The collective-risk social dilemma and the prevention of simulated dangerous climate change

Manfred Milinski*[†], Ralf D. Sommerfeld*, Hans-Jürgen Krambeck*, Floyd A. Reed[‡], and Jochem Marotzke[†§]

*Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany; [‡]Department of Biology, University of Maryland, College Park, MD 20742-5815; and [§]Department of ''The Ocean in the Earth System,'' Max Planck Institute for Meteorology, 21046 Hamburg, Germany

Edited by Robert May, Oxford University, Oxford, United Kingdom, and approved December 18, 2007 (received for review October 8, 2007)

Will a group of people reach a collective target through individual contributions when everyone suffers individually if the target is missed? This ''collective-risk social dilemma'' exists in various social scenarios, the globally most challenging one being the prevention of dangerous climate change. Reaching the collective target requires individual sacrifice, with benefits to all but no guarantee that others will also contribute. It even seems tempting to contribute less and save money to induce others to contribute more, hence the dilemma and the risk of failure. Here, we introduce the collective-risk social dilemma and simulate it in a controlled experiment: Will a group of people reach a fixed target sum through successive monetary contributions, when they know they will lose all their remaining money with a certain probability if they fail to reach the target sum? We find that, under high risk of simulated dangerous climate change, half of the groups succeed in reaching the target sum, whereas the others only marginally fail. When the risk of loss is only as high as the necessary average investment or even lower, the groups generally fail to reach the target sum. We conclude that one possible strategy to relieve the collective-risk dilemma in high-risk situations is to convince people that failure to invest enough is very likely to cause grave financial loss to the individual. Our analysis describes the social window humankind has to prevent dangerous climate change.

cooperation | public good | threshold

Whenever people have to maintain a public resource, such as avoiding overfishing the oceans or protecting the global climate, they find themselves in a social dilemma, which has been called "the tragedy of the commons" (1). It is well known that over six billion "players" take part in a global climate tragedy of the commons, "a game that we cannot afford to lose" (2). However, humankind faces a dramatic change of living conditions on Earth when the already-rising global temperature passes a certain threshold: Dangerous climate change will occur (3–6). To reduce the risk of dangerous climate change, global greenhouse gas emissions should be reduced to $\approx$50% of the present level by 2050 (7). A reduction of this magnitude is beyond the capability of any single country and requires international cooperation. Herein lies a dilemma: substantial emissions reductions are likely to have negative short-term economic effects, but failure to accomplish this reduction may well incur dangerous climate change later, resulting in substantial human, ecological, and economic losses (5). Thus, a special type of social dilemma has to be solved, which we call the collective-risk social dilemma. Will a group of people reach a collective target through individual contributions when everybody would suffer if the group fails to achieve the target?

The collective-risk social dilemma has characteristic features that, taken together, distinguish it from other social dilemmas: (*i*) people have to make decisions repeatedly before the outcome is evident, (*ii*) investments are lost (i.e., no refunds), (*iii*) the effective value of the public good (in this case, the prevention of dangerous climate change) is unknown, and (*iv*) the remaining private good is at stake with a certain probability if the target sum is not collected. Threshold public goods games have been studied intensely theoretically (8, 9) and experimentally (refs. 10–15; see ref. 16 for a review). Of the above-mentioned characteristics, several studies investigated the exclusion of refunding (see table 1 in ref. 16), and one study tested the effect of limited information about the value of the public good and its distribution (17). However, these economically focused studies are designed to simulate not a dilemma similar to the climate problem but rather the funding for bridges, public roads, railway lines (14), or a new coffee pot for a community office (8). In contrast to these examples, the climate game involves investing in a public good, not to realize a gain but to avoid a loss. Thus, we expect the strategies adopted in the climate game to be risk-averse. Furthermore, many public goods games focus on a conflict between individual and group interests, but a major component of the climate problem is also a conflict between short- and long-term interests. Therefore, we propose the collective-risk social dilemma as a framework to investigate the inherent problems of avoiding dangerous climate change and other problems of this type.

