SCIENTIFIC REPORTS

OPEN

Humans best judge how much to cooperate when facing hard problems in large groups

Andrea Guazzini[1,2], Federica Stefanelli[1,2], Enrico Imbimbo [1,2], Daniele Vilone[3,4], Franco Bagnoli[2,5] & Zoran Levnajić[6]

We report the results of a game-theoretic experiment with human players who solve problems of increasing complexity by cooperating in groups of increasing size. Our experimental environment is set up to make it complicated for players to use rational calculation for making the cooperative decisions. This environment is directly translated into a computer simulation, from which we extract the collaboration strategy that leads to the maximal attainable score. Based on this, we measure the error that players make when estimating the benefits of collaboration, and find that humans massively underestimate these benefits when facing easy problems or working alone or in small groups. In contrast, when confronting hard problems or collaborating in large groups, humans accurately judge the best level of collaboration and easily achieve the maximal score. Our findings are independent on groups' composition and players' personal traits. We interpret them as varying degrees of usefulness of social heuristics, which seems to depend on the size of the involved group and the complexity of the situation.

Cooperation allows to solve problems that are too complex for anyone to solve individually. It allows to concentrate the efforts of many individuals on tasks that do not give immediate reward to none of them. It is one of the key factors that lead humans to become possibly the most cooperative species in nature, able to collaborate at (almost) every level of our society[1,2].

But the benefits that cooperative behavior brings come with a price for the individual. This apparently odd dualism and the social mechanisms behind it has been the focus of research attention for decades[3]. Specifically, research identified several factors influencing pro-social behavior, including direct[4,5] and indirect[6,7] reciprocity, multilevel cooperation[8,9], kin-selection[10], and social network structure[11,12]. These mechanisms seem to be responsible for transforming the cooperation into an evolutionary stable strategy and spreading it within and among human societies[13].

In fact, we routinely rely on various collaboration strategies to solve problems of different nature in our daily lives. This instinctive drive for cooperation has long been a topic of interest[14–16], since understanding the factors behind this drive could promote and regulate successful collaboration among human beings. These processes are often studied in form of "crowdsourcing" or "wisdom of the crowds"[17–21]. In particular, it was found that cooperative organized actions can lead to original ideas and solutions of notoriously hard problems. These include improving medical diagnostics[22], designing RNA molecules[23], computing crystal[24] or predicting protein structures[25], proving mathematical theorems[26], or even solving quantum mechanics problems[27]. In all of these studies, the actual problem was simplified into an approachable computer game.

Another interesting question in this context is what kind of decision-making process leads to cooperation and which to selfishness? Does this depends on whether the decisions are based on rational calculation or are more intuitive? The decision-making in cooperative games is known to be influenced by time pressure[28,29], yet it

[1]University of Florence, Department of Educational Science and Psychology, Florence, 50126, Italy. [2]University of Florence, Center for Study of Complex Dynamics, Florence, Italy. [3]LABSS (Laboratory of Agent Based Social Simulation), Institute of Cognitive Science and Technology, National Research Council (CNR), Rome, 00185, Italy. [4]Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Madrid, Spain. [5]Department of Physics and Astronomy, University of Florence and INFN, via G. Sansone 1, Sesto Fiorentino, 50019, Italy. [6]Faculty of Information Studies in Novo mesto, Novo Mesto, 8000, Slovenia. Andrea Guazzini, Federica Stefanelli and Enrico Imbimbo contributed equally. Correspondence and requests for materials should be addressed to E.I. (email: enrico.imbimbo@unifi.it)

is likely to be also influenced by a series of other factors. Since the concept of collaboration is naturally linked to confronting problems in groups, two obvious factors are the *size of the group* confronting the problem, and the *complexity of the problem* itself. The group size as the variable is related to the *peer pressure* exerted by the group on the individual: when in small groups, humans behave differently than when in the crowd[30]. Actually, the relation between the group size and the benefit to the individual from cooperation has been largely studied[31–34]. On the other hand, the problem complexity as perceived by an individual can be real or imaginary. Yet regardless of this, it is well-known to objectively influence our decision-making[35,36].

The role of both these variables can be measured via suitable game-theoretic experiment: after decades of successful theoretical developments[2,19,37,38] and with the help of modern technology, game theorists can now realize diverse experiments under controlled conditions. They revealed profound aspects of human cooperation in simple models such as Prisoner's dilemma[39–44], and in intricate phenomena such as reputation[45], coordination[46], reward and punishment[47], inter-generational solidarity[48], and for various underlying social networks[41,49].

The question we ask here is: How much error we make when estimating the benefits of collaboration without having the relevant information for that decision available? Does this depend on the size of the social group and the complexity (difficulty) of the problem that our group is trying to solve? Precisely: *given a group of human players solving a problem of specific level of complexity, how much does the empirically found level of collaboration of an average player differs from the level of collaboration that leads to the best attainable score*? This question requires an experiment whose realization is associated with several difficulties. First, how to design a sequence of problems whose levels of complexity are equally perceived by all players? Second, the obtained results should be universal, and not attributable to individual characteristics of any player or the composition of any group of players. Third, we need strong evidence that players do not make their collaborative decisions (exclusively) via rational calculation. Fourth, we need to find the level of collaboration which leads to the best outcome and compare it to the empirically found level of collaboration.

