Agencies method for coalition formation in experimental games

John F. Nash[a,1], Rosemarie Nagel[b], Axel Ockenfels[c], and Reinhard Selten[d]

[a]Princeton University, Princeton, NJ 08544; [b]Institució Catalana de Recerca i Estudis Avançats, Department of Economics and Business, Universitat Pompeu Fabra, and BGSE, 08005 Barcelona, Spain; [c]University of Cologne, D-50923 Cologne, Germany; and [d]University of Bonn, D-53113 Bonn, Germany

Contributed by John F. Nash, September 20, 2012 (sent for review April 22, 2012)

In society, power is often transferred to another person or group. A previous work studied the evolution of cooperation among robot players through a coalition formation game with a noncooperative procedure of acceptance of an agency of another player. Motivated by this previous work, we conduct a laboratory experiment on finitely repeated three-person coalition formation games. Human players with different strength according to the coalition payoffs can accept a transfer of power to another player, the agent, who then distributes the coalition payoffs. We find that the agencies method for coalition formation is quite successful in promoting efficiency. However, the agent faces a tension between short-term incentives of not equally distributing the coalition payoff and the long-term concern to keep cooperation going. In a given round, the strong player in our experiment often resolves this tension approximately in line with the Shapley value and the nucleolus. Yet aggregated over all rounds, the payoff differences between players are rather small, and the equal division of payoffs predicts about 80% of all groups best. One reason is that the voting procedure appears to induce a balance of power, independent of the individual player's strength: Selfish subjects tend to be voted out of their agency and are further disciplined by reciprocal behaviors.

~~rules | reciprocity | fairness | institution~~

The evolution of human altruism and cooperation is a puzzle. Unlike other animals, people frequently cooperate even absent of any material or reputational incentive to do so. In this paper we show how a voting procedure to transfer power to another person successfully promotes cooperation by balancing the tension between short-term incentives to defect and long-term incentives to keep cooperation going. Our work is inspired by John Nash (1), who theoretically studied the evolution of cooperation among robot players through acceptance of an agency of another player.

Beyond Nash's (1) work, there is virtually no work on the agencies method in (experimental) economics as we apply it in our paper. The underlying idea is simple and important: Human subjects can transfer the power to an agency, who determines the final payoff distribution within the group.* Our game reflects that, often, efficiency requires people's willingness to accept the agency of others, such as political, social, or economic leaders (for voting of an expert, see ref. 7).

In Nash's (1) work, the robots employed optimal strategies, being the computational result of complex systems of equations. Motivated by Nash's paper, we study laboratory three-person coalition formation games with a noncooperative procedure of acceptance of an agency of another player. The base games are finitely repeated for 40 rounds with the same three subjects, allowing cooperation and coordination to evolve. In our games noncooperative game theory cannot organize behavior because it is basically consistent with any outcome. Thus, even the strategies of fully rational agents cannot be predicted by the theory.† We show, however, that the solution concepts of cooperative game theory together with the equal split solution provide some structure on the emergence of cooperation in our experiment. Yet understanding how cooperation is affected by decisions to transfer power to others requires theories that go beyond these approaches.

More specifically, our model specifies the coalition formation process in extensive form (for more details see *Methods* and Fig. 5). It consists of a coalition formation phase, and second phase in which the final agent distributes the coalition value. A given characteristic function specifies a value for all possible coalitions (Table 1 shows the 10 three-person characteristic function games used in the experiment). In phase 1 each player of a group of three can accept at most one other player as an agent to form a pair. In case nobody accepts, the phase is repeated until a pair is formed or a random break-up rule leads to zero payoffs for all. If one pair is formed, the accepting player becomes inactive and is represented by the accepted player who enters phase 2 together with the remaining player. In case more than one pair is formed in phase 1, a random draw decides which pair is decisive for the next step. In phase 2, each of the two active players has to decide whether to accept the other active player. In case no one accepts, the stage is repeated until a player accepts or a random break-up rule selects the pair of stage 1 as the final coalition. In the latter case the accepted player of stage 1 divides the value of his two-person coalition. In case a player accepts in the second stage, the accepted player divides the three-person coalition value among the three players. If there are two accepted players in the second stage, a random draw selects which of the two players can divide the coalition value.

For a long time the focus of attention in the analysis of coalition formation has been cooperative game theory. One underlying idea of cooperative game theory is that there are no restrictions on how agreements can be reached among players. The coalition formation process, making offers and counteroffers, can thus remain largely unspecified. In our cooperative games it is to the joint benefit of the group to form the grand coalition. Yet, which allocation of the corresponding coalition value is agreeable to all players? We mostly focus on three solution concepts of cooperative game theory: the core, Shapley value, and nucleolus.

Author contributions: J.F.N., R.N., A.O., and R.S. designed research; R.N. and A.O. performed research; R.N., A.O., and R.S. analyzed data; J.F.N., R.N., A.O., and R.S. contributed to the theory and the supporting material; and J.F.N., R.N., A.O., and R.S. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the ■■■ database (accession nos. ■■■).

[1]To whom correspondence should be addressed. E-mail: xkjfnj@princeton.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216361109/-/DCSupplemental.

*Unlike in our case, where there is no precommitment to a particular policy prior to voting, most models in the theoretical voting literature assume that candidates are fully committed to their campaign policy proposals. Thus, by assumption, when a candidate is elected, he implements the policies that he promised to his constituency during the campaign. There are a few papers that drop the assumption of full commitment and analyze the strategic policy choice of candidates after they are elected. In one-shot elections the only possible outcome is the implementation of the most preferred policy of the winning candidate (2). In repeated elections, the value of reputation allows candidates to make credible policy proposals in equilibrium (3). In the experimental literature an important exception is ref. 4; see also *Summary and Conclusion*. A related literature on voting experiments can be found in ref. 5 with experiments on voting for fully committed candidates, the voting paradox, and also some experiments on the issue of voting over redistribution; e.g., a proposal by one or more players, by the experimenter, or through the rules of the game is either accepted or dismissed (6).

†See ref. 8 for the history of experimental testing of game-theoretic hypotheses and different approaches to model human behavior (9–14). Unlike in the context of cooperation and bargaining base games with multiple equilibria, the evolution of behavior in simple coordination games appears to be relatively well understood since Schelling's (15) seminal work on coordination problems (refs. 16–18 and references therein). There is also a considerable literature on how reputation-building institutions may affect cooperation (refs. 19–21 and references therein).

PNAS proof
Embargoed

Table 1. Characteristic functions, nucleolus, and Shapley values for the 10 games used in the experiment

Characteristic function games				Nucleolus			Shapley value		
No.	*v*(*AB*)	*v*(*AC*)	*v*(*BC*)	Pay *A*	Pay *B*	Pay *C*	Pay *A*	Pay *B*	Pay *C*
1*	120	100	90	53.33	43.33	23.33	46.67	41.67	31.67
2*	120	100	70	66.67	36.67	16.67	53.33	38.33	28.33
3*	120	100	50	80.00	30.00	10.00	60.00	35.00	25.00
4*	120	100	30	93.33	23.33	3.33	66.67	31.67	21.67
5*	100	90	70	56.67	36.67	26.67	48.33	38.33	33.33
6†	100	90	50	70.00	30.00	20.00	55.00	35.00	30.00
7‡	100	90	30	83.33	23.33	13.33	61.67	31.67	26.67
8‡	90	70	50	60.00	40.00	20.00	50.00	40.00	30.00
9‡	90	70	30	72.50	32.50	15.00	56.67	36.67	26.67
10§	70	50	30	57.50	37.50	25.00	50.00	40.00	30.00

In columns 2, 3, and 4 we state for each of the 10 games the ~~payoffs~~ for the two-person coalitions, *V*(*XY*) = *V*(*AB*), *V*(*AC*), and *V*(*BC*), respectively, of the characteristic function; the three-person coalition (grand coalition) is always 120, and the one-person coalition is normalized to 0. Columns 5, 6, and 7 present the theoretical payoffs for players *A*, *B*, and *C*, respectively, according to nucleolus, and columns 8, 9, and 10 present those for the Shapley value in a one-shot cooperative game. These theoretical payoffs are always distributions of the grand coalition value. Player *A* is the strong player in all games, because with him the highest two-person coalition payoffs can be achieved compared to a coalition without him. For a similar reason *B* is the second strongest player. In all theoretical cooperative solutions player *A* receives always the highest payoff, followed by player *B*.

*Games where no core exists.

†Core has unique solution equal to nucleolus.

‡Core is multivalued with equal split outside the core.

§Core is multivalued with equal split inside the core; for core area, see Fig. 2.

Loosely speaking, divisions of the total return are called points of the core if they are stable in the sense that no coalition should have the desire and power to upset the agreement. If, e.g., a coalition of two members is assigned a smaller total payoff in the grand coalition proposal than what the two members coalition can achieve alone, this proposal is inherently instable and thus not in the core. However, one problem is that the core can consist of many points without distinguishing a preferred point, or it may even be empty. The Shapley value (22), on the other hand, assigns to each player a unique payoff ("value"), which may be interpreted as a measure of power of the respective player in the game. One way of arriving at the Shapley value is to suppose that the grand coalition is formed by each player entering into this coalition one by one. As each player enters, he receives a payoff equal to his marginal contribution to the grand coalition payoff. This contribution generally depends on the entering order. The Shapley value is the average payoff to the players if they enter in a random order.[‡] The nucleolus (23) is neither based on a stability concept like the core nor characterized by principles of power or fairness like the Shapley value. Rather it finds a unique solution of a cooperative game by computing the maximum dissatisfaction with a given allocation across all possible coalitions and then finding the allocation that minimizes the maximum dissatisfaction. So, loosely speaking, the nucleolus serves the most dissatisfied players first. (Dissatisfaction is measured by the maximum total payoff a coalition can reach minus the actual total payoff assigned by the allocation. The nucleolus is always in core, if nonempty.)

We discuss the performance of these solution concepts for our games in the next section (see Table 1 and Fig. 2 for the quantitative solutions for our base games and the theoretical discussion of the solution concepts in *SI Text A*).

Built on Nash's seminal papers in 1950 (24), since the 1980s there have been many attempts to understand coalition formation and the distributions of payoffs as equilibria of noncooperative games (25–27 and references therein). Starting in the field of industrial organization, extensive game models of oligopolistic competition and the analysis of their subgame perfect equilibria (28, 29) turned out to be a fruitful approach. Harsanyi's theory of incomplete information (30) opened further opportunities for noncooperative game models. Somewhat later noncooperative game modeling spread to many other fields of economic theory and much less attention was paid to cooperative games. There are exceptions. One is the "Gale–Shapley algorithm" (31) that rather recently turned out to be useful in practical market design; Roth (32) surveys the literature. However, the earliest attempt to develop noncooperative modeling of cooperation and bargaining was by Nash (33) on two-person bargaining. There he not only presented his axiomatic theory but also offered a noncooperative interpretation. Both Nash's model (1) and our agencies method ultimately build on this noncooperative approach.

A description of a strategic situation as a noncooperative game is much more detailed than for a cooperative game. For experimental purposes an extensive game procedure for coalition bargaining, as we have devised it, has the advantage that the players interact in a formal and anonymous way. Thereby one isolates the strategic situation from social influences like personal sympathies and easily protocols every decision. However, any coalition (including no coalition with zero payoffs for all) can be supported in a pure equilibrium of our base game. The final agent who can be any of the three players takes the entire coalition value. When repeating the base game, as in our experiment, noncooperative game theory imposes even less structure on behavior and outcomes: In the super-game almost any payoff division can be chosen in equilibrium, supported by a threat to convert to a one-shot base game equilibrium with no acceptances of any player in case of a deviation from the equilibrium path (*SI Text A*). Thus, it is not possible to derive predictions from noncooperative game theory. [The complexity involved in analyzing supergames was first emphasized by Nash in the context of the theoretically much less demanding repeated prisoner's dilemma (34) and has been confirmed in the laboratory (35–37).]