The ability of people to assemble a target common good, such as food reserves, or to build up a collective defense system, i.e., a fence or a trench, might have been important at a time when humans lived in small communities that had to survive the winter or to defend their village. The building of an emergency sandbag levee by neighbors to protect their community from a flood has been important over the centuries. Half a fence or sand bag levee is hardly better than none. When there is no attack or high flood, the investment is lost. If the fence or sandbag levee is not complete and an attack or high flood occurs, all private goods are at stake. Many similar scenarios exist. Thus, there might exist a rich toolbox of social strategies of testing, signaling, or encouraging motivation to help collect a target common good in a collective-risk social dilemma. Nevertheless, the collective-risk social dilemma has not yet been analyzed by evolutionary theory. Our approach is experimental and might be an incentive for theoretical modeling.

In our simulation of the collective-risk social dilemma, every subject faces the same tradeoff: The more he or she invests in the collective good, the higher the probability that the group reaches the target sum, but the less money remains in his or her personal account, which he or she is guaranteed to receive in cash after the session if the target sum has been reached. In contrast, failure to reach the target sum implies a risk that the remaining money in the personal account will be lost. The social dilemma of this scenario adds another tradeoff: The more others invest, the less a subject needs to invest for the group still to reach its target sum. In this simulation, we interpret not reaching the target sum as failing to prevent dangerous climate change. Would others

Author contributions: M.M., R.D.S., F.A.R., and J.M. designed research; M.M. and R.D.S. performed research; H.-J.K. contributed new reagents/analytic tools; M.M. analyzed data; and M.M., R.D.S., F.A.R., and J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 2261.

[†]To whom correspondence may be addressed. E-mail: milinski@mpil-ploen.mpg.de or jochem.marotzke@zmaw.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709546105/DC1.

Table 1. The expected account values at the end of the game under three pure strategies (all players share the same strategy for the entire game)

Loss probability, %	Free rider, €	Fair sharer, €	Altruist, €
90	4	20	0
50	20	20	0
10	36	20	0

Free riders contribute €0 each round, fair sharers €2, and altruists €4. At a 90% probability of account loss, the optimal strategy is to contribute €2 each round to the collective. At a 10% probability of loss, the Free Rider strategy is rational, and at 50%, both of these strategies provide identical expected earnings.

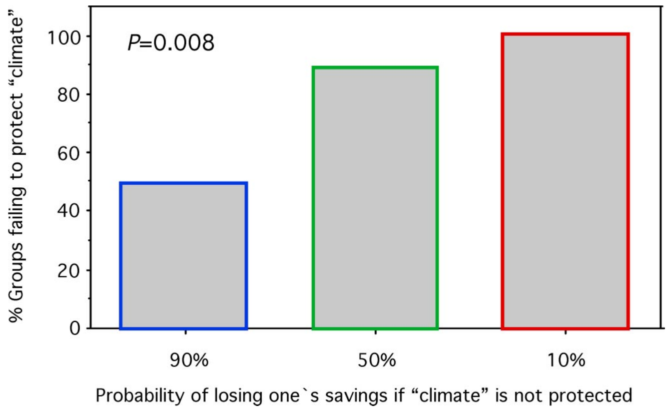

Fig. 1. Groups that either reach the target or fail. Displayed is the percentage of groups of six subjects each that fail to invest the target sum of at least €120 during 10 climate rounds, when they have a probability of 90%, 50%, or 10% of losing all their savings if the target sum is not reached. There are differences in not reaching the target among treatments ($P = 0.008$, $n = 30$, $df = 2{,}29$, $\chi^2 = 9.66$). The percentage of groups not reaching the target is significantly different from all groups reaching the target in each treatment (in each case: $n = 10$, $P < 0.0001$, binomial test).

compensate, that is, pay more than their fair share, if someone invested nothing or less than his or her fair share per round? Subjects were anonymous in the experiment, but all subjects could see how much money other subjects contributed after each round. Decisions from the same person could be recognized as such: Altruists, fair sharers, and free riders would become recognizable during the experiment, as well as subjects that change their strategy. Clearly, our groups of six subjects cannot represent all players of the climate game who individually decide how to travel, use fossil fuel, etc. If we find that all groups of six subjects collect the target sum, we cannot conclude anything for the climate game. However, if some of our small experimental groups fail to collect the target, larger groups would certainly fail with higher probability, and other solutions of the climate problems would have to be found.