We designed and performed just such an experiment and complement it by the equivalent computer simulation. As we report in what follows, this framework allows to circumvent the above difficulties to a sufficient degree and obtain interpretable findings (we explain later why we believe that players primarily follow their intuition). We found that humans match the optimal strategy only in large groups or when solving hard problems. In contrast, when playing alone, humans massively fail. In agreement with the previous studies, we conclude that in addition to time pressure[29,50–53], the onset of cooperative behavior is affected by the size of the social group and the complexity of the situation latane[28,54–56].

The experiment and the simulation

The experiment with human players. We recruited 216 participants on a voluntary basis (150 f, 66 m, age $19-51$, $M = 22.8$, $SD = 4.53$), all students at School of Psychology or School of Engineering of the University of Florence, Italy, who declared themselves native Italian speakers.

Our experiment involves "solving" the problems of increasing complexity in groups of increasing size. But how to design a "problem" whose level of complexity (difficulty) is objectively seen as equal by all participants[35,57]? We first note that we are here not interested in modeling the problem complexity *per se*, but in studying the *effects* that it has on participants (players). With this in mind we represent a "problem" as a probability p $(0 < p < 1)$, which models the chance of problem being "solved". In other words, we use p as a proxy for problem complexity, so that $p = 0$ models a problem impossible to solve (and get reward for it), whereas $p = 1$ means problem is automatically solved. For easier orientation we define the level of complexity for a problem as $R = 1 - p$, so that a problem with R close to 0 is "easy", while a problem with R close to 1 is "hard". Hence, solving a problem of complexity R means throwing a random number between 0 and 1: if it is smaller (bigger) than $p = 1 - R$, the problem is solved (not solved). In our experiment players make decisions to collaborate or not when solving the problems at four levels of complexity: $R = 0.9, 0.7, 0.5$ and 0.1 (details follow). Our representation of the problem complexity is crude and unrelated to player's cognitive abilities, but it guarantees that complexity is perceived equally by all players (more details in the Supplement). This is crucial for our experiment, as it establishes the grounds for assuming that players' collaborative decisions are based on a unique perception of the problem.

The experiment was organized in 18 sessions, each involving 12 different participants (players). Each participant was engaged in only one session ($18 \times 12 = 216$). One session consisted of two matches played by these 12 players, each match identified by its value of R. That is to say that each of the 12 participants only played at two out of the four levels of complexity R. Each of the four values of R was used in 9 different matches, played within 9 different sessions, each session attended by 12 different players. By shuffling the participants this way, we can guarantee that the results cannot be attributed to their personal characteristics. In each match the 12 players were divided in groups of equal size S. The experiment involved four different schemes of division into groups: $S = 1$ (12 groups with one player each), $S = 3$ (4 groups with 3 players), $S = 6$ (2 groups with 6 players) and $S = 12$ (all 12 players in one single group). A match consisted of a sequence of four games played consecutively, one for each division scheme defined by S. This implies that each complexity level R used in 9 matches is, in each match, played at all four division schemes.

A game consisted of participants in all groups playing 11 rounds. One round for a player means one decision to cooperate or not with other $S - 1$ players in his/her group, facing a problem of complexity R. By virtue of each player making a decision in a given round within his/her group, that group is split into cooperators (players that decided to cooperate in that round) and non-cooperators (players that decided otherwise). With a probability $p = 1 - R$ a player either solves the problem (wins) or fails to solve it (loses). This happens simultaneously in all groups which are part of a division scheme. A round is completed when all players in all groups (cooperators and non-cooperators) had either won or lost. In the next round all players in all groups choose again to cooperate or not, and the process repeats. Once all 11 rounds are completed, a new game starts, in which the 12 players are divided according to the next division scheme S, and the scoring is reset to zero. Once all four division schemes

are played, the match is completed, and a new match starts at a next complexity level R. Once two matches are completed, the session is closed, and a new session starts with 12 new participants invited into the experiment room. Each combination of R and S is played 9 times, each time with different players, making the results not attributable to the compositions of the groups.

Each player has an individual score that is increased in each round via two separate factors: (i) the number N_c^w of cooperators within the group that won in a given round, and (ii) the group bonus GB, a separate score related to the entire group, which increases by 1 each time at least one cooperator wins. The scoring per round for a cooperator is:

$$\text{Pay-off}_{\text{coop}} = \left(\frac{1 + GB}{S}\right) N_c^w,$$

(1)

where S is the number of players in the group (not the number of cooperators). GB increases by 1 each time $N_c^w > 0$. Note that the payoff of a cooperator does not depend on whether he/she wins, but just on N_c^w and GB.

In contrast to a cooperator, a non-cooperator is allowed to choose his/her own gain K from integers between 1 to 10, with the payoff depending directly on winning. The price for this is decreased winning probability: a non-cooperator wins with probability p^K (decreases exponentially with K). The scoring for a non-cooperator who wins is:

$$\text{Pay-off}_{\text{non-coop}}(\text{win}) = K + \left[\left(\frac{1 + GB}{S}\right) N_c^w\right],$$

(2)

and is not related to the number of non-cooperators that win. On the other hand, even when loses, a non-cooperator is guaranteed the same payoff as a cooperator:

$$\text{Pay-off}_{\text{non-coop}}(\text{lose}) = \left(\frac{1 + GB}{S}\right) N_c^w = \text{Pay-off}_{\text{coop}}.$$

(3)

So non-cooperators earn by exploiting cooperators' efforts, while cooperators have no benefit of non-cooperators' success. GB does not increase when non-cooperators win.