We find that the agencies method for coalition formation is an effective mechanism to promote efficient cooperation and balanced payoffs. In particular, we observe that even though the players' strengths differ, long-run payoffs aggregated over all rounds tend to converge to the equal division. This is consistent with Nash's (1) simulations with robots in a similar context, as well as with parts of the behavioral economics literature, indicating a general attraction for payoff equality in bargaining and cooperation games, especially when payoff comparisons between players are possible.[§]

However, equality is not generally the leading principle in each round separately. Here, many agencies seem to succumb to short-term incentives and allocate a significantly larger portion of the payoffs to themselves. Yet, they generally resist taking the whole surplus, as would have been predicted by noncooperative game theory for the one-shot version of the game. Rather, in the short run, many agents appear to be guided by strength comparisons as captured by the Shapley value and the nucleolus. So, although there is inequality at many snapshots, some of which is organized by cooperative game theory, the repeated and symmetric voting procedure makes sure that, ultimately, everybody is taken care of equally—even to the extent that it mitigates strong differences in the subjects' strengths.

‡The values can also be characterized by the Shapley axioms of fairness: (*a*) the sum of all players' values is the grand coalition value, (*b*) players contributing equally to any coalition have the same value, (*c*) players neither harming nor helping any coalition have a zero value, and (*d*) the value of two games played at the same time is equal to the sum of values played at different times. Note that the Shapley value is not necessarily in the core, even if the core is nonempty.

§See also, e.g., the role of equity considerations in coalition bargaining (38) and a survey of recent evidence in various noncooperative games (39). We caution that, because unlike Nash we implement a finite multistage game, Nash's simulation results are only suggestive for our data analysis, which is why we complement our ~~analyses~~ with hypotheses from cooperative and noncooperative game theory.

PNAS proof
Embargoed

Experimental Results

Each of the 10 games in Table 1 was played by 10 different groups of three. The same group interacted for 40 ~~periods~~, and all players maintained their strengths (associated with player roles A, B, and C) throughout. This procedure gives us 10 independent observations per game. We start our analyses with results on voting behavior, before studying payoff consequences. We generally do not find significant differences in behavior and outcomes across games. Thus, most of the time, we pool the data. If there are differences, we report them.

Coalition Building. Overall, cooperation is very successful. Only 1% of all 4,000 rounds end in no agreement, 7.5% in two-person coalitions, and 91.5% in the grand coalition.[¶] The high level of efficiency appears to be only to a small extent the result of the low exogenously given probability of conflict (10%) in case of an impasse, because the random mechanism (see Fig. 5, steps 4 and 10) was rarely used: An impasse occurred only in 5% of all possible cases in stage 1, of which in 75% voting was repeated only once or twice. Stage 2 saw more deadlocks, but still the random mechanism was used only in 14% of the rounds and at most twice in 8%. Although neither cooperative nor noncooperative game theory predicts how a grand coalition can emerge, one might speculate that the key to successful cooperation is a commonly accepted, stable agency.[‖] However, although player A has some advantage,[**] in no group is the same player always the final agent; in 14/100 groups, the same player (A in 8 of these groups) is the final agent 80% or more of the time. In 50 groups the same player (A in 27 of these groups) becomes the final agent between 50% and 80% of the time.

We conclude that the voting procedure does not strongly and consistently discriminate according to the players' strengths. Rather, the symmetric, procedural aspect of coalition formation in our game gives all players a chance in creating and leading the grand coalition, whether weak or strong.

Payoff Distribution. ***Equal splits.*** The equal split in the grand coalition has considerable attraction. We observe it in 54% of the 4,000 rounds. (As a contrast, in 3% we observe that a player takes all.) Fig. 1 shows the distribution of the 100 groups according to the number of times with exact equal splits of the 40 rounds. (Figs. S2 and S3 give a more detailed picture of both the evolution and the distribution of payoff distributions for each group.) In 29 of the 100 groups, the final agents chose the equal split between 36 and 40 times and thus are homogenous groups with respect to payoff distributions.[††] For the remaining 71 groups Fig. 1 reveals large differences in equal-splits usage. The next paragraphs investigate the guiding behavioral principles in these remaining groups.

[¶]There is no significant difference between the frequency of two-person coalitions in games 1–4 (~~146/301~~ rounds with two-person coalitions), with the two-person coalition implying full efficiency ($v(A, B) = 120$), and the frequency in games 5–10 (151 of 301), when full efficiency is reached only in the grand coalition (using the Mann–Whitney U test, using each individual group as an independent observation, $P > 0.12$). In 32% of all 100 groups there are never two-person coalitions, and in 18% it happens only once per group.

[‖]Recall, however, that noncooperative game theory predicts that, in every pure equilibrium of the base game where one player is voted for, this player will be voted for by either one or two players but the voted player will never vote for another player. In our supergame, we find that players A do not vote in 40% of the cases when they are voted for by either or both of the other players, whereas players B and C do not vote in 33% and 30% of the cases, respectively, with no significant differences between the players (sign tests based on group level with $P = 0.38$, 0.31, and 0.18 for the three comparisons).

[**]Over all groups, player A is most often the final agent in 42% of all rounds [16.7 rounds (SD = 10.4)], player B in 32% [12.8 rounds (SD = 9.1)], and player C in 25% [10.1 rounds (SD = 8.6)], and in 1% no coalition is formed. Rank ordering A, B, and C players in a group by number of being representative, there is a significant difference between the three players (Friedman's test, two-way analysis on ranks, $P < 0.01$).

[††]Twenty-three of these 29 groups choose the equal split from the very beginning, whereas only 6 groups manage to converge to the equal-split norm when the first three rounds are not equal-split proposals (see ref. 40 for how norms may emerge in competitive environments). However, even when the equal split is the start-off norm, 20 groups fail to maintain it throughout.

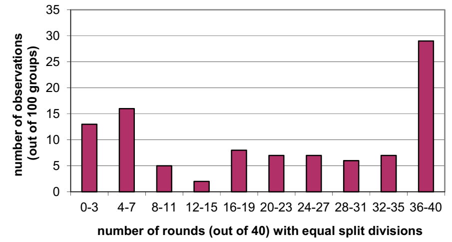


Fig. 1. Frequency of groups categorized according to the number of rounds of equal-split divisions. For each group we count how many times across all 40 rounds the final agreement is according to the equal split. We observe that about 30% of all groups use equal split in more than 35 rounds (right-hand column), whereas almost 15% use this rule in less than 4 rounds (left-hand column).

Payoff distributions in non~~pure~~ equal-split groups. Whereas noncooperative game theory is consistent with basically any payoff distribution in our games, cooperative game theory predictions are not. Our results demonstrate that some solution concepts from cooperative game theory can structure the data of the 71 nonhomogeneous groups in some (qualitative) ways. We concentrate on three concepts, the Shapley value, the nucleolus, and the core. We also consider the equal split.

The average payoff distributions across all rounds are typically closest to the equal split (72%, 51 groups), whereas 25% (18 groups) distribute near the Shapley value and 3% (2 groups) near the nucleolus (Fig. S1).[‡‡] The core distribution has predictive power only in game 10 with all theoretical solutions, the equal split, and all data points lying in the core. No actual payoff distribution of the other games lies, however, in the core. None of the payoff splits are close to the corners of the triangles, the selfish splits.

On a more qualitative level, we find that in 37% of all groups the payoffs are ordered as suggested by the cooperative solutions concepts, with player A earning the highest payoff and C the lowest (a random order would produce 17% of this particular order). In 24% of all groups B receives the highest payoff and in 7%, C. Further, in 32% A receives the highest payoff, but the order is reversed between B and C or one of these gets the same as A. Thus there is a significant difference between the payoff ordering between the three players, using the Friedman test ($P < 0.001$), with each group as an independent observation.

So far we have compared aggregate payoff distributions over all rounds with cooperative concepts and the equal split. However, it is also interesting to compare the theoretical concepts with the average proposal of the final agent. In all rounds the final agent either splits the coalition payoff in his favor or uses an equal split. Thus, the proposals of players B and C are inconsistent with the strengths of the players. Therefore it makes sense to compare only the proposal of the strong player, A, with the theoretical concepts. Fig. 2 shows the average payoff distributions proposed by A in each group and game with the theoretical solutions and equal split. Unlike aggregate comparisons over all rounds, the equal split best explains A's average payoff distribution 34% of the time, whereas the Shapley value reaches 40% and the nucleolus, 23%. The corner solution, with A earning (close to) 120, is the best descriptor 4% of the time.

Summing up, overall payoff differences mirror differences in the players' strengths and are thus qualitatively captured by cooperative concepts. In particular, focusing on the divisions by the strong player, outcomes are better organized by the Shapley value than by equality and selfishness. [Kahan and Rapoport (41) and references therein summarize many of the experiments that

[‡‡]Our measure of success for a particular prediction is the mean squared error (MSE) between the payoff vector of the coalition formation solution and the actual average payoff vector of a group. The theoretical concept having the smallest MSE to the actual average data best describes a particular group on average.

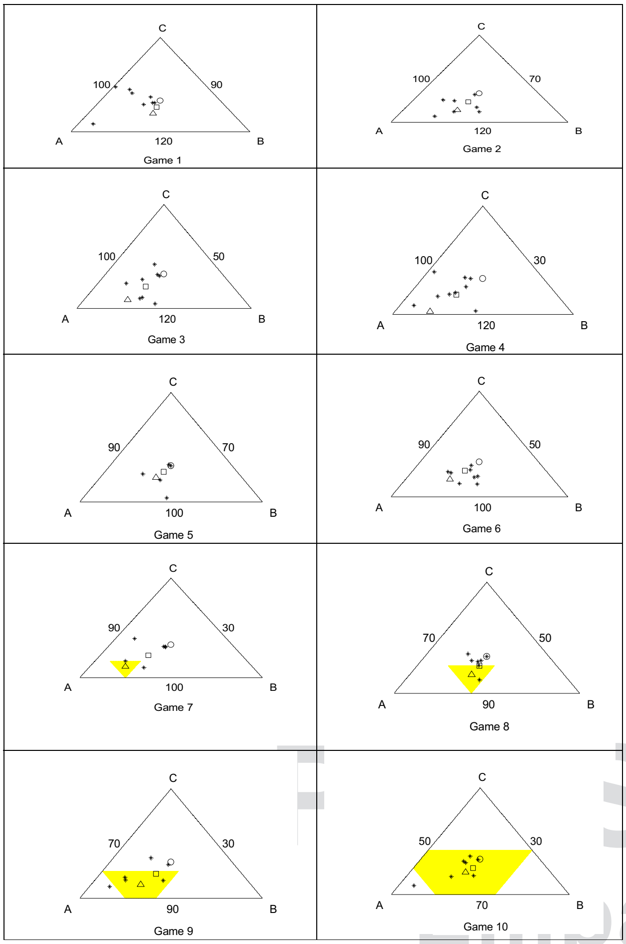


Fig. 2. Simplex for each game: Average payoff distribution for each group and game (stars), when *A* is agent and in non-equal-split groups (equal split in less than 36 rounds), and theoretical cooperative solutions (nucleolus, triangle; Shapley value, square; core area, in yellow; and equal split, circle). For all non-equal-split groups ~~(those who divide in less than 36 rounds according to equal split)~~, we compute the average payoff distribution when player *A* is agent for each group to compare his proposal with the theoretical concepts. The Shapley value and nucleolus assign the highest payoff to the strong player *A*. Furthermore, we plot the core, which is an area theory, allowing for several different splits (not in games 1–5, when the core does not exist, and in game 6 where the core is identical to the Shapley value). The equal split is the center of the triangle as it gives the same payoff to all members of a group. Because every agent within a group assigns himself at least the equal-split payoff, it makes no sense to compare the theoretical outcomes to proposals of *B* and *C* as these would be closest to equal split.

competitively test several solution concepts; interestingly, in this literature, the Shapley value is generally not supported.] Nucleolus also organizes a significant share of observations well. The core does not have predictive value. However, over all rounds the equal split outperforms other principles of behavior.

