LETTERS

Egalitarian motives in humans

Christopher T. Dawes[1], James H. Fowler[1], Tim Johnson[2,3], Richard McElreath[4] & Oleg Smirnov[5]

Participants in laboratory games are often willing to alter others' incomes at a cost to themselves, and this behaviour has the effect of promoting cooperation[1–3]. What motivates this action is unclear: punishment and reward aimed at promoting cooperation cannot be distinguished from attempts to produce equality[4]. To understand costly taking and costly giving, we create an experimental game that isolates egalitarian motives. The results show that subjects reduce and augment others' incomes, at a personal cost, even when there is no cooperative behaviour to be reinforced. Furthermore, the size and frequency of income alterations are strongly influenced by inequality. Emotions towards top earners become increasingly negative as inequality increases, and those who express these emotions spend more to reduce above-average earners' incomes and to increase below-average earners' incomes. The results suggest that egalitarian motives affect income-altering behaviours, and may therefore be an important factor underlying the evolution of strong reciprocity[5] and, hence, cooperation in humans.

Scarce resources create selective pressure for behaviours that influence how resources are divided in animal societies[6]. When the availability of resources is independent of the choices involved in acquiring them, such behaviours—for example, aggression[6] and begging[7]—can be understood as explicitly motivated by preferences for specific resource divisions. In potentially cooperative encounters, however, resources are produced through organisms' choices to cooperate or defect; thus, behaviours that alter resource allocations produced in cooperative encounters can be viewed as either promoting a cooperation norm or satiating a taste for particular resource divisions[4]. This has sparked considerable debate about how to model social choices[8–10] and has provoked questions[4] concerning the ultimate source of behaviours—such as the reward of contributors[3] and costly punishment of free-riders[1]—that promote cooperation.

For example, in the standard model of multi-person cooperation—the public goods game—cooperation and payoff are correlated. Individuals are endowed with a resource that can be contributed to a common pool; if contributed, the resource increases in value and is divided equally among group members. Social welfare is maximized if all group members contribute, whereas personal wealth is greatest when an individual retains her endowment and others contribute. Past research suggests that individuals are willing to punish those who do not contribute to the common pool[1] and to reward those who do[3]. This behaviour has been interpreted as cooperative norm enforcement[1], but because a player's contribution to the public good is proportional to her payoff from the public good, decreasing the payoff of a defector also has the effect of retrieving economic equality.

Attempts to separate norm enforcement from the pursuit of equality have been incomplete. For instance, one approach[11] alters the efficiency of punishment by making punishment costs equal to the amount punishment reduces incomes. Although this prohibits an individual from reducing inequality between herself and the punished individual, it does not prevent reduction of the standard deviation from the group mean. If a player possessing above average income reduces the income of a wealthier player, then the income difference between high earners and below average earners decreases. Even players with below average income can reduce total inequality if their income is closer to the group mean than the above average earner's income. Thus, even though the inequality between the punisher and the punished player remains the same, punishment can still serve an egalitarian motive in this design[11].

To separate motives, we use a simple experimental design to examine whether individuals reduce or augment others' incomes when there is no cooperative norm to advance (see Methods). We call these behaviours 'taking' and 'giving' instead of 'punishment' and 'reward' to indicate that income alteration cannot change the behaviour of the target. Subjects are divided into groups having four anonymous members each. Each player receives a sum of money randomly generated by a computer. Subjects are shown the payoffs of other group members for that round and are then provided an opportunity to give 'negative' or 'positive' tokens to other players. Each negative token reduces the purchaser's payoff by one monetary unit (MU) and decreases the payoff of a targeted individual by three MUs; positive tokens decrease the purchaser's payoff by one monetary unit (MU) and increase the targeted individual's payoff by three MUs. Groups are randomized after each round to prevent reputation from influencing decisions; interactions between players are strictly anonymous and subjects know this. Also, by allowing participants more than one behavioural alternative, the experiment eliminates possible experimenter demand effects[12]—if subjects were only permitted to punish, they might engage in this behaviour because they believe it is what the experimenters want.

Over the five sessions income alteration was frequent. Among participants, 68% reduced another player's income at least once, 28% did so five times or more, and 6% did so ten times or more. Also, 74% of participants increased another player's income at least once, 33% did so five times or more, and 10% did so ten times or more. Most (71%) negative tokens were given to above-average earners in each group, whereas most (62%) positive tokens were targeted at below-average earners in each group.