Thirty groups of six students each took part in a public goods game (18) modified to simulate the collective-risk social dilemma via an interactive computer program. We provided the students with an endowment of €40 each, and the students knew they would be asked in each of 10 rounds to invest anonymously €0, €2, or €4 in a "climate account." They knew that, if the total contributions of the group reached or surpassed €120 at the end of the 10 rounds (equivalent to €2 per person per round on average), all group members would receive what they had not invested in climate change mitigation strategies in cash (for example, €20 if a person had invested €2 per round). If the group failed to reach this collective goal, the computer "threw dice," so that with a probability of 90% all group members would lose all their savings. In two other treatments, also with 10 groups of six students each, this probability was 50% and 10%, respectively (see *Methods*). With these treatments, failure to reach the target sum is still interpreted as analogous to incurring dangerous climate change but with a much-reduced risk of affecting individuals.

Independent of the treatment, groups who exactly reached the target sum collected, on average, €20 per player; the players then, on average, had invested their fair share. To maximize income, the best possible strategy would thus be to collect the target sum when failure would result in complete loss of money with a probability of 90%: If no one invests anything, any subject would receive his or her saved €40 only in one of 10 cases, that is, €4 on average (Table 1). This is much less than if the fair personal share is invested and the average €20 would be gained after the session. The groups taking part in the 50% treatment, on average, would gain €20 per person, either by investing the fair share and safely collecting €20 each or by investing nothing and on average gaining €20, that is, €40 in half of the cases. Groups in the 10% treatment on average gained most from investing nothing: If no one invested anything, any subject would receive €36 on average, because he or she would lose the saved €40 only in one of 10 cases (Table 1).

In the 90% treatment, each course of the game that leads to exactly reaching the target sum of €120, irrespective of who contributes how much as long as each player invests less than €36, is a Nash equilibrium: No single player can gain by deviating from his or her strategy (19–21). Once a single subject irrationally switches either to a strategy that jeopardizes reaching the target sum or that may exceed the target sum (both of which lowers his or her expected earnings), the remaining subjects' best strategies also involve a switch [Table 1; a description of equilibrium points and rational decisions is available as supporting information (SI) *Appendix*]. Depending on the current public sum and stage in the game, this switch might be either to "cut their losses" and also become free riders or to "maintain their investment" and increase the level of altruism. The latter is expected when one knows that the target might be missed because of some free riders; one invests more to secure some money. This represents a "snow drift game" scenario (22), where the best strategy is to cooperate when you know your partner will defect. Irrational responses such as the first switch stated above (that is, those that lower personal expectations) have been reported in several other game designs (e.g., ref. 23). These responses appear to be related to a sense of fairness, that is, situations that are perceived as unfair often result in players making less than optimal choices. In addition to purely self-interested people, there appears to be a fraction of people who are also motivated by these fairness considerations (24). One explanation for this behavior is that humans are better adapted to repeated game situations where fairness can become adaptive (e.g., ref. 25). A lot of opportunity exists in this climate protection game for unfair choices to be made, that is, free rider strategies can be adopted requiring other subjects to choose altruist strategies for the group still to reach the target sum (see *SI Appendix*), which, because of human behavior, jeopardizes the goal of preventing dangerous climate change.