Reciprocity. How can average payoff vectors chosen by *A* be successfully organized by the Shapley value and, partly, the nucleolus, whereas equality is the dominant principle across all rounds and players? Before starting the experiment, our hypotheses were guided by the simulation in Nash (1) that cooperation can emerge only when no demand is selfish or when a selfish demand is matched with forgiving play. As mentioned above, 29 groups of the 100 groups showed equal splits across most rounds. Yet, in the remaining 71 groups our voting procedure produces just the opposite pattern: The more aggressive the demand of one player is, the more aggressive are those of the others. This kind of reciprocity is possible, because bargaining strength is offset by the voting procedure as shown below.

Fig. 3 illustrates the reciprocal behavioral pattern among all three players. The strong positive relationship indicates that the gifts of the three players in a group, made across rounds, are positively correlated. The Spearman rank correlation coefficient of mutual payoff gifts across all rounds is 0.50 [$P < 0.0001$~~; rounds 1–20 (0.54); rounds 21–40 (0.35, $P < 0.01$)~~] for *A* and *B*; 0.75 [$P < 0.0001$~~; (0.67); (0.67)~~] for *A* and *C*, and 0.57 ($P < 0.0001$~~; (0.67); (0.54)~~] for *B* and *C*. We ~~thus~~ also observe that these correlations are high from the very beginning and do not increase over time. An analogous analysis regarding each player's demands for himself when being the final agent yields very similar patterns. The Spearman rank correlation coefficient across all rounds is 0.24 [$P < 0.05$ ~~(rounds 1–20, 0.35; $P < 0.005$); (rounds 21–40, 0.3; $P < 0.02$)~~] for *A* and *B*, 0.41 [$P < 0.001$~~; (0.27; $P < 0.03$); (0.43; $P < 0.001$)~~] for *A* and *C*, and 0.30 [$P < 0.02$~~; (0.37, $P < 0.005$); (0.32, $P < 0.05$)~~] for *B* and *C*. T~~hus,~~ this coefficient increases over time ~~only~~ for *A* and *C* demands.

The reciprocal relationship between gifts and demands as revealed by the correlations shows that payoff mitigation is made possible through a "fair" voting mechanism that disciplines too selfish demands. In particular, Fig. 4 illustrates the negative relationship between average payoff demand and the number of times being the final agent [the Spearman rank correlation coefficient is −0.24 ($P < 0.05$) for *A*, −0.26 ($P < 0.05$) for *B*, and −0.36 ($P < 0.005$) for *C*].

Summing up, whereas the strong player's behavior is better organized by the Shapley value and, partly, by the nucleolus, reciprocity explains the strong prominence of the equal split in the aggregate. The three players mimic each other, so that both gifts and demands are highly correlated between players.

Summary and Conclusion

The agencies method by Nash (1) is very effective in promoting human cooperation and fair outcomes: Full efficiency is almost always reached in our laboratory coalition formation game, and the divisions of payoffs across rounds are much less extreme than one might expect from a noncooperative analysis of the base game. The tension between short-term incentives of not sharing the coalition value with others and the long-term concern to keep cooperation going is, by the strong player, often solved approximately in line with the Shapley value and the nucleolus. Also, the players' average payoff differences reflect the different strengths of players as measured by these concepts. However, over all rounds the payoff differences are rather small, and the equal division is the concept best describing 80% of all average payoff vectors. One reason is that the symmetry of the voting procedure induces a balance of power: Selfish agents tend to be voted out of

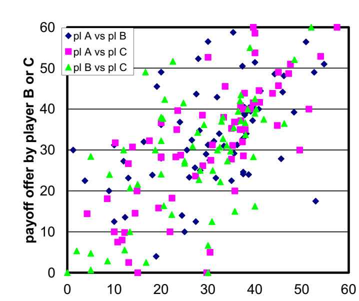


Fig. 3. Average payoff gifts within a group, pairwise comparing between the three players (only non-equal-split groups equal the equal split in less than 36 rounds). Each blue diamond is the average payoff offer from *A* to *B* and vice versa of a specific group, when they are agents. Thus the diamond coordinate (52, 18) means that agent *A* offers *B*, both from the same group, on average 52 and this agent *B* offers this same *A* 18. Each pink square shows offers between *A* and *C* and each green triangle shows offers between *B* and *C* of a particular group.

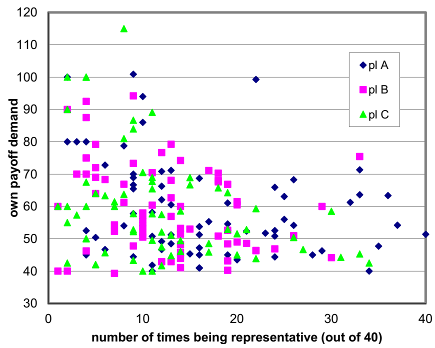


Fig. 4. Average demand and becoming the final agent per player (only non-equal-split groups ~~equal~~ the equal split in less than 36 rounds). Each blue diamond means how much on average a final agent *A* demands for himself and how many times he is voted as the final agent within the 40 rounds within a specific group; other symbols shown are similar for players *B* and *C*.

their agency and are disciplined by reciprocal behavior. In fact, all players have a good chance to become the final agent. As a result, even if the short-run round payoffs are dispersed, long-run average payoffs tend to converge.

We use the noncooperative approach to clearly define and control the coalition formation process. Yet noncooperative theory does not structure the behavior as the base game solutions are inconsistent regarding final payoffs and voting behavior. This complements earlier research in one-shot characteristic function games, where a great number of different extensive game procedures have been employed [see, in particular, the work on demand commitment models (41–44)]. For instance, the noncooperative theoretical analysis of these procedures suggests that the results depend strongly on procedural details. In fact, however, human behavior depends less on such details than predicted. Humans often seem to analyze the situation more in the flavor of cooperative game theory, ignoring the strategic consequences of the specific procedures used (38). Similarly, in earlier work on repeated asymmetric cooperation games, behavior could not be explained by optimizing behavior but rather by fairness criteria and cooperative goals (37).

The cooperative solution concepts, on the other hand, can help us organize the payoff division data, but they do not capture the effect of the underlying institutions and procedures. Whereas the strength of the players captures some of the average payoff differences when the strong player is in charge, voting and long-run distribution behavior was essentially independent of the characteristic function. Here, the repeated voting procedure, which gives all an equal weight when transferring power to an agency, leads to rather equal total payoffs. This mitigating effect of the voting procedure is not captured by theory. (The distribution of power across subjects in our experiments—as is generally the case in experimental economics—was random, which may also contribute to the attractiveness of the equal split.)

We conclude that other approaches to modeling human cooperation and coalition formation are needed, models that take people's cognitive and motivational limits in dealing with institutions and other players seriously. In this connection, an interesting related experimental study is the "three-person cooperative game with no side payments" by Kalisch, Milnor, Nash, and Nering (4). This study is one of the first experimental economics studies of negotiation and characteristic function games. In one treatment (section IV of their paper), two players could vote for another player; yet a player attracting two votes could not choose the distribution but was automatically awarded 40 monetary units, whereas the other two lost 20 each (otherwise, all payoffs were zero). They observed, like we do, that in the long run players typically equalized payoffs. Sometimes this was accomplished by randomization and sometimes by sequential reciprocity ("if you vote for me, I'll vote for you").

In the same paper the authors suggested to investigate these two mitigating mechanisms in an asymmetric setting as a robustness check for their findings. Although our experiment differs in some other ways too, we implement asymmetric characteristic functions—and observe the same two basic mechanisms at work in the following sense. Randomizing can be interpreted as a fair procedure, because it equalizes expected payoffs in the (base) game, where a deterministic equal outcome is not feasible (45, 46). Voting in our experiment can similarly be qualified as a fair procedure, balancing negotiation power in an otherwise asymmetric situation, because "one man, one vote" levels the playing field for everybody independent of a player's strength. Q:21

Given the repeated structure both in our study and in ref. 4, reciprocity comes in as an additional, dynamic balancing mecha-

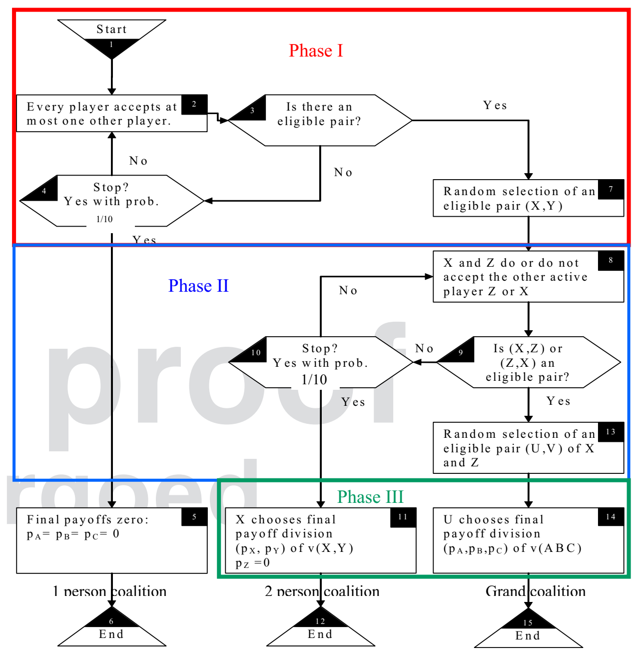


Fig. 5. Flowchart of the experimental stage game: Each stage game (within a round) consists of two voting phases and a distribution phase. *X*, *Y*, and *Z* denote players among *A*, *B*, and *C*, such that each of the three players *A*, *B*, and *C* receives exactly one of the names *X*, *Y*, and *Z*. Similarly *U* and *V* are the two players *X* and *Z* with $U = X$ and $V = Z$ or alternatively $U = Z$ and $V = X$. The small rectangles ~~describe~~ the steps of the process. A rhomboid represents a switch with two exits for answers Yes and No to the question inside the rhomboid. As in rhomboids 4 and 10 the answer may be a realization of a random event. The answers Yes and No are written above or at the right of the lines representing the exits from a switch. The start and the possible ends are represented by triangles. The arrows indicate the direction of flow. Specifically, after the start of the base game in a given round, players can accept at most one other player as one's agent (step 2 in phase I). An ordered pair (*X*, *Y*) is eligible, if *Y* has accepted *X* as his ~~or her~~ agent. If there is no eligible pair (switch 3), a random procedure decides whether the formation process stops, which happens with probability 1/10 (switch 4). "Stop" means a break-off of current negotiations on agent relationships and leads, at this point, to the normalized zero payoffs for all players (one-person coalition). If, however, there are one or more eligible pairs after step 2, one of them, (*X*, *Y*) is chosen randomly with equal probabilities (step 7), which ends phase I. A player is active, if he has no agent. After step 7 player *Y* is not active anymore; *X* and *Z* are the only active players left who enter phase II. Each of them accepts or not the other active player as his ~~or her~~ agent (step 8). Then, a procedure analogous to the one after step 2 starts. The process ends either with a two-person coalition with payoffs p_X and p_Y distributed by player *Y* such that $p_X \geq 0$ and $p_Y \geq 0$ as well as $p_X + p_Y = v(XY)$ hold (step 11; see Table 1 for the values) or with a three-person coalition with a payoff division (p_A, p_B, p_C) distributed by player *U* with nonnegative components and $p_A + p_B + p_C = v(ABC)$ = 120 (step 14). The symbol *v* denotes a superadditive zero-normalized characteristic function for the player set {*A*, *B*, *C*}. Superadditivity requires $v(ABC) \geq v(XY) \geq 0$ for every two-person coalition *XY*. Q:33

PNAS proof
Embargoed

nism, further reinforcing the convergence of power and payoffs in the groups. This suggests that the interaction of "fair institutions," such as voting, randomization, and reciprocity might be a key ingredient of the evolution of cooperation. However, it is captured neither by cooperative nor by noncooperative theory, and it has been rarely studied outside these two papers. (An exception is ref. 47, which assumes a role of "fair chance" in coalition games and on this basis applies a probabilistic choice model for light guessing behavior to coalition choice problems.) We hope that our findings and our framework inspire more research in this field.