All decisions and scores are stored throughout the experiment. To simplify our analysis, we focus only on the two following variables.

- *Average probability of cooperation C*: the fraction of positive cooperative decisions made by a given player in one game (11 rounds–11 decisions),
- *Agent fitness AF*: the final cumulative payoff of a given player after completing one game.

The values of C and AF are measured in all games. For each combination of R and S we thus obtain 9 values of C and AF, each for one game played by a different player. We average these 9 values to obtain the universal values of C and AF for the corresponding R and S, which are not attributable to characteristics of players/groups.

The entire experiment was based on online communications, since they are easier to record than face-to-face communications and are less sensitive to non-verbal interactions[58], which can distort the results of an experiment (more technical details in the Supplement). Participants also completed a standard socio-demographic and psychological questionnaire. Ethical Commission of the University of Florence was made aware of our experiment, yet under current ethical regulations no explicit approval was required.

The equivalent computer simulation. Due to its game-theoretic nature, our experiment can be directly translated into a computer simulation, where human players are modeled as evolving agents[20]. This was done by letting the (virtual) agents evolve in a competitive environment (similarly to a genetic algorithm). Each agent is defined by a fixed collaboration probability. Unsuccessful agents are replaced by new agents, where collaboration probability is selected at random. This allows us to extract the collaboration probability C_{best} that leads to theoretically best agent fitness AF_{best}. We computed C_{best} and AF_{best} by applying our evolutionary algorithm to a large population of evolving agents (other details in the Supplement). This is done separately for each combination of R and S, assuming that agents have no knowledge about each others decisions, in accordance with the experimental setting.

Results

We present our findings by comparing the experimental and simulated values of agent fitness (AF vs. AF_{best}) and of average probability of cooperation (C vs. C_{best}).

Humans systematically earn less than they could and reach their full potential only when confronting the hardest problems or working in the largest group. We first examine the differences between AF and AF_{best} in relation to R and S. This is also to confirm that our computer simulation did indeed extract the strategy that leads to the best attainable AF_{best}. However, AF and AF_{best} cannot be immediately compared, since due to the evolutionary nature of the simulations the two scoring schemes are not normalized in the same way. In order to make an interpretable comparison, we run additional independent simulations as described in the Supplement, and obtain new values of AF and AF_{best}, reported in the Fig. 1, top two panels.