Results and Discussion

Despite the difficulties outlined above, we found that 5 of the 10 groups in the 90% treatment successfully collected the target sum. This is significantly different ($n = 10$, $P < 0.0001$, binomial test) from the rational outcome of all groups reaching the target. Only one group in the 50% treatment and, as expected, none of the groups in the 10% treatment achieved the collective goal of preventing simulated dangerous climate change (Fig. 1). Surprisingly, almost all groups provided €2 per person in the first round, an all fair-sharer situation. However, in subsequent rounds, some subjects supplied nothing, which others clearly noticed, as stated in a questionnaire after the game; probably as a consequence, the

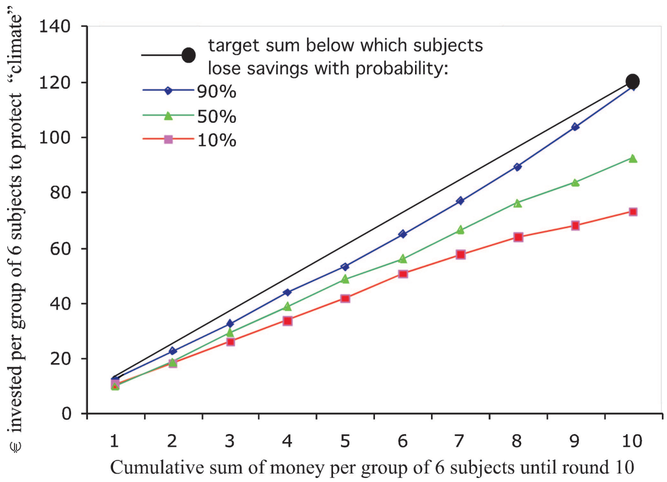

Fig. 2. Cumulative sum of money per group and round provided for the climate account. The target sum to be achieved after 10 rounds was €120; the treatments differed in the probability, i.e., 90%, 50%, and 10%, with which all subjects in a group lost their individual savings when the group did not supply the target sum for the climate account.