Methods

Subjects. We invited 300 subjects, mostly economics students, into the Cologne Laboratory for economic research. For each game (Table 1) we ran one session with 10 independent groups of 3 subjects. Each group interacted via computer terminals for 40 rounds without knowing the identity of other subjects, using the coalition formation procedure explained below. Each subject could participate only in one supergame, maintaining the same position [strong (*A*), medium (*B*), or weak (*C*) player], ~~all of~~ which was known to the subjects. At the end each subject was paid individually according to the points obtained throughout the 40 rounds.

Experimental Design and Task. In each round of each game, each group of three bargained in two steps to elect an agent (or representative). If no member wished to be represented by another group member, all members received a payoff of zero for this round. If only one member wished to be represented, the representative could divide the corresponding coalition payoff among himself and the represented member; the third member receives a payoff of zero for this round. If two members were represented by the third member, the representative could divide the grand coalition payoff among the three members of the group.

More specifically, in step 1 each member could select at most one other agent to form a pair. If no member wished to be represented by any other (no pair was formed), the round ended with probability 10% with zero payoffs for all three; otherwise (with probability 90%) the first step was repeated. If there is more than one pair of a member and his representative, one of the pairs is chosen with equal probability. In this case, or when there is only one pair, step 1 ends and the represented group member remains passive for the rest of this round.

In step 2, each of the two active members had the option to choose the other as his representative. If no active member chose the other as his representative, the second step was repeated with probability 90%; otherwise (with probability 10%) the round ended and the representative chosen in step 1 could divide the corresponding coalition payoff among himself and the represented member. In this case, the third member received a payoff of zero for this round. If both opted for the other as their representative, the actual representative was chosen with equal probability. In this case, or when only one representative was voted for in step 2, this final agent divided 120 experimental currency units (ECU) at his discretion among all three members of the group. After this, the first step of the next round began.

The flowchart in Fig. 5 illustrates more formally our base game as implemented in the experiment. See *SI Text C* for the instructions to subjects and screenshots.

ACKNOWLEDGMENTS. We thank Felix Lamouroux for research assistance and help with the collection of data and Aniol Llorente, Riccardo Pedersini, and Jonathan Hersh for technical help. A.O. acknowledges financial support from the German Science Foundation (Deutsche Forschungsgemeinschaft) through the Leibniz Program and through the research group "Design & Behavior." R.N. thanks the Spanish Ministry of Education (SEJ2005-08391 and ECO2008-01768), ~~the Barcelona CREA program,~~ Barcelona Graduate School of Economics, and the Generalitat de Catalunya for support.

PNAS proof
Embargoed

1. Nash JF (2008) The agencies method for modelling coalitions and cooperations in games. *Int Game Theory Rev* 10(4):539–564.
2. Harrington JE (1992) The revelation of information through the electoral process: An exploratory analysis. *Econ Polit* 4:255–275.
3. Aragones E, Palfrey T, Postlewaite A (2007) Political reputations and campaign promises. *J Eur Econ Assoc* 5(4):846–884.
48. Kalisch GK, Milnor JW, Nash JF, Nering ED (1954) Some experimental n-person games. *Decision Processes*, eds Thrall RM, Coombs CH, Davis RL (Wiley, New York), pp 301–327.
Palfrey TR (2009) Laboratory experiments in political economy. *The Handbook of Experimental Economics*, eds Kagel J, Roth AE (Princeton Univ Press, Princeton), Vol 2.
5. Cabrales A, Nagel R, Rodriguez X ~~(■■■) *J Exp Econ*.~~
6. Mohr LM (1992) Authority in organizations: On the reconciliation of democracy and expertise. *EurAmerica* 22(4):13–36.
7. Roth AE (1995) Introduction to experimental economics. *Handbook of Experimental Economics*, eds Kagel JH, Roth AE (Princeton Univ Press, Princeton), pp 3–110.
8. Roth AE, Kagel J (1995) *Handbook of Experimental Economics*, eds Kagel JH, Roth AE (Princeton Univ Press, Princeton).
Camerer CF (2003) *Behavioral Game Theory* (Princeton Univ Press, Princeton).
10. Gigerenzer G, Selten R, eds (2002) *Bounded Rationality: The Adaptive Toolbox* (MIT Press, Cambridge, MA).
11. Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc Natl Acad Sci USA* 106(23):9163–9168.
12. Bolton GE, Ockenfels A (2000) ERC - A theory of equity, reciprocity and competition. *Am Econ Rev* 90:166–193.
13. Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868.
14. Schelling TC (1960) *The Strategy of Conflict* (Harvard Univ Press, Cambridge, MA).
15. Ochs J (1995) Coordination. *Handbook of Experimental Economics*, eds Kagel JH, Roth AE (Princeton Univ Press, Princeton), pp 195–251.
16. Van Huyck JB, Battalio RC, Beil RO (1990) Tacit coordination games, strategic uncertainty. *Am Econ Rev* 80:234–248.
17. Heinemann F, Nagel R, Ockenfels P (2004) Global games on test: Experimental analysis of coordination games with public and private information. *Econometrica* 72:1583–1599.
18. Camerer CF, Weigelt K (1988) Experimental test of a sequential equilibrium reputation model. *Econometrica* 56:1–36.
19. Greif A (1993) Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *Am Econ Rev* 83:525–548.
20. Bolton GE, Katok E, Ockenfels A (2004) How effective are online reputation mechanisms? An experimental investigation. *Manage Sci* 50:1587–1602.
21. Shapley LS (1953) A value for n-person games. *Contributions to the Theory of Games*, Vol II. Annals of Mathematics Studies, Vol 28, eds Kuhn HW, Tucker AW (Princeton Univ Press, Princeton).
22. Schmeidler D (1969) The nucleolus of a characteristic function game. *SIAM J Appl Math* 17(6):1163–1170.
23. Nash JF (1950) Equilibrium points in n-person games. *Proc Natl Acad Sci USA* 36(1):48–49.
24. Selten R (1981) A noncooperative model of characteristic function bargaining. *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, eds Bohm V, Nachtkamp H, in Gesellschaft, Recht, Wirtschaft, Band 4, (Wissenschaftsverla[illegible] Bibliographisches Institut, Mannheim-Wien-Zürich, Zurich) pp 131–151.
25. Chatterjee K, Dutta B, Ray D, Sengupta K (1993) A noncooperative theory of coalitional bargaining. *Rev Econ Stud* 60(2):463–477.
26. Ray D, Vohra R (1999) A theory of endogenous coalition structures. *Games Econ Behav* 26:286–336.
27. Selten R (1965) Spieltheoretische Behandl[illegible]s Oligopolmodells mit Nachfrageträgheit – Teil I: Bestimmung des dynamische[illegible]eisgleichgewicht. *Z Gesamte Staatswiss* 121:301–324.
28. Selten R (1965) Spieltheoretis[illegible]ndlung eines Oligopolmodells mit Nachfrageträgheit – Teil II: Eigenschaften de[illegible]ynamischen Preisgleichgewichts. *Z Gesamte Staatswiss* 121:667–689.
29. Harsanyi JC (1968) Games with incomplete information played by Bayesian players. *Manage Sci* 14:159–182, 320–334, 486–502.
30. Gale D, Shapley LS (1962) College admissions and the stability of marriage. *Am Math Mon* 69:9–15.
31. Roth AE (2008) Deferred acceptance algorithms: History, theory, practice, and open questions. *Int J Game Theory* 36:537–569.
32. Nash J (1953) Two-person cooperative games. *Econometrica* 21:128–140.
33. Flood MM (1958) Some experimental games. *Manage Sci* 5:5–26.
34. Selten R (1978) The chain store paradox. *Theory Decis* 9:127–159.
35. Selten R, Stöcker R (1986) End behavior in sequences of finite prisoner's dilemma supergames. *J Econ Behav Organ* 7:47–70.
36. Selten R, Mitzkewitz M, Uhlich GR (1997) Duopoly strategies programmed by experienced players. *Econometrica* 65:517–555.
37. Selten R (1987) Equity and coalition bargaining in experimental three-person games. *Laboratory Experimentation in Economics*, ed Roth AE (Cambridge Univ Press, Cambridge, MA), pp 42–98.
38. Cooper DJ, Kagel J (■■■) Other-regarding preferences. *The Handbook of Experimental Economics*, eds Kagel J, Roth AE (Princeton Univ Press, Princeton), Vol 2, pp ■■■–■■■.
39. Bettenhausen K, Murnighan JK (1985) The emergence of norms in competitive decision making groups. *Adm Sci Q* 30:350–372.
40. Kahan JP, Rapoport A (1984) *Theories of Coalition Formation* (Erlbaum, Hillsdale, NJ).
41. Selten R (1992) A demand commitment model of coalition bargaining. *Rational Interaction: Essays in Honor of John C. Harsanyi*, ed Selten R (Springer, Berlin), pp 245–28.
42. Selten R, Kuon B (1993) Demand commitment bargaining in three-person quota game experiments. *Int J Game Theory* 22:261–277.
43. Winter E (1994) The demand commitment bargaining and snowballing cooperation. *Econ Theory* 4:255–273.
44. Abbink K, et al. (2003) the fisherman's problem: Exploring the tension between co-operative and non-cooperative concepts in a simple game. *J Econ Psychol* 24:425–445.
45. Bolton GE, Brandts J, Ockenfels A (2005) Fair procedures: Evidence from games involving lotteries. *Econ J* 115:1054–1076.
46. Ofshe R, Ofshe L (1970) Choice behavior in coalition games. *Behav Sci* 15:337–349.
47. Siegel S, Siegel AE, Andrews JM (1964) *Choice, Strategy, and Utility* (McGraw-Hill, New York).

Supporting Information

Nash et al. 10.1073/pnas.1216361109

SI Text A

Some Theory. Noncooperative game theory distinguishes between the one-shot base game and finitely repeated base games, also called finite supergames. In our case of the base game, in any pure equilibrium, any coalition structure can be a final outcome, i.e., no coalition with zero payoffs for all, any two-person coalition and the three-person coalition with any of the players of a coalition being the final agent. (This is an equilibrium because somebody who deviates from this strategy by accepting another player gains nothing from this strategy.) Regarding payoffs, the final division always gives the dividing player the entire coalition payoff.