Looking at the top left panel in Fig. 1 (four plots, a–d), within each plot we show the values for four problem complexities R. For simple problems (a) humans earn considerably less than they best can, specially for the group consisting of half of players. Exception is the case of largest group (12 players), where humans actually match the
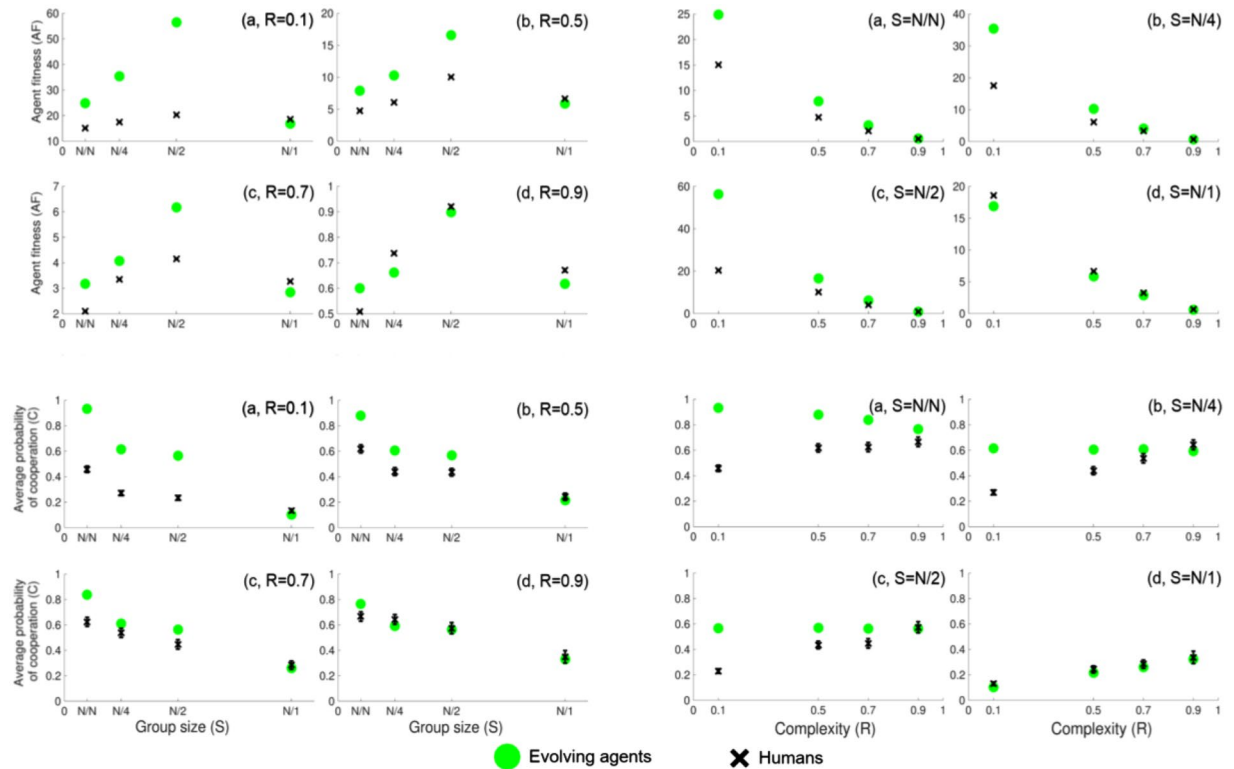
Figure 1. Comparison of the experimental and the simulated values of agent fitness and average probability of cooperation. Top two panels: Comparison of experimental AF (black crosses) and best attainable AF_{best} (green circles). Top left panel: four plots (**a**–**d**) show the comparison over four values of problem complexity R. Within each plot we show the values for four group sizes S, where N/N indicates the group composed of a single player, while $N/4$, $N/2$ and $N/1$ respectively indicate the divisions into 4, 2 and 1 group (since experiment and simulation have different total number of players/agents, for easier comparison we use the notation involving number of groups). Top right panel: four plots (**a**–**d**) show the same comparison, but this time over four group sizes S. When dealing with simple problems and when playing in small groups, humans earn much less than they could. This performance improves as problems get harder and/or as humans play in larger groups. Finally, when facing hardest problems or when playing all together in one group, humans earn as much as they could. Humans somewhere appear to earn slightly more than evolving agents: this is an artifact coming from the statistical nature of the simulations (see Supplement). Bottom two panels: Comparison of experimental C (black crosses) and C_{best} leading to $_{best}$ agent fitness (green circles). Bottom left panel: four plots (**a**–**d**) show the average level of cooperation C for four values of problem complexity R. Within each plot we show the values for four group sizes S, as in the top panels. Bottom right panel: four plots (**a**–**d**) show the same comparison, but this time over four group sizes S. When in isolation/small groups and/or when confronting simple problems, humans cooperate far less that it would be best for their benefit. As groups become bigger and problems become harder, average human level of collaboration increases. Finally, when all in one group or when facing hardest problems, humans correctly judge the best level of cooperation. All differences between experimental and simulated values in all panels are statistically significant. All values have error bars (very small) that express standard deviations.

best score. As problems get harder (b and c), the performance of humans is systematically closer to the best per-formance. The payoffs in largest group are always matching the best payoffs. When facing the hardest problems (d), humans basically earn the maximum possible payoff in groups of any size. In the top right panel in Fig. 1 we show the same values as in the top left panel, but now presenting them via group size S. For the smallest group (a) consisting of one player who plays alone, humans largely under-perform, but their performance improves as the problems become harder. As the group size increases (b and c), human performance increasingly matches the best performance–a trend that is more pronounced for harder problems. When playing in the group comprising all 12 players (d), humans exhibit almost ideal performance regardless of the problems complexity.

As R approaches 1 the values of AF and AF_{best} go to zero, so our result may appear trivial there. But as we show below, the average probability of cooperation C does not go to zero in this limit. We also found that the ratio between AF_{best} and AF is close to 1 in this limit independently of S, implying that this is not a mere artifact.

These results indicate a clear correlation between AF and both R and S. The performance of humans systemat-ically improves as the problems get harder and as the groups get bigger, eventually reaching the optimum for the hardest problems and largest group. This points to cooperation being more useful in larger groups and for harder problems. In contrast, humans considerably under-perform when solving easy problems and/or when playing in isolation or small groups, which are the situations where cooperation appears not to work to the benefit of

players. As we show in what follows, this is related to excessive opportunistic behavior and intuitive of the benefits of collaboration.

Humans systematically collaborate less than what would lead to the maximal payoff, correctly judging the best collaboration level only when confronting the hardest problems or working in the largest group. We now examine the differences between C and C_{best} in relation to R and S, as reported in the Fig. 1, bottom two panels. No additional normalization is needed here.

In the bottom left panel of Fig. 1 (four plots, a–d), within each plot we show the values for four problem complexities R. When problems are simple (a), humans largely fail in estimating the optimal collaboration level, and do well only when in the largest group comprising all players. As problems get harder (b and c), humans increasingly better estimate the best collaboration level. In fact, the estimations are systematically better for larger groups. For hardest problems (d), humans correctly guess the best cooperation level regardless of the group size. In the bottom right panel in Fig. 1 we show the same values as in the bottom left panel, but now presenting them via group size S. When playing alone (a), humans generally fail to find the best cooperation level, but their estimation improves as the problems become harder. For groups of moderate size (b and c), humans increasingly better guess the optimal cooperation level. In fact, their guesses are improving as the problems get harder, and for hardest problems the estimations are correct. When playing all together in one group (d), humans collaborate at the optimal level for problems of any complexity. Interestingly, humans are never found to collaborate more than the optimal level (which could likewise go against their benefit).

