el pago si nos proporcionó una dirección incorrecta.

Para su comodidad, estas instrucciones estarán a su disposición durante todo el experimento.

**Figure 4.** Screenshot of decision screen for the condition  $N_C = 10$ .

*You will be paid at the end of the experiment. Your earnings will be converted to money and transferred to your PayPal account, in a maximum period of 7 days counting from the end of the experiment. Please make sure the address of the Paypal account you provide us is correct, since it will be the address used for the payment. If you want to correct it, please write to us by email before the end of the experiment. We will not be able to repeat the payment if the account address is incorrect. For your convenience, these instructions will be available throughout the experiment.”*

At the beginning of each session, the instructions were shown, and afterwards participants had to complete four control questions to facilitate the understanding of the experiment. They had five opportunities to give the right answer, and afterwards, the correct answer was shown. The questions were:

- Q1: Imagine you decide to join group A, and the number of people that also decided to join A (size A) is 4 people. How many points will you earn?
- Q2: Imagine you decide to join group B, and the number of people that also decided to join A (size A) is 8 people. How many points will you earn?
- Q3: Imagine you decide to join group A, and the number of people that also decided to join A (size A) is 15 people. How many points will you earn?

**Table 2.** Payoff beta functions showed in the experiment:  $N_C = 10$  (left) and  $N_C = 20$  (right).

Size of Group A	Payoff of A	Payoff of B	Size of Group A	Payoff of A	Payoff of B
0	0	10	0	0	10
1	5	15	1	5	15
2	10	20	2	10	20
3	15	25	3	15	25
4	20	30	4	20	30
5	25	35	5	25	35
6	30	40	6	30	40
7	35	45	7	35	45
8	40	50	8	40	50
9	45	55	9	45	55
10	50	60	10	50	60
>10	50	60	11	55	65
			12	60	70
			13	65	75
			14	70	80
			15	75	85
			16	80	90
			17	85	95
			18	90	100
			19	95	105
			20	100	110
			>20	100	110

- Q4: Imagine you decide to join group B, and the number of people that decided to join A (size of A) is 20 people. How many points will you earn?

During the experiment, the only information available to the participants was the group size they were playing every time and the beta function (as a function of A and B decisions) presented by means of Table 2. Participants played online through a web browser in a computer, tablet or mobile phone; and could make their decisions during eight hours each session day. After each of the sessions, which were all one-shot game sessions, groups were randomly formed only with participants that made their decision in order to compute the payoffs. Participants membership to these randomly formed groups (of size  $N$ ) did not remain constant across sessions, but new random groupings were created after each session. For participants in incomplete groups, we assigned a payoff by randomly sampling among participants in the corresponding session that made the same decision (cooperate/not cooperate). The payoffs of all rounds were only shown at the end of the experiment.

## Accession codes

Data is available in an structured way at Zenodo public repository with DOI <http://doi.org/10.5281/zenodo.2565197>.

## References

1. Gintis, H. & Bowles, S. *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton University Press, Princeton, 2011).
2. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
3. Milinski, M., Semman, D., & Kranbeck, H.-J. Reputations helps solve the tragedy of the common. *Nature* **415**, 424–426 (2002).
4. Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. In search of the uniquely human. *Behav. Brain Sci.* **28**, 721–727 (2005).
5. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
6. Rand, D. G., & Nowak, M. A. Human cooperation. *Trends Cogn. Sci.* **17**, 413–425 (2013).

7. Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., Szolnoki, A. Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51 (2017).
8. Sánchez, A. Physics of human cooperation: experimental evidence and theoretical models. *J. Stat. Mech. Theory Exp.* **2**, 02401 (2018).
9. Boehm, C. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard University Press, Cambridge, 2000).
10. Molina, J.L., Lubbers, M.J., Valenzuela-García, H. & Gómez-Mestres, S. Cooperation and competition in social anthropology. *Anthropol. Today* **33**, 11–14 (2017).
11. Wellman, B. & Rainie, L. *Networked: The New Social Operating System* (MIT Press, Cambridge, 2012).
12. Olson, M. *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard University Press, 1965).
13. Hardin, R. *Collective Action* (Johns Hopkins University Press, Baltimore, 1982).
14. Chamberlin, J. Provision of public goods as a function of group size. *Am. Polit. Sci. Rev.* **68**, 707–716 (1974).
15. Oliver, P.E. & Marwell, G. The paradox of group size in collective action: A theory of the critical mass, III. *Am. Sociol. Rev.* **53**, 1–8 (1988).
16. Dawes, R., McTavish, J., & Shaklee, H. Behavior, communication, and the assumptions about other people's behavior in a commons dilemma situation. *J. Pers. Soc. Psychol.* **35**, 1–11 (1977).
17. Komorita, S. S., & Lapworth, W. C. Cooperative choice among individuals versus groups in an n-person dilemma situation. *J. Pers. Soc. Psychol.* **42**, 487–496 (1977).
18. Baland, J. M., & Platteau, J. P. The ambiguous impact of inequality on local resource management. *World Dev.* **27**, 773–788 (1999).
19. Whelan, S.A. Group size, group development, and group productivity. *Small Group Research* **40**, 247–262 (2009).
20. Grujić, J., Eke, B., Cabrales, A., Cuesta, J. A., Sánchez, A. Three is a crowd in iterated prisoner's dilemmas: Experimental evidence on reciprocal behavior. *Sci. Rep.* **2**, 638 (2012).
21. Nosenzo, D., Quercia, S. & Sefton, M. Cooperation in Small Groups: The Effect of Group Size. *Exp. Econ.* **18**, 4–14 (2015).
22. Barcelo, H. & Capraro, V. Group size effect on cooperation in one-shot social dilemmas. *Sci. Rep.* **5**, 7937 (2015).
23. McGuire, M. C. Group size, group homogeneity, and the aggregate provision of a pure public good under Cournot behavior. *Public Choice* **18**, 107–126 (1974).
24. Isaac, M. R., Walker, J. M. & Williams, A. W. Group size and the voluntary provision of public goods. *J. Public Econ.* **54**, 1–36 (1994).
25. Haan, M., & Kooreman, P. Free riding and the provision of candy bars. *J. Public Econ.* **83**, 277–291 (2002).
26. Agrawal, A., & Chhatre, A. Explaining success on the commons: Community forest governance in the Indian Himalaya. *World Dev.* **34**, 149–166 (2006).
27. Masel, J. A. A Bayesian model of quasi-magical thinking can explain observed cooperation in the public good game. *J. Econ. Behav. Organ.* **64**, 216–231 (2007).
28. Zhang, X. Q., & Zhu, F. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Am. Econ. Rev.* **101**, 1601–1615 (2011).
29. Szolnoki, A., & Perc, M. Group-size effects on the evolution of cooperation in the spatial public goods game. *Phys. Rev. E* **84**, 047102 (2011).
30. Capraro, V., & Halpern, J. Y. Translucent players: Explaining cooperative behavior in social dilemmas. *Proceedings of the 15th Conference in Theoretical Aspects of Rationality and Knowledge* (2015).
31. Mao, A., Mason, W., Suri, S. & Watts, D. J. An Experimental Study of Team Size and Performance on a Complex Task. *PLoS ONE* **11**, e0153048 (2016).
32. Gautam, A. P. Group size, heterogeneity and collective action outcomes: Evidence from community forest in Nepal. *Int. J. Sustain. Dev. World Ecol.* **14**, 547–583 (2007).
33. Pecorino, P., & Temimi, A. The group size paradox revised. *J. Public Econ. Theory* **10**, 785–799 (2008).
34. Rustagi, D., Engel, S., & Kosfeld, M. Conditional cooperation and costly monitoring explain success in forest commons management. *Science* **330**, 961–965 (2011).