Somewhat more structure can be predicted about voting behavior: In every pure equilibrium of the base game where one player is voted for, this player will be voted for by either one or two players. The player who is voted for does not vote for another player. To see this, we exclude the possibility that in voting for an agent two players choose each other. Suppose that this happens in an equilibrium play; then it would be advantageous for each of both players to deviate by not choosing the other because then he will become an agent with higher positive probability than without this deviation. It is important that the other players are not informed about this deviation from equilibrium play, which means that their later equilibrium behavior remains the same. We can also exclude the case that there is an equilibrium play with a circular pattern of voting in the first stage. Let i, j, and k be the three players and suppose that i votes for j and j votes for k and k votes for i; then one of the players, say player i, could deviate by not voting for any player and thereby he will be an agent with a higher probability. This is clearly advantageous if the three-person coalition formation in the second stage is not reached. Suppose that it is reached. In this case one of the two remaining players does not choose the other player. It can also be seen that the agent of the two-person coalition of the first stage will be the one who does not vote for the other, because in the case that the other becomes the agent, the deviator will receive zero whereas he receives at least the two-person coalition otherwise. Consequently the circular voting pattern cannot occur in a pure equilibrium play. It can also not happen that i votes for j and j votes for k and k does not vote, because it would be advantageous for player j not to vote. Thus only one player is voted for by either one or two players.

That said, as far as pure strategy equilibria are concerned, the overall noncooperative analysis of voting, coalition, and payoff outcomes is completely independent of the players' strengths, which are the critical attributes according to cooperative game theory.

When the base game gets repeated, as in our experiment, noncooperative game theory imposes even less structure on behavior and outcomes: In the supergame almost any payoff division can be chosen in equilibrium, supported by a threat to convert to a one-shot base game equilibrium with no acceptances of any player in the case of a deviation from the equilibrium path. For instance, anything that is in equilibrium in the one-shot base game can also be supported as equilibrium play in each of the games played in the supergame. Moreover, in two-person coalitions, say, players A and B can accept each other as agents in some manner and receive both positive payoffs on average. In the final ~~periods~~, they alternate who is being accepted in the first stage and the final agent receives the entire coalition payoff. The third player never accepts and never gets accepted and so does not earn anything. Also, the grand coalition can be formed in each ~~period~~ of the supergame, and all players get on average positive payoffs. In fact, all payoff distributions, including all payoff vectors predicted by cooperative solution concepts, can be supported as equilibrium outcomes.

Cooperative game theory, on the other hand, predicts payoff divisions among the involved players for a given characteristic function (see survey, ref. 1). In our games there are always three player types with high, medium, and low strength according to the payoffs of the different two-person coalitions. Because in our experiments mostly grand coalitions are formed, we compare the actual payoff distributions with predictions to the core (2–7) and the nucleolus (8). The core of a game contains all payoff profiles that are stable in the sense that no (sub)coalition can profitably deviate and achieve a higher payoff for all of its members. The Shapley value measures each player's expected marginal contribution to a (randomly specified) coalition he could be contained in. Whereas the nucleolus is a more abstract concept, it can be shown that it is unique and always in the core if the core exists (the Shapley value and the nucleolus always exist). We describe the nucleolus and its mathematical derivation for our experimental games, along with some more general results, in *SI Text B*.

Table 1 together with Fig. 1 and Fig. S1 show the three cooperative predictions mentioned above. Reflecting the different strengths of players, both the Shapley value and the nucleolus always give most to player A and least to player C. Moreover, the nucleolus always gives more to the first player than the Shapley value, but typically less to player B than the Shapley value. Thus the third player always receives more through the Shapley value, such that the Shapley value is always closer to the equal split than the nucleolus. [See also Gomes (9), who implements either the Shapley value or the nucleolus, depending on the structure of the characteristic function game. In our experiment his solution would be the Shapley value, which indeed is chosen more often than the nucleolus.] Observe also that no core exists in games 1–5, and in game 6 it is unique and coincides with the nucleolus. We also state the quota that is calculated in *SI Text B*. The adjusted quota is equal to the nucleolus when the core does not exist and sometimes even farther away from the equal split than the nucleolus when the core exists.

We also compare our results to the equal division payoff. There are three reasons. First, in his experiments with robot players, Nash (10) found that equilibrium payoffs tend to be equal. Second, Schelling (11), among others, observed in an experiment that the equal split is a "prominent" outcome and may thus facilitate coordination and cooperation among players. And finally, the recent literature on social preferences suggests that payoff equality also has a motivational side in that some human subjects seem to dislike inequality (e.g., refs. 12, 13).

Regarding average payoffs across ~~periods~~, cooperative game theory does not distinguish between the base game and the supergame, whereas noncooperative game theory in principle allows differences. Neither approach, however, yields predictions of how the final outcome is reached.

SI Text B. Nucleolus of Our Experimental Games

1. Introduction. It is the main purpose in this section to determine the nucleolus of the 10 characteristic function games used in our experimental study. In addition to this, some general facts about the nucleolus of three-person quota games are described. The nucleolus (8) is a well-explored concept and no claim of originality is intended.

PNAS proof
Embargoed

2. Definitions and Notations. A three-person game in characteristic function form, referred to in abbreviated form as a game in the following, is a function v, which assigns a real number $v(C)$ to every nonempty subset of the player set $N = \{1, 2, 3\}$. (Here we use the notation 1, 2, 3 for the respective players A, B, and C.) The nonempty subsets of N are called coalitions. Instead of $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$, we write in abbreviated form 1, 2, 3, 12, 13, 23, and 123, respectively. All games considered here are assumed to be zero-normalized; i.e.,

$$v(1) = v(2) = v(3) = 0.$$

We use the following notation:

$$g = v(123),\ \ a = v(12),\ \ b = v(13),\ \ \text{and}\ \ c = v(23).$$

We consider only games in the class K of all games with

$$g \geq a \geq b \geq c \geq 0 \quad \text{and} \quad b + c \geq a \quad \text{as well as} \quad g > 0.$$

The first of the three inequalities involves a property called superadditivity implying that a, b, and c are nonnegative and not greater than g. In addition to this $a \geq b \geq c$ is assumed without loss of generality. The players can always be renumbered in such a way that $a \geq b \geq c$ becomes valid. The inequality $g > 0$ excludes the case that the game is inessential in the sense that $v(C) = 0$ holds for every coalition C. The equation system

$$\begin{aligned} q_1 + q_2 &= a \\ q_1 + q_3 &= b \\ q_2 + q_3 &= c \end{aligned}$$

has the unique solution

$$\begin{aligned} q_1 &= \frac{a + b - c}{2} \\ q_2 &= \frac{a + c - b}{2}. \\ q_3 &= \frac{b + c - a}{2} \end{aligned}$$

These numbers q_1, q_2, and q_3 are called the quotas of players 1, 2, and 3, respectively. If and only if the inequality $b + c \geq a$ holds, all quotas are nonnegative. Such games are called quota games. Define

$$Q = q_1 + q_2 + q_3 = \frac{a + b + c}{2}.$$

We call Q the quota sum. An imputation is a vector $x = (x_1, x_2, x_3)$ with real components and the following properties:

$$x_i \geq 0 \quad \text{for} \quad i = 1, 2, 3 \quad \text{and} \quad x_1 + x_2 + x_3 = g.$$

The core is the set of all implications with

$$\begin{aligned} x_1 + x_2 &\geq a \\ x_1 + x_3 &\geq b\,. \\ x_2 + x_3 &\geq c \end{aligned}$$

It can be seen without difficulty that the core is nonempty if and only if we have $Q \leq g$.

For every imputation x and every coalition C the sum of all x_i with $i \in C$ is denoted by $x(C)$:

$$x(C) = \sum_{i \in C} x_i.$$

The *excess* $e(C,x)$ of a coalition C in an imputation x is defined by

$$e(C,x) = v(C) - x(C).$$

If a game has a thick core, i.e., a core that has more than one element, then this core is a two-dimensional subset of the plane $x_1 + x_2 + x_3 = g$. The *excess* $e(C,x)$ is negative if x is an imputation in the interior of a thick core. Because the determination of the nucleolus is somewhat more difficult in games with a thick core than in other games, we prefer to work with positive numbers in these cases and therefore look at the *surplus* $s(C,x)$ defined by

$$s(C,x) = -e(C,x)$$

instead of the excess.

3. Nucleolus. The set of all coalitions C different from $N = \{1, 2, 3\}$ is denoted by T. As before we look at a fixed but arbitrary three-person characteristic function game v with $g \geq a \geq b \geq c \geq 0$ and $b + c \geq a$ as well as $g > 0$. All definitions are relative to v. In our case T has six elements, namely 1, 2, 3, 12, 13, and 23.

For every imputation x we construct a surplus vector

$$s(x) = (s_1(x), \ldots, s_6(x)).$$

The components $s_k(x)$ of $s(x)$ are the six numbers $s(C,x)$ with $C \in T$, ordered in such a way that we have

$$s_1(x) \leq s_2(x) \leq \ldots \leq s_6(x)$$

Let X_1 be the set of all imputations. The nucleolus is the imputation resulting from the following process of lexicographic maximization of $s(x)$: One first determines

$$s_1 = \max_{x \in X_1} s_1(x)$$

and the set X_2 of all imputations $x \in X_1$ with $s_1(x) = s_1'$. One then maximizes $s_2(x)$ within X_2 and obtains the maximum s_2 and the set X_3 of all imputations $x \in X_2$ with $s_2(x) = s_2$. In this way one continues until one obtains s_6 and X_6. Formally for $k = 2, \ldots, 6$ we have

$$s_k = \max_{x \in X_k} s_k(x)$$

and for $k = 1, \ldots, 5$ the set X_{k+1} is the set of all imputations $x \in X_k$ with $s_k(x) = s_k$. It can be shown that X_6 contains exactly one imputation. A proof of this is not given here. The nucleolus of the three-person characteristic function game under consideration is the unique imputation in X_6.

The original definition of the nucleolus by Schmeidler (8) is in terms of excess vector $e(x) = -s(x)$. The nucleolus is the result of the lexicographical minimization of $e(x)$. Our definition is an equivalent mirror image of the original one.

4. Maximizing the Minimal Surplus. We refer to $s_1(x)$ as the minimal surplus of x and to s_1 as the maximin surplus. In the following (i, j, k) is always a permutation of ~~the triple Eqs. **S1–S3**~~. The set of all six permutations of this is denoted by P. As we shall see, we have to maximize $s(x)$ under the constraints Eqs. **S1–S4** to find s_1:

PNAS proof
Embargoed

$$x_i + x_j \geq v(ij) + s \quad \text{for} \quad (i,j,k) \in P \qquad \textbf{[S1]}$$

$$x_i \geq s \quad \text{for} \quad i = 1,2,3 \qquad \textbf{[S2]}$$

$$x_i \geq 0 \quad \text{for} \quad i = 1,2,3 \qquad \textbf{[S3]}$$

$$x_1 + x_2 + x_3 = g. \qquad \textbf{[S4]}$$

The relationship Eqs. **S3** and **S4** expres~~ses~~ the condition that $x = (x_1, x_2, x_3)$ is an imputation. The constraints Eqs. **S1** and **S2** make sure that we have $s \leq s(x) \leq s(C,x)$ for every $x \in X_1$ and every $C \in T$. It follows that s_1 is the maximal s satisfying Eqs. **S1**–**S4**. In view of Eq. **S4** the sum $x_i + x_j$ in Eq. **S1** can be replaced by $g - x_k$. This yields

$$g - x_k \geq v(ij) + s \quad \text{for} \quad (i,j,k) \in P,$$

or equivalently

$$x_k \leq g - v(ij) + s \quad \text{for} \quad (i,j,k) \in P.$$

With the help of the last inequality the constraints Eqs. **S1**–**S4** can Q:5 be written as follows:

$$x_1 + x_2 + x_3 = g \qquad \textbf{[S4]}$$

$$\max[0,s] \leq x_i \leq g - v(jk) - s \quad \text{for} \quad (i,j,k) \in P. \qquad \textbf{[S5]}$$

Because these conditions are equivalent to Eqs. **S1**–**S4**, the maximin surplus s_1 is the maximal s satisfying Eqs. **S4** and **S5**.