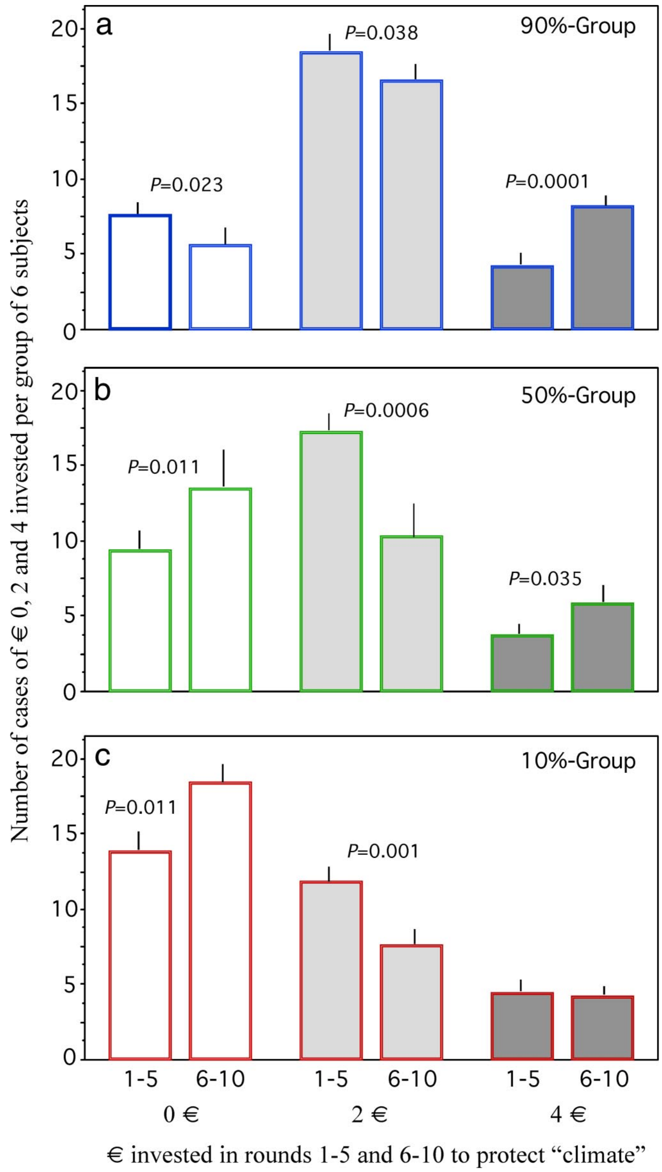

Fig. 3. Strategic behavior of subjects depending on the risk of losing all savings. Number of selfish acts, i.e., providing €0 (unfilled columns); fair-share acts, i.e., providing €2 (light-gray shading); and altruistic acts, i.e., providing €4 (dark-gray shading), per group of six subjects in the first (rounds 1–5, left in column pair) and second halves (rounds 6–10, right in column pair). The probability of losing all savings if the target sum is not reached was 90%, 50%, or 10%, for *a*, *b*, and *c*, respectively. See text for statistics.

motivation to contribute declined steadily. This was particularly noticeable for the 50% and 10% treatments, where many subjects reduced investments after round six. Providing €118.2 ± 1.9 (mean ± SE), the groups in the 90% treatment just failed on average to reach the target sum of €120 (Fig. 2). The five groups that did not reach the target sum collected on average (€112.8 ± 1.2). Just failing to reach the target represents the worst possible outcome: low individual savings and no collective benefit. The groups involved in the 50% and 10% treatments missed the target sum by far with €92.2 ± 9.0 and 73.0 ± 4.4, respectively (difference among the three treatments: $P = 0.0001$, $n = 30$, $df = 2.29$, $F = 13.784$). However, even the groups in the 10% treatment did not try to gain most on average by investing nothing (Fig. 2); against the rational strategy, they collected >50% of the target sum. This irrational behavior may indicate a high potential motivation also to protect the real climate and suggests future experiments where responses to other motivation factors could be studied. Perhaps this surprising behavior also stems from strong risk aversion, which has been shown to result in departures from standard Nash-equilibrium predictions (e.g., ref. 26). Some free riders did invest nothing and in 8 of 10 groups won their big savings. Group members invested differently in the climate account. Of the six members of each group, there were 3.3 ± 1.3, 2.1 ± 0.6, and only 1.1 ± 0.3 subjects in the 90%, 50%, and 10% treatments, respectively ($P = 0.003$, $n = 30$, $df = 2.29$, $F = 7.137$), that provided at least their fair share of €2 per round. The rest were free riders; the extreme free rider per group provided only €1.4 ± 0.1, 0.9 ± 0.2, and 0.5 ± 0.2 per round in the 90%, 50%, and 10% treatments, respectively ($P = 0.0003$, $n = 30$, $df = 2,29$, $F = 11.137$).

Interestingly, the dynamics of selfish acts (providing €0), fair-share acts (providing €2), and altruistic acts (providing €4) differed among and within treatments (Fig. 3). In the 90% treatment, selfish acts decreased from the first to the second half of the session ($P = 0.023$, $z = 2.739$, paired t test, two-tailed), as did fair-share acts ($P = 0.038$, $z = 2.433$), but altruistic acts strongly increased during the session ($P = 0.0001$, $z = -6.325$). This is mirrored by the slightly U-shaped curve describing the cumulative approach of the target sum: It is almost reached after round 10 on average (Fig. 2). This shift of strategies during the 90% treatment also illustrates the importance of short-term, in contrast to long-term, interests in individual strategy decisions during game play. Conversely, in the 50% treatment, selfish acts increased ($P = 0.011$, $z = -3.188$), but fair-share acts strongly decreased ($P = 0.0006$, $z = 5.161$), and altruistic acts slightly increased ($P = 0.035$, $z = -2.473$) (Fig. 3). In the 50% treatment, altruists obviously tried to compensate free riding of others, but usually in vain. In the 10% treatment, selfish acts increased from the first to the second half of the session; from an already high level ($P = 0.011$, $z = -3.196$), fair-share acts decreased ($P = 0.001$, $z = 4.776$), and altruistic acts were maintained at a low level ($P = 0.8$, $z = 0.259$).