35. Ledyard, J.O. Public goods: A survey of experimental research, in *Handbook of Experimental Economics*, edited by J. Kagel and A. Roth (Princeton University Press, 1995).
36. Weimann, J., Brosig-Koch, J., Hennig-Schmidt, H., Keser, C. & Stahr, C. Public good experiments with large groups. FEMM Working Papers 120009, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management. Available at <https://ideas.repec.org/p/mag/wpaper/120009.html> (2012). (Date of access: 11/11/2018).
37. Zelmer, J. Linear public goods experiments: A meta-analysis. *Exp. Econ.* **6**, 299–310 (2003).
38. Capraro, V., Jordan, J. J., & Rand, D. G. Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Sci. Rep.* **4**, 6790 (2014).
39. Engel, C., & Zhurakhovska, L. When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives. *Applied Econ. Lett.* **23**, 1157–1161 (2016).
40. Peña, J. Group size diversity in public goods games. *Evolution* **66**, 623–636 (2011).
41. Peña, J. Group size effects in social evolution. *J. Theor. Biol.* **457**, 211–220 (2018).
42. Gokhale, C. S., & Traulsen, A. Evolutionary games in the multiverse. *Proc. Nat'l. Acad. Sci. USA* **107**, 5500–5504 (2010).
43. Han, T.A., Traulsen, A., & Gokhale, C. S. On equilibrium properties of evolutionary multi-player games with random payoff matrices. *Theor. Popul. Biol.* **81**, 264–272 (2012).
44. Duong, M. H., & Han, T. A. Analysis of the expected density of internal equilibria in random evolutionary multi-player multi-strategy games. *J. Math. Biol.* **73**, 1727–1760 (2016).
45. Capraro, V. & Barcelo, H. Group Size Effect on Cooperation in One-Shot Social Dilemmas II: Curvilinear Effect. *PLoS ONE* **10**, e0131419 (2015).
46. Capraro, V. A Model of Human Cooperation in Social Dilemmas. *PLoS ONE* **8**, e72427 (2013).
47. Poteete, A. R., & Ostrom, E. Heterogeneity, group size and collective action: The role of institutions in forest management. *Dev. Change* **35**, 435–461 (2004).
48. Agrawal, A., & Goyal, S. Group size and collective action – Third-party monitoring in common-pool resources. *Comp. Polit. Stud.* **34**, 63–93 (2001).
49. Agrawal, A. Small is beautiful, but is larger better? Forest management institutions in the Kumaon Himalaya, India. In: Gibson, C., McKean, M. A., & Ostrom, E. (Eds.) *People and Forests: Communities, Institutions, and Governance*. Cambridge, MA: MIT Press, 57–85 (2000).
50. Yang, W., Liu, W., Viña, A., Tuanmu, M.-N., He, G., & Dietz, T. Nonlinear effects of group size on collective action and resource outcomes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10916–10921 (2013).
51. Cinner, J. E., MacNeil, M. A., Basurto, X., & Gelcich, S. Looking beyond the fisheries crisis: Cumulative learning from small-scale fisheries through diagnostic approaches. *Global Environ. Change* **23**, 1359–1365 (2013).
52. Kenna, R., & Berche, B. Critical mass and the dependency of research quality on group size. *Scientometrics* **86**, 537–540 (2011).
53. Kenna, R., & Berche, B. Critical masses for academic research groups and consequences for higher education research policy and management. *High. Educ. Manag. Pol.* **23**, 9–29 (2011).
54. Kenna, R., & Berche, B. Managing research quality: Critical mass and academic research group size. *J. Manag. Math.* **23**, 195–207 (2012).
55. IBSEN (Bridging the gap: from Individual Behaviour to the Socio-technical Man) project webpage. <http://www.ibsen-h2020.eu> (last accessed November 19, 2018).
56. Croson, R., & Gneezy, U. Gender differences in preferences. *J. Econ. Lit.* **47**, 448–474 (2009)
57. Molina, J. A., Giménez-Nadal, J. I., Cuesta, J. A., Gracia-Lázaro, C., Moreno, Y. & Sánchez, A. Gender differences on cooperation: Experimental evidence on high school students. *PLoS ONE* **8**, e83700 (2013).
58. Brañas-Garza, P., Capraro, V., & Rascón-Ramírez, E. Gender differences in altruism on Mechanical Turk: Expectations and actual behavior. *Econ. Lett.* **170**, 19–23 (2018).
59. Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. Social heuristics and social roles: Intuition favors altruism for women but not for men. *J. Exp. Psychol. Gen.* **145**, 389–396 (2016).
60. Rand, D. G. Social dilemma cooperation (unlike Dictator Game giving) is intuitive for men as well as women. *J. Exp. Soc. Psychol.* **73**, 164–168 (2017).

61. Cialdini, R. Influence, the Psychology of Persuasion. New York: Harper Collins (1984).
62. Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. Paying to be nice: Consistency and costly prosocial behavior. *Manag. Sci.* **58**, 179-187 (2012).

## **Acknowledgements**

This work was partially supported by MINECO/FEDER (Spain) through grant FIS2015-64349-P VARIANCE (AS) and by the EU through FET Open Project IBSEN (contract no. 662725, AS).

## **Author contributions statement**

MP, VC, and AS conceived the original idea for the experiment; MP wrote the software interface for the experiment and carried out the experiment; MP and VC analyzed the data; MP, VC, and AS discussed the analysis results and wrote the paper.

**Competing interests.** The authors declare no competing interests.