The size of income alterations varied with the relative income of the recipient (Fig. 1). Individuals who earned considerably more than other members of their group were heavily penalized. Subjects who earned ten MUs more than the group average received a mean of 8.9 negative tokens compared to 1.6 for those who earned at least ten MUs less than the group. In contrast, individuals who earned considerably less than other group members received sizeable gifts. Subjects who earned ten MUs more than the group average received a mean of 4 positive tokens compared to 11.1 for those who earned at least ten MUs less than the group. Individual spending decisions also suggest that subjects were influenced by concerns for inequality. On average, the bottom earner in each group spent 96% more on

[1]Department of Political Science, University of California, San Diego, California 92093, USA. [2]Center for Adaptive Behaviour and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany. [3]Department of Political Science, Stanford University, Palo Alto, California 94305, USA. [4]Department of Anthropology, University of California, Davis, California 95616, USA. [5]Department of Political Science, University of Miami, Coral Gables, Florida 33124, USA.

negative tokens than the top earner and the top earner spent 77% more on positive tokens than the bottom earner (both differences significant, Student's *t*-test, one-tailed, $P < 0.0008$).

Because choices to reduce or to augment others' incomes were costly and yielded no material gain, self-interested subjects had no incentive to engage in it. Those behaviours therefore might decline over time as subjects learn they are not profitable. However, period-specific taking and giving (Fig. 1) shows no consistent pattern over time. Mann–Whitney tests fail to reject the null hypotheses that the number of negative tokens received in period five by above-average earners is the same as that received in periods one to four ($P = 0.38$, two-tailed) or that the number of positive tokens received in period five by below-average earners is the same as that received in periods one to four ($P = 0.86$, two-tailed). Therefore, subjects' income-altering behaviour persists even after acquiring experience playing the game.

To explore how income affects these behaviours, we conduct Tobit regressions of negative and positive tokens received as a function of the positive and negative deviation of one's income from the average income assigned to other group members. We employ robust standard errors clustered on each experimental session. This method allows us to account for the fact that observations are independent only across sessions and that costly taking and giving are censored variables. When examining costly taking, the regression coefficient on 'negative deviation' is –0.45 ($z = -5.11$, $P < 0.001$), and 'positive deviation' is 0.74 ($z = 4.43$, $P < 0.001$); subjects' payoffs are reduced by nearly three-quarters of an MU for each additional MU of income they receive above the average income of other group members. The average income of other group members, when included in the regression, is insignificant ($z = 0.41$, $P = 0.69$). In a model of costly giving, the regression coefficient on 'negative deviation' is 0.83 ($z = 7.56$, $P < 0.001$) and 'positive deviation' is –0.22 ($z = -2.43$, $P = 0.02$); subjects' payoffs are increased by more than eight-tenths of an MU for each additional MU of income below the average income of other group members. The average income of other group members, when included in the regression, is insignificant ($z = -1.41$, $P = 0.18$).