5. Adjusted Quota Vector. Consider the vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ with

$$\tilde{x}_i = q_i + \frac{1}{3}(g - Q) \quad \text{for} \quad i = 1,2,3. \qquad \textbf{[S6]}$$

Obviously we have

$$\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = g.$$

We call $\tilde{x}$ the adjusted quota vector. The adjusted quota vector results from the quota vector (q_1, q_2, q_3) by the addition of the same amount to each quota. This amount is chosen in such a way that the sum of the $\tilde{x}_i$ is g.

We now examine the question under which conditions the adjusted quota vector is an imputation. In the view of $a \geq b \geq c$ we have $q_1 \geq q_2 \geq q_3$. Therefore $\tilde{x}_1 \geq \tilde{x}_2 \geq \tilde{x}_3$ holds. It follows that the adjusted quota vector is an imputation if and only if we have

$$\tilde{x}_3 \geq 0.$$

Equivalent transformations of this inequality yield

$$q_3 + \frac{1}{3}(g - Q) \geq 0$$

$$\frac{b+c-a}{2} - \frac{a+b+c}{6} + \frac{1}{3}g \geq 0$$

$$\frac{b+c}{3} - \frac{2}{3}a + \frac{1}{3}g \geq 0$$

$$\frac{b+c-a}{3} + \frac{g-a}{3} \geq 0.$$

In view of $b + c \geq a$ and $g \geq a$ it follows that for the games considered here the adjusted quota vector is always an imputation.

Result 1. *For the games in the class K of games considered here, the adjusted quota vector is always an imputation.*

6. Minimal Surplus of the Adjusted Quota Vector. In view of $q_i + q_j = v(i,j)$ we have

$$\tilde{x}_i + \tilde{x}_j - v(ij) = \tilde{x}_i + \tilde{x}_j - q_i - q_j = \frac{2}{3}(g - Q)$$

for every $(i,j,k) \in P$. This yields

$$s_1(\tilde{x}) \leq \frac{2}{3}(g - Q).$$

We now examine the question under which conditions the following is true:

$$s_1(\tilde{x}) = \frac{2}{3}(g - Q).$$

In view of $\tilde{x}_1 \geq \tilde{x}_2 \geq \tilde{x}_3$ this is the case if and only if we have

$$\tilde{x}_3 \geq \frac{2}{3}(g - Q).$$

Equivalent transformations yield

$$q_3 + \frac{1}{3}(g - Q) \geq \frac{2}{3}(g - Q)$$

$$q_3 \geq \frac{1}{3}(g - Q)$$

$$\frac{b+c-a}{2} + \frac{b+c+a}{6} - \frac{g}{3} \geq 0$$

$$\frac{2(b+c)}{-3} - \frac{a+g}{-3} \geq 0$$

$$b + c \geq \frac{g+a}{2}.$$

We have obtained the following result.

Result 2. *For the games in the class K the minimal surplus of the adjusted quota vector $\tilde{x}$ is*

$$s_1(\tilde{x}) = \frac{2}{3}(g - Q)$$

if and only if

$$b + c \geq \frac{g+a}{2} \qquad \textbf{[S7]}$$

holds.

Remark: We have seen that Eq. **S7** is equivalent to the condition

$$\tilde{x}_3 \geq \frac{2}{3}(g - Q).$$

If this condition does not hold, then we have

$$s_1(\tilde{x}) = s_1(3, \tilde{x}) = \tilde{x}_3 < \frac{2}{3}(g - Q)$$

PNAS proof
Embargoed

in view of $\tilde{x}_1 \geq \tilde{x}_2 \geq \tilde{x}_3$ and $s_1(i,\tilde{x}) = \tilde{x}$, as well as the fact that in $\tilde{x}$ every two-person coalition has the surplus $2(g-Q)/3$. It follows that we have

$$s_1(\tilde{x}) = \tilde{x}_3$$

if and only if Eq. **S7** does not hold.

7. Conditions Under Which the Adjusted Quota Vectors Are the Nucleolus. Summation of the three inequalities (Eq. **S5**) for $i = 1, 2, 3$ yields

$$3\max[0, s] \leq g \leq 3g - 2Q - 3s.$$

It follows that we must have

$$3s \leq 2(g-Q)$$

or equivalently

$$s \leq \frac{2}{3}(g-Q)$$

and therefore

$$s_1 \leq \frac{2}{3}(g-Q).$$

Consequently $s_1(\tilde{x})$ is the maximum surplus s_1 if

$$s_1(\tilde{x}) = \frac{2}{3}(g-Q)$$

holds. In view of *Result 2* we can conclude that the following is true.

Result 3. *For the games in the class K we have*

$$s_1 = \frac{2}{3}(g-Q)$$

if Eq. **S7** *holds.*

We now examine the set X_2 of all imputations with $s_1(x) = s_1$ under the assumption that Eq. **S7** holds. It is argued that under this assumption $s_1(\tilde{x})$ is the only element of X_2. Let $x = (x_1, x_2, x_3)$ be an imputation different from $\tilde{x}$. Then for one i among the numbers 1, 2, and 3 we must have $x_i > \tilde{x}_i$ and consequently

$$x_j + x_k < \tilde{x}_j + \tilde{x}_k \text{ for a permutation } (i,j,k) \in P.$$

It follows that

$$s_1(x) \leq x_j + x_k - v(jk) < \frac{2}{3}(g-Q) = s_1$$

holds. Therefore x does not belong to X_2. We can conclude that $\tilde{x}$ is the only element of X_2. Because for $k = 1, \ldots, 5$ the set X_{K+1} is a subset of X_K, it follows that $\tilde{x}$ is also the only element of X_6. Consequently $\tilde{x}$ is the nucleolus. We have obtained the following result.

Result 4. *For the games in the class K the adjusted quota vector $\tilde{x}$ is the nucleolus if the inequality Eq.* **S7** *is satisfied.*

Remark: The experimental games 1 ~~and 2~~ satisfy inequality Eq. **S8**. By *Result 4* the nucleolus of each of these games is the adjusted quota vector of the concerning game. Inequality Eq. **S8** is not satisfied for games 9 and 10. In the remainder of this paper the nucleolus of these games is determined.

8. Maximum Surplus If Eq. S7 Does Not Hold. From now on we assume that Eq. **S7** does not hold. Then we have

$$b + c < \frac{g+a}{2}. \qquad \textbf{[S8]}$$

Under this assumption the *Remark* after *Result 2* in *SI Text B*, section 6 comes to the conclusion that $s_1(\tilde{x}) = \tilde{x}_3$ holds. Of course, the minimal surplus $s_1(x)$ may not be maximal at $x = \tilde{x}$, but in any case we must have

$$s_1 \geq s_1(\tilde{x}) = \tilde{x}_3 \geq 0.$$

This shows that the maximization of s under the constraints Eqs. **S4** and **S5** can be restricted to nonnegative values of s. If this is done, the inequality Eq. **S5** assumes the following form:

$$s \leq x_i \leq g - v(j,k) - s \quad \text{for} \quad (i,j,k) \in P.$$

The upper bound in this inequality is smallest for $i = 3$ in the permutation (i, j, k). In this case we have $v(j, k) = a$. Therefore the following condition must hold for every nonnegative s and every imputation $x = (x_1, x_2, x_3)$ satisfying the constraints:

$$0 \leq s \leq x_3 \leq g - a - s.$$

This yields

$$s \leq \frac{g-a}{2} \leq x_3.$$

Let $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$ be the nucleolus of the game under consideration. What has been shown up to now permits the following conclusions.

Result 5. *For the games in class K the inequality*

$$0 \leq s_1 \leq \frac{g-a}{2} \leq \hat{x}_3 \leq g - a - s_1$$

holds for the maximum surplus s_1, where $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$ is the nucleolus of the game.

We now examine under which conditions we have

$$s_1 = \frac{g-a}{2}.$$

In view of *Result 5* this equation is correct, if and only if we can find an imputation $x = (x_1, x_2, x_3)$ with

$$s_1(x) = \frac{g-a}{2}$$

such that Eqs. **S4** and **S5** are satisfied for $s = s_1(x)$ and x. This is the case if and only if we have $x_1 + x_2 + x_3 = g$ and

$$\frac{g-a}{2} \leq x_1 \leq g - c - \frac{g-a}{2} = \frac{g+a}{2} - c$$

$$\frac{g-a}{2} \leq x_2 \leq g - b - \frac{g-a}{2} = \frac{g+a}{2} - b$$

$$\frac{g-a}{2} \leq x_3 \leq g - a - \frac{g-a}{2} = \frac{g-a}{2}.$$

The inequality for x_3 is satisfied if and only if

$$x_3 = \frac{g-a}{2}$$

PNAS proof
Embargoed

holds. It follows by Eq. **S4** that

$$x_1+x_2=g-\frac{g-a}{2}$$

or equivalently

$$x_1+x_2=\frac{g+a}{2}$$

must hold for the imputation $x=(x_1,x_2,x_3)$. Further necessary conditions can be derived by the inequalities for x_1 and x_2:

$$g-a\le\frac{g+a}{2}\le g+a-b-c.$$

In view of Eq. **S8** the second part of this inequality is satisfied. The first part requires

$$\frac{g}{2}\le\frac{3}{2}a$$

or equivalently

$$a\ge\frac{g}{3}. \qquad \textbf{[S9]}$$

In the following it is assumed that this necessary condition for $s_1=(g-a)/2$ is satisfied. This is the case for games 9 and 10.

It is now shown that under the conditions Eqs. **S8** and **S9** the imputation $x=(x_1,x_2,x_3)$ satisfies Eq. **S5**:

$$x_1=a$$

$$x_2=\frac{g-a}{2}$$

$$x_3=\frac{g-a}{2}.$$

It is the consequence of $a\ge g/3$ that

$$\frac{g-a}{2}\le a$$

holds:

$$a\le\frac{g+a}{2}-c$$

follows by Eq. **S8**. Therefore the inequality for x_1 in Eq. **S5** is satisfied. The inequality for x_2 in Eq. **S5**,

$$\frac{g-a}{2}\le g-b-\frac{g-a}{2}$$

or equivalently

$$g-a\le g-b,$$

is implied by $a\ge b$. As we have seen before, the inequality for x_3 in Eq. **S5** holds, too. We can conclude that $x=(x_1,x_2,x_3)$ is an imputation with $s_1(x)=(g-a)/2$. In view of *Result 5* this shows that we have obtained the following result.

Q:6 **Result 6.** *For a game in class K with the additional properties*

$$b+c<\frac{g+a}{2} \qquad \textbf{[S8]}$$

and

$$a\ge\frac{g}{3} \qquad \textbf{[S9]}$$

the equation

$$s_1=\frac{g-a}{2}$$

holds for the maximum surplus s_1. *Moreover*

$$x_3=\frac{g-a}{2}$$

and

$$x_1+x_2=\frac{g+a}{2}$$

hold for an imputation $x=(x_1,x_2,x_3)$ *with* $s_1(x)=s_1$.

Remark: Generally the set X_2 of all imputations x with $s_1(x)=s_1$ has more than one element. If this is the case, a further step in the process of lexicographic surplus maximization is necessary.

8. Further Steps in the Lexicographic Maximization of the Minimal Surplus. Q:7 We continue to assume Eqs. **S8** and **S9**. We also assume

$$a+2c\ge g. \qquad \textbf{[S10]}$$

This inequality is also satisfied by games 9 and 10. In view of *Result 6* we have

$$s(12,x)=x_1+x_2-b=\frac{g+a}{2}-a=\frac{g-a}{2}=s_1=s(1,x)$$

for every imputation x with $s(x)=s_1$ or in other words for every $x\in X_2$. Therefore the second component s_2 of the surplus vector $s(x)=(s_1,\ldots,s_6)$ is equal to s_1 and we have $X_3=X_2$.