These results are consistent with the findings for agent fitness from the two top panels in Fig. 1, suggesting a correlation between C and both R and S. Humans estimate the optimal collaboration level systematically better as the groups become bigger and the problems become harder. When playing in the largest group or when facing the hardest problems, humans correctly judge the best collaboration level. In contrast, when playing in isolation or when facing easy problems, humans largely fail in their collaborative strategy. These observations explain why the earning of humans are considerably smaller than what they could be, except in cases of largest group and/or hardest problems. This also indicates that collaboration–whose usefulness we understand via discrepancy between theoretical and real behavior–is more beneficial in larger groups and for harder problems. These seem to be the conditions when humans attain their full collaborative potential.

We also observe that both humans and evolving agents cooperate more in small groups than in large groups regardless of the problem complexity. Such behavior is known as *social loafing* and is often found in experiments[54,59–61]. Humans quickly lose interest in solving a problem when inserted in a group of increasing size, but instead count on others to solve it (free riding). On the other hand, both humans and evolving agents cooperate more as the problems get harder. In other words, humans behave opportunistically when facing easy problems, while they resort to collective actions when dealing with complex problems, which holds regardless of the group size. Also interesting is the group involving half of the players (division into 2 equal groups), where humans most drastically under-perform (except in the case of hardest problems). Social loafing here seems to have the strongest negative effect.

Humans systematically generate smaller group bonus than they could, reaching the maximal values only for the hardest problems or when working in the largest group. One may think that while under-performing for the case of their individual benefit (measured by AF), human choose the level of collaboration that actually maximizes the common benefit, measured by group bonus GB. This however is not the case: we compare the human GB with the optimal GB_{best} in the Supplement, and find a very similar pattern to what found for AF: human level of collaboration generates best GB_{best} only when facing hardest problems or when working in the largest group, and fails under other conditions. Therefore, social collaboration seems to be equally useful to humans for both individual and common benefit.

Psychological variables only weakly correlate with the propensity for collaboration. We also examined if the psychological variables measured via survey predict the collaborative behavior. We examined standard "Big 5" personality dimensions (Neuroticism, Surgency, Agreebleness, Closeness, and Conscientiousness), Honesty-Humility scale, State-Trait Anxiety Inventory, General Self-Efficacy Scale, and Social Community scale (see Supplement for details). Using standard statistical analysis we show that only a small part of the collaboration variance can be explained via psychological variables. This further confirms that R and S are indeed the only relevant variables affecting the collaborative behavior of players and the findings cannot be attributed to players' personal characteristics.

Discussion

In sum, we found that in our experiment humans perform best when joined in large groups to solve problems they see as difficult. Under these conditions humans most accurately estimate the level of collaboration that will maximize their benefit. In contrast, the same approach fails when humans act in small groups or face challenges they perceive as easy. These findings are robust to players' and groups' characteristics including personal traits, and appear to depend exclusively on the two game-theoretic variables in our setup: problem complexity and group size.

We interpret these findings as a result of players' intuitive (heuristic) decision-making. In fact, recent and not-so-recent results in cognitive psychology consistently show that humans are not always rational[62,63]. This was found in a series of experiments, indicating that our processing of information can follow either a fast and economic *heuristic* path or a slower and more resource-demanding *rational* path[56,64,65]. The choice between these two paths depends on many factors, such as emotional involvement with the situation[66,67] or time pressure to react in a situation[29,68,69]. Specifically, our (often) irrational drive for cooperation was recently articulated as *Social*

Heuristics Hypothesis[28]: It posits that through life experience we devise simple rules-of-thumb that enable us to instinctively adopt "socially best" choices, as opposed to a more rational, slow and often selfish consideration. And indeed, when the decisions are taken under time pressure, intuitive reasoning leads to more cooperation in social dilemma games[70,71] and in dictator games[72]. In contrast, with no rush to decide, humans perform a more accurate evaluation of their individual interests and cooperate less[73]. We appear to internalize and learn[74] simple intuitive strategies that are advantageous in our routine lives and "mistakenly" transplant them into less typical situations, where we would be (individually) better off with a more rational approach. Guiding ourselves on heuristics generally works to our benefit, but under specific conditions can also work against our benefit.

Coming back to our experiment, we cannot conclusively prove the absence of deliberative rational thinking in players' decision-making. However, below we discuss four independent lines of evidence pointing to this.

First, humans are known to generally apply intuition when limited information suggests that it could be the best strategy. Because time pressure regulates the speed by which subjects have to process the available information, so it can be seen as a proxy for the availability of information. Along the same lines, time pressure can also be seen as a way to manipulate the perceived complexity of the problem: solving a fixed problem under varying time constraints generates varying perception of the difficulty of that problem. In our experiment we simply use a different way to regulate the perceived complexity of the problem–we manipulate the probability for a player to win in a given round. This is a good proxy for a regulable problem complexity, much like the time pressure. So from this point of view we believe that our experimental setup is not that different from the setups in earlier studies on SHH.