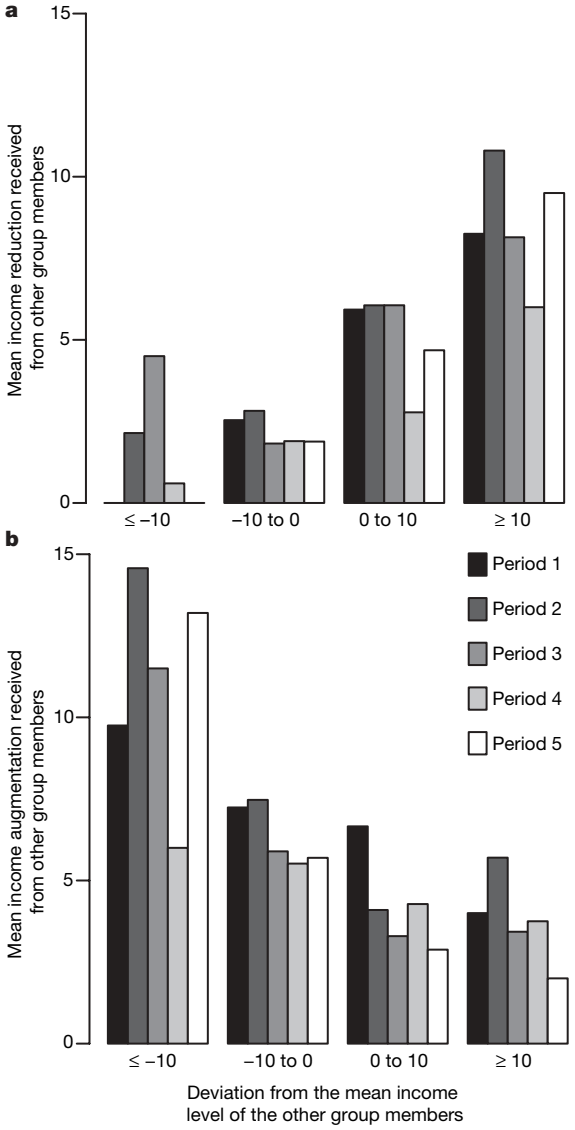

Figure 1 | Mean reduction (a) and augmentation (b) of income by other players in each period as a function of the deviation from the mean income level of the other group members. Income is assigned randomly to each group member by the computer. Reduction and augmentation are costly to the sender—each MU spent on 'costly taking' decreased the recipient's income by three MUs and each MU spent on 'costly giving' increased the recipient's income by three MUs.

We emphasize that income alteration provides no material benefit and, moreover, that a desire for revenge or reimbursement cannot explain choices to reduce or to augment others' incomes. Subjects were told that they never meet the same person twice, so they cannot satisfy, in future rounds, a desire to reciprocate negative or positive tokens assigned to them. To be sure that reciprocation was not a motivation, we conducted additional Tobit regressions. Results show that negative tokens sent were not significantly affected by negative tokens received in the previous round ($z = -0.30$, $P = 0.76$) and positive tokens sent were not significantly affected by positive tokens received in the previous round ($z = -1.17$, $P = 0.24$). Nonetheless, we did observe some behaviours that could not be explained by egalitarian motives. For example, below-average earners sent negative tokens to other below-average earners 12.2% of the time, while below-average earners sent above-average earners positive tokens 16.9% of the time (see Supplementary Information).

In our experiment there is no normative behaviour, so we wondered why people alter incomes. Others[1] show that experimental subjects feel anger towards free-riders in a public goods setting and this anger may motivate punishment. Also, negative emotions inspire the destruction of earned resources when an undeserving party aims to usurp those resources[13] and non-pecuniary expressions of anger satiate the desire to punish individuals who choose not to share a resource equally in experimental games[14]. Income levels are determined by subject behaviour in these experiments, so it is unclear whether resource distributions or anti-social behaviours cause the anger. One possibility is that inequality itself arouses negative emotions. If so, in our experiment we should observe annoyance and anger at high earners; these sentiments should increase as inequality increases and they should be associated with subjects' income-altering behaviours.

To elicit emotional reactions, we presented subjects hypothetical scenarios in which they encountered group members who obtained higher payoffs than they did (see Methods). Subjects were then asked to indicate on a seven-point scale whether they felt annoyed or angry (1, 'not at all'; 7, 'very') by the other individual. In the 'high-inequality' scenario, subjects were told they encountered an individual whose payoff was considerably greater than their own. This scenario generated much annoyance: 75% of the subjects claimed to be at least somewhat annoyed, whereas 41% indicated a high level (4 or more) of annoyance. Many subjects (52%) also indicated that they felt at least some anger towards the top earner. In the 'low-inequality' scenario, differences between subjects' incomes was smaller, and there was significantly less anger (Wilcoxon signed rank test, $P < 0.0001$) and annoyance ($P < 0.0001$). Only 46% indicated they were annoyed and 27% indicated they were angry. Individuals apparently feel negative emotions towards high earners and the intensity of these emotions increases with income inequality.

These emotions seem to influence behaviour. Subjects who said they were at least somewhat annoyed or angry at the top earner in the high-inequality scenario spent 26% more to reduce above-average earners' incomes than subjects who said they were not annoyed or angry. These subjects also spent 70% more to increase below-average earners' incomes. Mann–Whitney tests of both differences indicate that they are significant (one-tailed, $P = 0.05$ and $P = 0.001$, respectively). Emotional reactions towards high earners—even when the source of income is known to be purely random—cause individuals to engage in costly acts that promote equitable resource distributions.

The evidence here indicates that social inequality arouses negative emotions that motivate both the reduction and augmentation of others' incomes. This finding supports research that indicates humans are strongly influenced by egalitarian preferences[7,8]. Furthermore, the results distinguish between models of inequality aversion[8,9]: models that specify which players' incomes will be altered for egalitarian reasons[8] capture subject behaviour better than models that do not[9]. Finally, the results are also consistent with the punishment of non-contributors[1] and the reward of contributors in public good games[3]. Although concerns for equality are clearly not the only motive for human behaviour in these contexts, our results suggest that egalitarian motives may underlie strong reciprocity[11] and, thus, play an important role in the maintenance of cooperation.