We now have to determine s_3. We have $X_3=X_2$ and

$$s_3=\max_{x\in X_1}:\min_{C\in T}[s(1,x),s(2,x),s(13,x),s(23,x)].$$

It will be shown that this maximal minimum is achieved at one and only on imputation $\hat{x}=(\hat{x}_1,\hat{x}_2,\hat{x}_3)$. This imputation is the nucleolus of the game under consideration. We first determine

$$s_3'=\max_{x\in X_1}:\min_{C\in T}[s(13,x),s(23,x)]$$

and the uniquely determined imputation $x'=(x_1',x_2',x_3')$ at which this maximal minimum is achieved. Obviously we must have

$$s_3'\ge s_3.$$

By *Result 6* we have

$$x_1=\frac{g+a}{2}-x_2.$$

With the help of this relationship $s(13,x)$ and $s(23,x)$ can be expressed as functions of x_2:

$$s(13,x)=\frac{g+a}{2}-x_2+\frac{g-a}{2}-b=g-b-x_2$$

$$s(23,x)=x_2+\frac{g-a}{2}-c.$$

PNAS proof
Embargoed

Because $s(13,x)$ is decreasing and $s(23,x)$ is increasing in x_2, the minimum of both surpluses is achieved at the value of x_2' at which both are equal provided that Eq. **S5** is satisfied at $x' = (x_1', x_2', x_3')$ with $x_1' = (g+a)/2 - x_2'$ and $x_3' = (g-a)/2$. The equation

$$x_2' + \frac{g-a}{2} - c = g - b - x_2'$$

yields

$$x_2' = \frac{g+a}{4} - \frac{b-c}{2}$$

and therefore

$$x_1' = \frac{g+a}{4} + \frac{b-c}{2}$$

and with these values for x_1 and x_2 one obtains

$$s_3' = s(13,x') = s(23,x') = \frac{g+a}{4} - \frac{b-c}{2} + \frac{g-a}{2} - c$$

$$s_3' = \frac{3g - a - 2(b+c)}{4}.$$

With the help of Eq. **S8** and $g \geq a$ it can be seen that s_3' is positive. The inequalities Eq. **S5** for x_1' and x_2' are satisfied for $s = s_3'$ if we have

$$\begin{gathered} s_3' \leq x_1' \leq g - c - s_3' \\ s_3' \leq x_2' \leq g - b - s_3' \end{gathered}.$$

The difference between the upper bounds for x_1' and x_2' is $b - c$. We also have $x_1' - x_2' = b - c$. In view of Eq. **S8** the sum of both upper bounds is greater than $(g+a)/2$, the sum of x_1' and x_2'. It follows that x_1' and x_2' are below their respective upper bounds by the same amount. Therefore the inequality for x_1' is satisfied if the inequality for x_2' is satisfied. We have to examine whether

$$\frac{3g - a - 2(b+c)}{4} \leq \frac{g + a - 2(b-c)}{4} \leq g - b - \frac{3g - a - 2(b+c)}{4}.$$

The first part of this inequality requires

$$3g - a - 2c \leq g + a + 2c$$

$$g \leq a + 2c.$$

This is nothing else than assumption Eq. **S10**. The second part of the inequality for x_2' requires

$$g + a - 2(b+c) \leq g + a - 2b + 2c$$

or equivalently

$$0 \leq 4c.$$

We can conclude that $x' = (x_1', x_2', x_3')$ satisfies Eq. **S5** with $s = s_3$. We now want to argue that x' is the nucleolus. It remains to show that $s_3' = s_3$ holds. For all imputations $x = (x_1, x_2, x_3)$ in X_2 we have $x_3 = (g-a)/2$ and $x_1 + x_2 = (g+a)/2$. Therefore x_3 is fixed in X_3 and x_1 is a function of x_2. Suppose that the maximum of $s_3(x)$ in X_3 is attained at an x_2 with $x_2 \neq x_2'$. Then at x_2 either $s(13,x)$ or $s(23,x)$ is smaller than s_3' because in X_3 the surplus $s(13, x)$ is decreasing in x_2, whereas $s(23, x)$ is increasing in x_2. This shows that s_3 is maximized in X_3 at $x = x'$. Moreover, this maximum is not attained at any other imputation than x'. Therefore x' is the only imputation in X_4, X_5, and X_6. It follows that x' is the nucleolus $\hat{x}$.

Result 7. *For the games in class K the imputation* $x = (x_1, x_2, x_3)$ *with*

$$\hat{x}_1 = \frac{g+a}{4} + \frac{b-c}{2}$$

$$\hat{x}_2 = \frac{g+a}{4} - \frac{b-c}{2}$$

$$\hat{x}_3 = \frac{g-a}{2}$$

is the nucleolus, if Eqs. **S8–S10** *are satisfied.*

Tabel~~le~~ S1. The experimental games:
Explanations.

$$a = v(12),\ \ b = v(13),\ \ c = v(23),$$

$$g = v(123) = 120,\ \ v(i) = 0 \quad \text{for} \quad I = 1,2,3$$

$$q_1 = \frac{a+b-c}{2},\ \ q_2 = \frac{a+c-b}{2},\ \ q_3 = \frac{b+c-a}{2}.$$

The nucleolus of games 1–8 is the adjusted quota vector (*SI Text B*, section 5). The nucleolus of games 9 and 10 is the vector $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$ described in *Result 7*.

SI Text C. Instructions

Welcome and thank you for participating in this experiment. Please read these instructions carefully. All participants received identical instructions.

Please raise your hand, if you have any questions. All questions will be answered at your desk. From now on all communication with other participants is strictly prohibited. Please switch off your mobile phones now.

Any amount of money will be expressed in experimental currency units (ECU) during the experiment and will be paid out to you in cash at the end of the experiment, using an exchange rate of 1 euro (EUR) = X ECU. You receive 2.50 EUR at the beginning of the experiment as a show-up fee. In the course of the experiment you will be able to earn more money. How much you will earn depends on your actions and the actions of the other participants. All your decisions will be treated confidentially and anonymously.

Bargaining Situation. In the course of the experiment you will be interacting within a group of three members in total. You will be informed about your member name (*A*, *B*, or *C*) before being asked to make your first decision. The members of the group and their names will remain unchanged during the entire experiment. The experiment will be carried out over ~~100~~ rounds. In each round the group will be presented with an identical bargaining situation. The bargaining proceeds in predefined steps, in each of which members can elect representatives from within the group.

If no member wishes to be represented by any other, all members will receive a payoff of zero for this round.

If only one member wishes to be represented, the representative can divide a predefined sum (see graphic and enumeration below) among himself and the represented member; the third member receives a payoff of zero for this round.

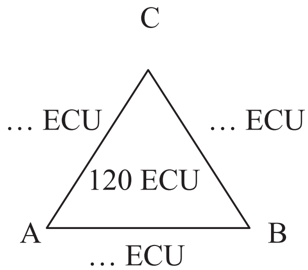
PNAS proof Embargoed

If A represents B or vice versa, the sum to be divided among these two members amounts to XX ECU (C receives no payoff).

If A represents C or vice versa, the sum to be divided among these two members amounts to XX ECU (B receives no payoff).

If B represents C or vice versa, the sum to be divided among these two members amounts to XX ECU (A receives no payoff).

If two members are represented by the third member, the representative can divide 120 ECU among the three members of the group.



Steps in the Bargaining Process. Each round consists of a maximum of three steps.

Step 1. In the first step (Screen 1) each member of the group has the option to choose at most one of the other two members as his representative; e.g., A has the option to choose B or C or neither of these two as his representative.

If more than one member has chosen another member as his representative, each of the pairs consisting of one member and one representative has an equal chance of being picked; e.g., if A has chosen B and B has chosen C as his representative, each of these two constellations will be chosen with a probability of 50%. The represented member remains passive for the rest of this round; he or she will not make any further decisions in this round. The representative and the third member remain active.

If no member chooses a representative, the round ends with a probability of 1% with a payoff [illegible] ECU for all members; otherwise (with a probability of 99%) the first step is repeated.

If step 1 ends with one representative being chosen, step 2 (Screen 2) begins.

Step 2. The two active members have now the option to choose the other or nobody as their representative; e.g., if member B has been chosen in step 1 to represent A, B and C can choose each other or nobody as their representative. If both active members choose the other as their representative, the representative will be chosen at random with both members having an equal chance of being picked as representative. If only one active member has chosen the other as his or her representative, the other will immediately be picked as representative.

If no active member chooses the other as his or her representative, the second step is repeated with a probability of 99%; otherwise (with a probability of 1%) the round ends here and the representative chosen in step 1 (in the above example, B) can divide the defined sum for this pair (in the above example, XX ECU) among himself or herself and the represented member (Screen 3). The third member (in the above example, C) receives a payoff of zero for this round.

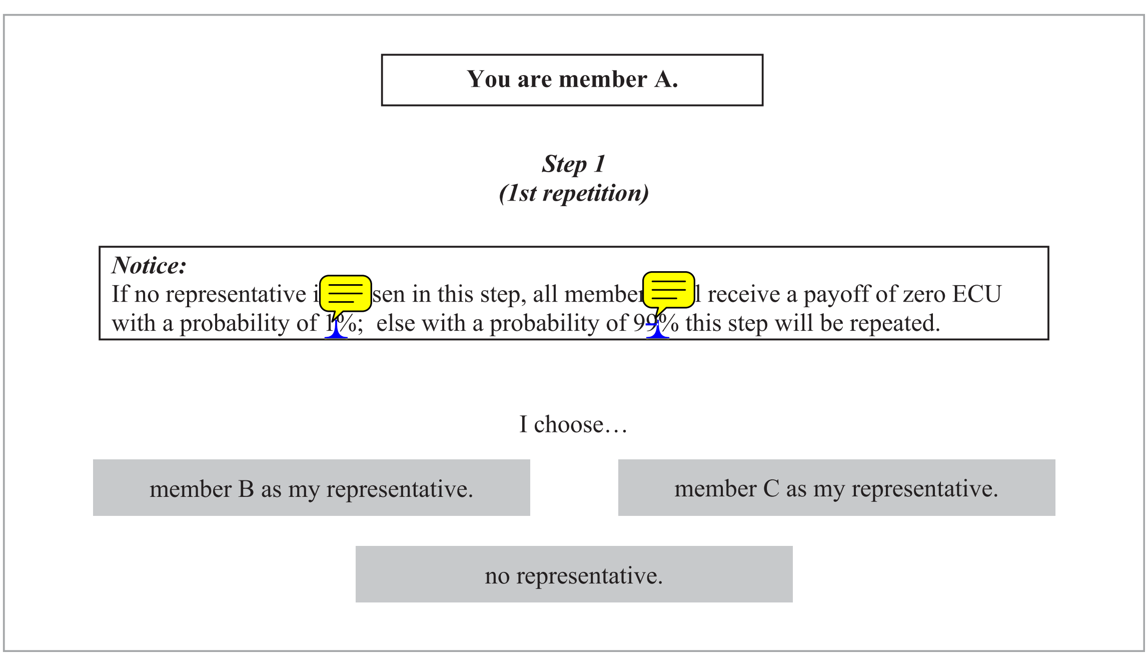
If step 2 ends with one representative being chosen, step 3 (Screen 4) begins.

Step 3. The representative chosen in step 3 can divide 120 ECU at his discretion among all three members of the group.