Second, in our scoring scheme cooperators contribute not only to the common payoff, but also to the payoff of non-cooperators, who can further increase their individual payoffs by striving for a higher reward at the price of solving the harder task alone. In these conditions we expect that the cooperative decisions are taken via same heuristics as in the real life: by sharing a piece of information a player can increase the benefit for those who exploit this wisdom, whereas by choosing to work alone, a player can profit of the collective knowledge (articulated as group bonus) without contributing to it. As a side effect, this complicates the scoring scheme and makes it impossible for players to calculate any payoffs beforehand. In addition, players cannot infer the strategies of their peers since they have no access to their history of cooperative decisions. So, not having much to go on, we expect that players mainly followed their instincts from round to round. In other words, limited information about other players should preclude the application of any deliberative reasoning.

Third, when the community is divided into groups (that could also be competing among them), rational reasoning would make the collaboration level depend primarily on the group one belongs to and on the total number of groups, rather than on the problem complexity. In fact, we such behavior is found for evolving agents, whose cooperation deceases with increase of R, except in the case of the largest group ($S = N/1$, cf. Fig. 1), where the cooperation actually increases with increase of R. Humans instead display a different pattern: they increase their cooperation with increase of R for all group sizes, including the largest group, which is hard to ascribe to any systematic rational reasoning.

Fourth, the typical reaction time of players did not change significantly over different rounds and games. In fact, when players reason rationally they are known to be slower to make decisions[28,29]. Therefore, rational thinking would in our case make the reaction time vary or even increase as the game unfolds, since players "learn" how to play in their best interest. However, as we show in the Supplement, players' reaction times are not correlated with their propensity to collaborate, meaning that no learning effects are present. In other words, players in all rounds seem to make their decisions on the same ground as in the first round. But if players are not learning as the game unfolds, it is reasonable to hypothesize that they rely on a previously learned strategy. And all heuristics, including the social heuristics, are well-known to be socially learned via life experience[75,76]. So, while we cannot determine conclusively that players relied *exclusively* on heuristics and not on rational thinking, above body of evidence strongly points to this interpretation.

Game-theoretic nature of our experiment allowed us to use computer simulation to exactly reproduce the experimental conditions. But how realistic is to approximate a human player as an evolving agent defined by the constant cooperation probability? Human players do not make their decisions "randomly", but based on some strategy or feeling, which makes this approximation crude. However, as we explained above, cooperative decisions taken in separate rounds cannot be statistically distinguished across players nor across rounds. Hence, as far as the external observer (experimenter) is concerned, a player can be well characterized by a single number, which expresses how likely is that he/she will cooperate in any given round. This also indicates that even more a sophisticated simulation of our experiment is not likely to reproduce more accurate estimation of the best strategy.

To what degree are these findings dependent on the choice of the game, especially in relation to the opposition between individual and common benefit? We designed our game to best reproduce a cognitive scenario that would force the players to act as much as possible on intuition. A game with a strict opposition between individual and common benefit would place a "simpler" dilemma in front of the players, making it harder for us to exclude the rational thinking. In fact, replication of our experiment using other games, possibly with varying balance of individual and common benefit, is the prime direction of future work. One could also examine the influence of a combination of several factors on the usefulness of collaboration heuristics, for example, a combination of time pressure, group size and problem complexity. Other directions include replication of the experiment with larger groups and more realistic (ecological) problems. While the former is generally doable, the latter is associated with already discussed difficulty (operationalization of problem complexity). Actually, we admit that the ultimate interpretation and the value of our results might be depended on our definition of problem complexity. We warmly welcome any further work where players would be solving actual problems of gradually increasing complexity.

In conclusion, we argue that humans have evolved to solve demanding challenges best when joining forces in large groups, which is in agreement with the existing studies[21,77,78]. This explains why errors in our collaboration

judgment are made more often and more drastically in smaller groups and when dealing with easier problems. At a certain point in our evolutionary path[79] communities whose members have been able to cooperate in the appropriate measure have thrived, as opposed to those communities whose members collaborated too much or not enough. Actually, there are numerous hypotheses on the evolutionary mechanisms perpetuating the human cooperative behavior: group/kin selection, direct/indirect reciprocity, etc.[4,7,13,80]. We here argue that an additional factor has been the need to find the level of collaboration that would best benefit both the community and the individual.