METHODS

The design and procedures of the experiment closely approximate a widely cited public good experiment[1]. One hundred and twenty ($n = 120$) students from the University of California at Davis volunteered to participate in the experiment. Recruitment of subjects was conducted in several different departments to maximize the chance that subjects did not know one another; any student who was at least 18 years old was eligible to take part in the study. Twenty subjects attended each of the six experimental sessions and each session involved five periods. Every period, subjects were randomly placed in groups of four subjects. At the beginning of each period subjects received a random payoff and were shown the payoffs for all four members of their group. To maintain comparability with other public good games, random payoffs were drawn from the empirical distribution of payoffs in the first stage of a widely cited public good game with punishment[1]. Subjects were then given an opportunity to either help or harm any member of the group by purchasing up to ten positive and ten negative tokens for each player. At the end of each period, subjects learned the amount of positive and negative tokens they received and their new payoff. The experiment lasted 30 minutes and on average subjects earned approximately ten US dollars per session.

All activity in the experiment was completely anonymous. Group composition changed every period so that no one played with the same person more than once. The subjects were ignorant of other players' experimental history: neither past payoffs nor past decisions were known. Different group composition each period and the absence of any history of play ensured that subjects could neither develop reputations nor target other subjects for revenge.

At the beginning of each session subjects were asked to read experiment instructions on their individual computer screens (see Supplementary Information), and they also had a paper copy available for reference. The instructions explained all features of the experiment, including how payoffs are determined, how group composition is altered every period, and how anonymity of individual decisions and payoffs in the experiment is preserved. In order for the experiment to start, subjects had to answer correctly several test questions designed to ensure full understanding of how choices in the game generate payoffs. At the end of the experimental session, subjects were asked to complete a survey about their demographic characteristics and a questionnaire concerning emotions. The experiment was programmed using GameWeb software written by R.McE.

The emotions questionnaire presented two hypothetical scenarios to subjects: "You receive 23 [19] tokens. The second group member receives 25 [21] and the third 21 [17] tokens. Suppose the fourth member receives 37 [22] tokens. You now accidentally meet this member. Please indicate your feelings towards this person." (Unbracketed numbers were used in the 'high-inequality' scenario and bracketed numbers were used in the 'low-inequality' scenario.) After reading each scenario, subjects were asked to indicate on a seven-point scale whether they felt annoyed or angry (1, 'not at all'; 7, 'very').

Received 8 November 2006; accepted 5 February 2007.

1. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415,** 137–140 (2002).
2. Andreoni, J., Harbaugh, W. & Vesterlund, L. The carrot or the stick: rewards, punishments, and cooperation. *Am. Econ. Rev.* **93,** 893–902 (2003).
3. Sefton, M., Shupp, R. & Walker, J. The effect of rewards and sanctions in the provision of public goods. *Center for Applied Economics and Policy Research Working Paper* **2006–005,** 1–43 (2006).
4. Fowler, J. H., Johnson, T. & Smirnov, O. Egalitarian motive and altruistic punishment. *Nature* **433,** E1, doi:10.1038/nature03256 (2005).
5. Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13,** 1–25 (2002).
6. Maynard Smith, J. & Price, G. The logic of animal conflict. *Nature* **146,** 15–18 (1973).
7. Bergstrom, C. T. & Lachmann, M. Signaling among relatives. III. Talk is cheap. *Proc. Natl Acad. Sci. USA* **95,** 5100–5105 (1998).
8. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114,** 817–868 (1999).
9. Bolton, G. & Ockenfels, A. ERC: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **90,** 166–193 (2000).
10. Falk, A. & Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **54,** 293–315 (2006).
11. Falk, A., Fehr, E. & Fischbacher, U. Driving forces behind informal sanctions. *Econometrica* **73,** 2017–2030 (2005).
12. Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am. Psych.* **17,** 776–783 (1962).
13. Bosman, R. & van Winden, F. Emotional hazard in a power-to-take experiment. *Econ. J.* **112,** 147–169 (2002).
14. Xiao, E. & Houser, D. Emotion expression in human punishment behaviour. *Proc. Natl Acad. Sci. USA* **102,** 7398–7401 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Center for Adaptive Behaviour and Cognition at the Max Planck Institute for Human Development and the UC Davis Institute of Government Affairs for generous research support.

Author Contributions The authors are listed alphabetically because each author contributed equally to the design, implementation, analysis and communication of this research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.F. (jhfowler@ucsd.edu).