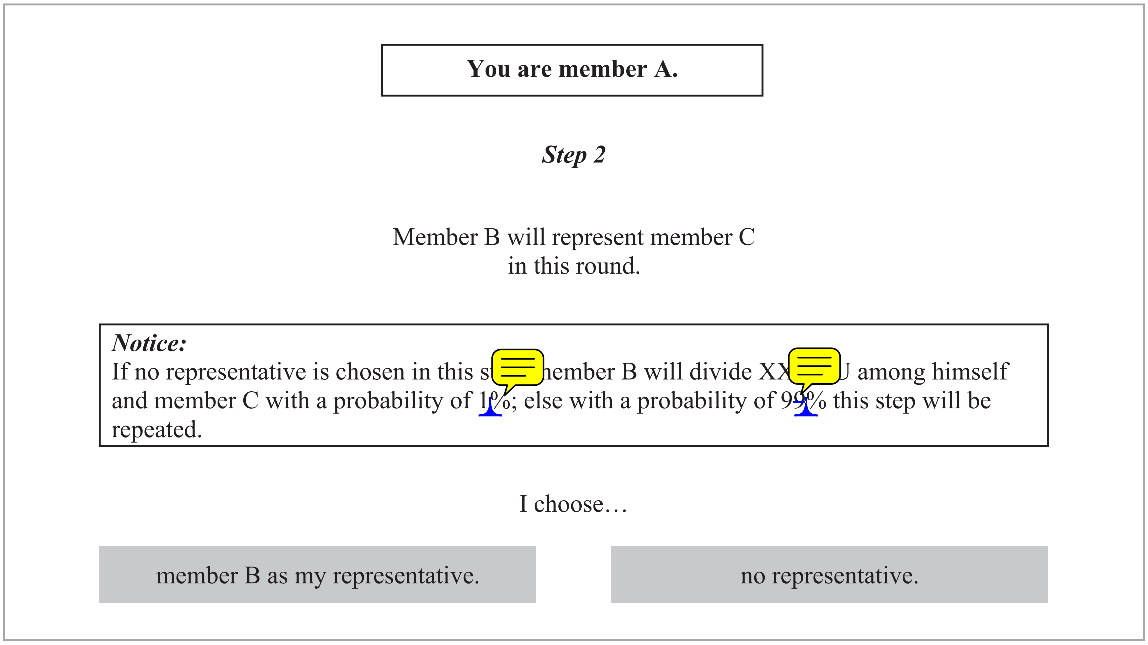
After this, the first step of the next round begins. After [illegible] rounds, earnings of all rounds will be paid out in cash.

PNAS proof
Embargoed

Information and Decision Screens During a Round.

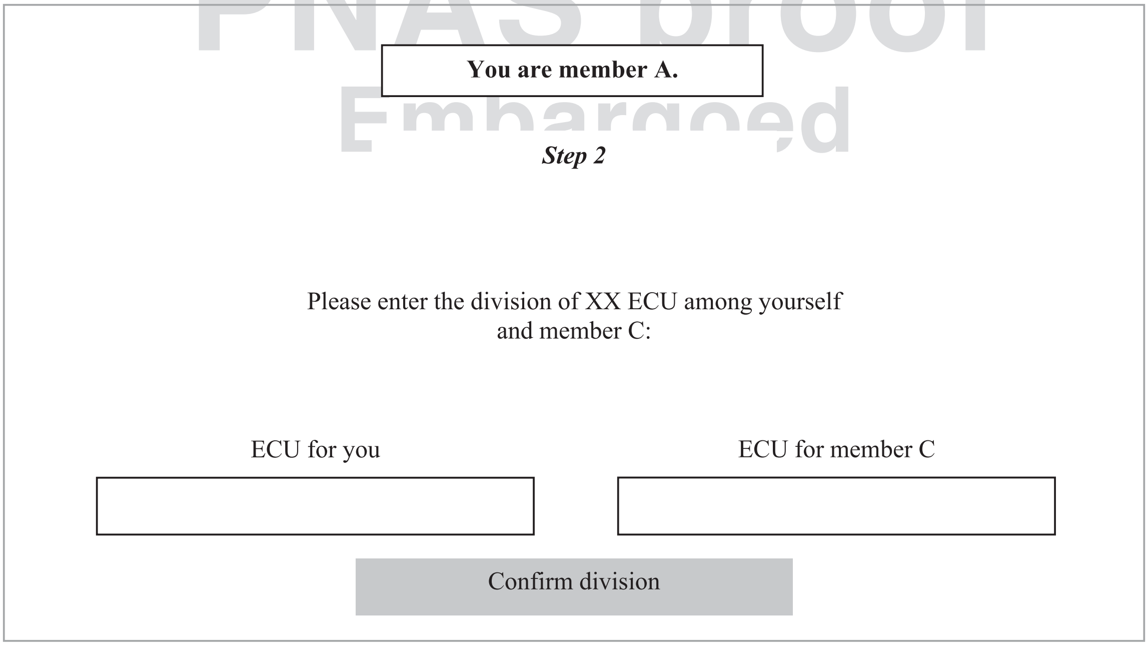


Screen 1. On the first screen you decide whether you want one of the other members or no other member to be your representative. You will also be informed about how many times this step has been repeated already in this round.



Screen 2. You will be informed about which member is represented by whom. Active members can choose the other or no active member as their representative.

PNAS proof
Embargoed



Screen 3. If the round has been terminated after step 2, the representative chosen in step 1 will divide the defined amount for this pair among himself and the passive member.

You are member C.

Step 3

Please enter the division of 120 ECU among yourself and the other members:

ECU for you	ECU for member A	ECU for member B

Confirm division

Screen 4. You will be informed about which active member has been chosen as representative. The representative then chooses the division of 120 ECU among all three members.

PNAS proof
Embargoed

You are member A.

30th Round result

Member C has divided 120 ECU among A,B,C as follows:

ECU for member A	ECU for member B	ECU for member C
10	**10**	**100**

Round	Representative	Represented	Payoff A	Payoff B	Payoff C
30	C	A,B	10	10	100
29	B	C	-	50	10
28	Nobody	Nobody	-	-	-
27	C	A,B	0	20	100
26	B	A	20	40	-

OK

Round end. All members are informed about the round payoffs of all members. In addition, the results of all previous rounds are displayed. Please confirm this screen by clicking OK to begin the next round.

Payoff. The sum of all round payoffs will be displayed at the end of the experiment in ECU as well as in EUR, including the show-up fee of 2.50 EUR.

If you have any questions now or in the course of the experiment, feel free to raise your hand. We will answer all questions at your desk. Be reminded that any communication with other participants is strictly prohibited.

We thank you for your participation in this experiment!

1. Maschler M (1992) The bargaining set, kernel, and nucleolus. *Handbook of Game Theory with Economic Applications*, eds Aumann RJ, Hart S (Elsevier/North-Holland, New York), Vol 1, pp 592–667.
2. Gillies DB (1959) Solutions to general non-zero-sum game. *Contributions to the Theory of Games*, eds Tucker AW, Luce RD (Princeton Univ Press, Princeton), Vol 4, pp 47–85.
3. Shapely LS (1959) The solutions of a symmetric market game. *Contributions to the Theory of Games*, Annals of Mathematics Studies 40, eds Tucker AW, Luce RD (Princeton Univ Press, Princeton), Vol 4, pp 145–162.
4. Shubik M (1959) Edgeworth market games. *Contributions to the Theory of Games*, Annals of Mathematics Studies 40, eds Tucker AW, Luce RD (Princeton Univ Press, Princeton), Vol 4, pp 267–278.
5. Aumann RJ (1964) Markets with a continuum of traders. *Econometrica* 32:39–50.
6. Scarf HE (1967) The core of an n-person game. *Econometrica* 35:50–69.
7. Shapely LS (1953) A value for n-person games. *Contributions to the Theory of Games*, Annals of Mathematics Studies 28, eds Kuhn HW, Tucker AW (Princeton Univ Press, Princeton), Vol 2, pp 307–317.
8. Schmeidler D (1969) The nucleolus of a characteristic function game. *SIAM J Appl Math* 17:1163–1170.
9. Gomes A (2005) Multilateral contracting with externalities. *Econometrica* 73:1329–1350.
10. Nash JF (2008) The agencies method for modelling coalitions and cooperations in games. *Int Game Theory Rev* 10(4):539–564.
11. Schelling TC (1957) Bargaining, communication, and limited war. *J Conflict Resolut* 1: 19–36.
12. Fehr E, Schmidt KM (1999) A theory of fairness, competition, *and cooperation*. *Q J Econ* 114:817–868.
13. Bolton GE, Ockenfels A (2000) ERC - A theory of equity, reciprocity and competition. *Am Econ Rev* 90:166–193.

Table S1. Characteristic functions, unadjusted quota, and nucleolus for the 10 games used in the experiment

	Coalition values			Quotas			Nucleolus		
Game no.	a	b	c	q_1	q_2	q_3	x_1	x_2	x_3
1	120	100	90	65	55	35	$53\frac{1}{3}$	$43\frac{1}{3}$	$23\frac{1}{3}$
2	120	100	70	75	45	25	$66\frac{2}{3}$	$36\frac{2}{3}$	$16\frac{2}{3}$
3	120	100	50	85	35	15	80	30	10
4	120	100	30	95	25	5	$93\frac{1}{3}$	$23\frac{1}{3}$	$3\frac{1}{3}$
5	100	90	70	60	40	30	$56\frac{2}{3}$	$36\frac{2}{3}$	$26\frac{2}{3}$
6	100	90	50	70	30	20	70	30	20
7	100	90	30	80	20	10	$83\frac{1}{3}$	$23\frac{1}{3}$	$13\frac{1}{3}$
8	90	70	50	55	35	15	60	40	20
9	90	70	30	65	25	5	$72\frac{1}{2}$	$32\frac{1}{2}$	15
10	90	50	30	45	25	5	$57\frac{1}{2}$	$37\frac{1}{2}$	25

In columns 2–4 we state for each of the 10 games the ~~payoffs~~ for the two-person coalitions, $V(XY) = V(AB)$, $V(AC)$, or $V(BC)$, respectively, of the characteristic function; the three-person coalition (grand coalition) is always 120, and the one-person coalition is normalized to 0. Columns 5, 6, and 7 present the theoretical payoffs for players *A*, *B*, and *C*, respectively, according to quota, and columns 8, 9, and 10 those for the nucleolus in a one-shot cooperative game. The quota for each player is calculated through the system of equations given by the two-person coalition values (*SI Text B*, section 2). The theoretical payoffs of the nucleolus are always distributions of the grand coalition value. Player *A* is the strong player in all games, because with him the highest two-person coalition payoffs can be achieved compared to a coalition without him. For a similar reason *B* is the second strongest player. In all theoretical cooperative solutions player *A* receives always the highest ~~player~~ payoff, followed by player *B*.

PNAS proof
Embargoed

Fig. S1. Average payoff distributions across all ~~periods~~ per group and game. Actual distributions (asterisks, only from the 71 groups that divided in less than 36 of 40 ~~periods~~ according to the equal split), theoretical cooperative solutions (core, yellow area; nucleolus, triangles; Shapley, squares), and equal split (circles) are shown. Note that the nucleolus and the core coincide in game 6.

Fig. S1

Fig. S2. Evolution of payoff divisions per group and game in each of the 40 ~~periods~~. Each line shows 10 different groups per game. It is a coarse visual impression of the payoff division in every ~~period~~. In about 30% of all groups the representative agents divide according to the equal split in almost all ~~periods~~. In the remaining groups there are a lot of unequal payoff divisions in most of the ~~periods~~.

Fig. S2

Fig. S3. (*A*–*J*) Payoff distribution for the three players for each of the 10 games and each of the 10 groups, separately. Here we show precisely which payoff divisions were proposed. Triangles indicate two-person coalition and the open circles show three-person coalition. The circles indicate the weight over 40 ~~periods~~ for a specific payoff vector, where the center represents the payoff vector. The payoff to player *C* in the grand coalitions is always 120 minus payoff to *A* minus payoff to *B*. The bigger the circle is around a dot, the higher the frequency that the specific payoff division was chosen.

Fig. S3