References

1. Nowak, M. & Highfield, R. *Supercooperators: Altruism, evolution, and why we need each other to succeed.* (Simon and Schuster, 2011).
2. Perc, M. *et al.* Statistical physics of human cooperation. *Physics Reports* **687**, 1–51 (2017).
3. Helbing, D. *et al.* Saving human lives: What complexity science and information systems can contribute. *Journal of statistical physics* **158**, 735–781 (2015).
4. Rand, D. G., Ohtsuki, H. & Nowak, M. A. Direct reciprocity with costly punishment: Generous tit-for-tat prevails. *Journal of theoretical biology* **256**, 45–57 (2009).
5. Delton, A. W., Krasnow, M. M., Cosmides, L. & Tooby, J. Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences* **108**, 13335–13340 (2011).
6. Panchanathan, K. & Boyd, R. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499 (2004).
7. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437**, 1291 (2005).
8. Traulsen, A. & Nowak, M. A. Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences* **103**, 10952–10955 (2006).
9. Szolnoki, A. & Perc, M. Emergence of multilevel selection in the prisoner's dilemma game on coevolving random networks. *New Journal of Physics* **11**, 093033 (2009).
10. Lieberman, D., Tooby, J. & Cosmides, L. The architecture of human kin detection. *Nature* **445**, 727 (2007).
11. Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502 (2006).
12. Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences* **108**, 19193–19198 (2011).
13. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
14. Bowles, S. & Gintis, H. The origins of human cooperation. In P. Hammerstein, D. w. r. (ed.) *Genetic and cultural evolution of cooperation*, 429–44 (Cambridge, MA: MIT Press, 2003).
15. Jordan, J. J., Hoffman, M., Nowak, M. A. & Rand, D. G. Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences* **113**, 8658–8663 (2016).
16. Mann, R. P. & Helbing, D. Optimal incentives for collective intelligence. *Proceedings of the National Academy of Sciences* **114**, 5077–5082 (2017).
17. Howe, J. The rise of crowdsourcing. *Wired magazine* **14**, 1–4 (2006).
18. Surowiecki, J. *et al.* The wisdom of crowds. *American Journal of Physics* **75**, 190–192 (2007).
19. Szolnoki, A., Wang, Z. & Perc, M. Wisdom of groups promotes cooperation in evolutionary social dilemmas. *Scientific Reports* **2**, 576 (2012).
20. Guazzini, A., Vilone, D., Donati, C., Nardi, A. & Levnajić, Z. Modeling crowdsourcing as collective problem solving. *Scientific reports* **5**, 16557–16557 (2014).
21. Prelec, D., Seung, H. S. & McCoy, J. A solution to the single-question crowd wisdom problem. *Nature* **541**, 532–535 (2017).
22. Kurvers, R. H. *et al.* Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences* **113**, 8777–8782 (2016).
23. Lee, J. *et al.* Rna design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences* **111**, 2122–2127 (2014).
24. Horowitz, S. *et al.* Determining crystal structures through crowdsourcing and coursework. *Nature Communications* **7**, 12549 (2016).
25. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756 (2010).
26. Gowers, T. & Nielsen, M. Massively collaborative mathematics. *Nature* **461**, 879 (2009).
27. Sørensen, J. J. W. *et al.* Exploring the quantum speed limit with computer games. *Nature* **532**, 210 (2016).
28. Rand, D. G. *et al.* Social heuristics shape intuitive cooperation. *Nature communications* **5**, 3677 (2014).
29. Gallotti, R. & Grujic, J. From intuitive altruism to rational deliberations - a neuroscience view on the learning process in game theory experiments. *eprint arXiv:1807.07866* (2018).
30. Reicher, S. D. The psychology of crowd dynamics. In Tindale, M. A. H. R. S. (ed.) *Blackwell handbook of social psychology: Group processes*, 182–208 (Oxford, UK:Blackwell, 2001).
31. Isaac, R. M. & Walker, J. M. Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics* **103**, 179–199 (1988).
32. Isaac, R. M., Walker, J. M. & Williams, A. W. Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of public Economics* **54**, 1–36 (1994).
33. Barcelo, H. & Capraro, V. Group size effect on cooperation in one-shot social dilemmas. *Scientific reports* **5** (2015).
34. Capraro, V. & Barcelo, H. Group size effect on cooperation in one-shot social dilemmas ii: Curvilinear effect. *PloS one* **10**, e0131419 (2015).
35. Campbell, D. J. Task complexity: A review and analysis. *Academy of management review* **13**, 40–52 (1988).
36. Byström, K. & Järvelin, K. Task complexity affects information seeking and use. *Information processing & management* **31**, 191–213 (1995).
37. Helbing, D., Szolnoki, A., Perc, M. & Szabó, G. Evolutionary establishment of moral and double moral standards through spatial interactions. *PLoS Comput Biol* **6**, e1000758 (2010).
38. Nosenzo, D., Quercia, S. & Sefton, M. Cooperation in small groups: The effect of group size. *Experimental Economics* **18**, 4–14 (2015).
39. Wong, R. Y.-m. & Hong, Y.-y. Dynamic influences of culture on cooperation in the prisoner's dilemma. *Psychological science* **16**, 429–434 (2005).
40. Grujić, J., Fosco, C., Araujo, L., Cuesta, J. A. & Sánchez, A. Social experiments in the mesoscale: Humans playing a spatial prisoner's dilemma. *PloS one* **5**, e13749 (2010).
41. Gracia-Lázaro, C. *et al.* Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma. *Proceedings of the National Academy of Sciences* **109**, 12922–12926 (2012).
42. Capraro, V., Jordan, J. J. & Rand, D. G. Heuristics guide the implementation of social preferences in one-shot prisoner's dilemma experiments. *Scientific reports* **4**, 6790 (2014).
43. Grujić, J. *et al.* A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance. *Scientific reports* **4**, 4615 (2014).

44. Perc, M. Phase transitions in models of human cooperation. *Physics Letters A* **380**, 2803–2808 (2016).
45. Cuesta, J. A., Gracia-Lázaro, C., Ferrer, A., Moreno, Y. & Sánchez, A. Reputation drives cooperative behaviour and network formation in human groups. *Scientific reports* **5**, 7843 (2015).
46. Antonioni, A., Sanchez, A. & Tomassini, M. Global information and mobility support coordination among humans. *Scientific reports* **4**, 6458 (2014).
47. Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. Positive interactions promote public cooperation. *Science* **325**, 1272–1275 (2009).
48. Hauser, O. P., Rand, D. G., Peysakhovich, A. & Nowak, M. A. Cooperating with the future. *Nature* **511**, 220 (2014).
49. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proceedings of the National Academy of Sciences* **111**, 17093–17098 (2014).
50. Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous giving and calculated greed. *Nature* **489**, 427–430 (2012).
51. Cone, J. & Rand, D. G. Time pressure increases cooperation in competitively framed social dilemmas. *PLoS one* **9**, e115756 (2014).
52. Capraro, V. Does the truth come naturally? time pressure increases honesty in one-shot deception games. *Economics Letters* **158**, 54–57 (2017).
53. Capraro, V., Schulz, J. & Rand, D. G. Time pressure increases honesty in a sender-receiver deception game. Available at SSRN: https://doi.org/10.2139/ssrn.3184537 (2018).
54. Latane, B., Williams, K. & Harkins, S. Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology* **37**, 822 (1979).
55. Nowak, A., Szamrej, J. & Latané, B. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review* **97**, 362 (1990).
56. Gigerenzer, G. & Goldstein, D. G. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review* **103**, 650 (1996).
57. Maynard, D. C. & Hakel, M. D. Effects of objective and subjective task complexity on performance. *Human Performance* **10**, 303–330 (1997).
58. Walther, J. B. Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication* **4**, 443–479 (2011).
59. Chidambaram, L. & Tung, L. L. Is out of sight, out of mind? an empirical study of social loafing in technology-supported groups. *Information Systems Research* **16**, 149–168 (2005).
60. Piezon, S. L. & Ferree, W. D. Perceptions of social loafing in online learning groups: A study of public university and us naval war college students. *International Review of Research in Open and Distance Learning* **9**, 1–17 (2008).
61. Abraham, J. & Trimutiasari, M. Sociopsychotechnological predictors of individual's social loafing in virtual team. *International Journal of Electrical and Computer Engineering (IJECE)* **5**, 1500–1510 (2015).
62. Simon, H. A. Theories of bounded rationality. *Decision and organization* **1**, 161–176 (1972).
63. Kahneman, D. & Frederick, S. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment* **49**, 49–81 (2002).
64. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *science* **185**, 1124–1131 (1974).
65. Shah, A. K. & Oppenheimer, D. M. Heuristics made easy: an effort-reduction framework. *Psychological bulletin* **134**, 207 (2008).
66. Slovic, P., Finucane, M. L., Peters, E. & MacGregor, D. G. The affect heuristic. *European journal of operational research* **177**, 1333–1352 (2007).
67. Jaspersen, J. G. & Aseervatham, V. The influence of affect on heuristic thinking in insurance demand. *Journal of Risk and Insurance* **84**, 239–266 (2017).
68. Rand, D. G. Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science* **27**, 1192–1206 (2016).
69. Bobadilla-Suarez, S. & Love, B. C. Fast or frugal, but not both: Decision heuristics under time pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **44**, 24 (2018).
70. Rand, D. G. & Kraft-Todd, G. T. Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience* **8**, 300 (2014).
71. Capraro, V., Corgnet, B., Espín, A. M. & Hernán-González, R. Deliberation favours social efficiency by making people disregard their relative shares: evidence from usa and india. *Royal Society open science* **4**, 160605 (2017).
72. Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V. & Barcelo, H. Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General* **145**, 389 (2016).
73. Belloc, M., Bilancini, E., Boncinelli, L. & D'Alessandro, S. A social heuristics hypothesis for the stag hunt: Fast-and slow-thinking hunters in the lab. *CESifo Working Paper Series No.* 6824 (2018).
74. Capraro, V. & Cococcioni, G. Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making. *Proc. R. Soc. B* **282**, 20150237 (2015).
75. Gigerenzer, G. & Todd, P. M. Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart*, 3–34 (Oxford University Press, 1999).
76. Chaiken, S. & Eagly, A. H. Heuristic and systematic information processing within and beyond the persuasion context. In Uleman, J. S. & Bargh, J. A. (eds) *Unintended Thought*, 212–252 (Guilford Press, 1989).
77. Tausczik, Y. R., Wang, P. & Choi, J. Which size matters? effects of crowd size on solution quality in big data q&a communities. In *ICWSM*, 260–269 (2017).
78. Huang, X. & Tausczik, Y. Does group size affect problem solving performance in crowds working on a hidden profile task? In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, LBW027 (ACM, 2018).
79. Bowles, S. Group competition, reproductive leveling, and the evolution of human altruism. *Science* **314**, 1569–1572 (2006).
80. Bowles, S. & Gintis, H. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical population biology* **65**, 17–28 (2004).

Acknowledgements

Author Contributions

A.G. and Z.L. envisaged the problem, A.G. and F.B. designed the experiment, A.G., D.V. and E.I. designed the computer simulation, F.S. and E.I. conducted the experiment, A.G. and Z.L. analyzed the data, Z.L., F.S., E.I. and A.G., wrote the manuscript, all authors reviewed it.

Additional Information