In the 90% treatment, failure to achieve the target sum sometimes occurred in an extremely irrational way. Occasionally in the last round, it became clear that all of the contributions to the cumulative sum would have been in vain unless a large proportion of the players made a maximal contribution. Nevertheless, an insufficient number of players made this contribution, and the group just failed to reach the target sum (see also *SI Appendix*). Not contributing more at the end of the game, despite the risk of personal loss, may stem from a reluctance to reward behavior that is perceived as unfair.

When the risk of loss was no higher than the necessary average investment, our experimental groups generally failed to reach the target sum, which represented a threshold for preventing dangerous climate change. In contrast, groups with a high probability (90%) of losing their savings either succeeded in preventing simulated dangerous climate change or, half the time, came close to preventing it. Groups of six subjects may not represent the real climate game; our findings probably underrate the problems of the global challenge, because achieving cooperation is more difficult for larger than for smaller groups (27). Nevertheless, a proportion of the target could be achieved even in larger groups.

However, the number of countries with political representatives at climate summits or G8 summits is small compared with all individual players in the game. Perhaps our experimental setup may best be interpreted as representing such summits, with their smaller group size. Furthermore, in contrast to our experiments, climate or G8 summits are not anonymous, and it has been shown that anonymity is a strong impediment to preserving a common good (28). Despite its conceptual simplicity, our climate game thus holds promise for representing the social dilemma inherent in preventing dangerous climate change.

The strict cutoff of €120 for any common benefit to be realized may not be realistic for some hazards of climate change. However, many elements of climate change have "threshold" properties, consistent with our game design; adaptation to these abrupt changes is expected to be particularly difficult (6). Unlike in our game, real climate change, which is already underway, will not affect everyone equally; some countries and regions might even profit from it (5). However, this is different when dangerous climate change occurs, which is expected to cause such disruption that in a globalized world, no one would remain unaffected by it (3–5).

The collective-risk social dilemma seems to be a frequent phenomenon of human social life. Preventing dangerous climate change is but one example, albeit important, of this type of social dilemma. We therefore tried to simulate the inherent problems of preventing dangerous climate change in the present experiment. There are three conclusions from our experiment: (*i*) To achieve an effective level of voluntary individual cooperation, people have to be convinced of the very high probability that individuals will be affected by dangerous climate change, should the set target for the reduction of global greenhouse gas emissions not be met by the set date (here, 50% of the present level by 2050) (3, 7). If they believe in a lower probability, climate protection may fail, as it did in our 50% and 10% treatments. The high motivation to invest in the collective good that we found even in the 10% treatment suggests that informing people is worth the effort. (*ii*) We cannot rely on people to always behave rationally (23) (see also *SI Appendix*). Climate protection programs that appeal to a human sense of fairness (24), that is, all players contribute a "fair share" to the collective goal, are more likely to avoid irrational self-detrimental behavior. (*iii*) Because even our small experimental groups had difficulty preventing simulated dangerous climate change, the many players in the global game will certainly have more problems and may fail to prevent dangerous climate change. Climate or G8 summits may well have increased probability of success because of their smaller groups sizes. Also, incentives for investing in climate protection, other than the fear of suffering from dangerous climate change in the future, need to be offered, such as improved social reputation with its inherent benefits (28) or a combination of reputation and sanction institutions (29). Otherwise, the short-term advantages of free riding may fulfill Hardin's (1) prediction that "freedom in the commons brings ruin to all."

Methods

We tested 180 undergraduates who participated in 30 groups of six subjects each in a computerized experiment (e.g., refs. 28 and 29) at the University of Cologne and the University of Bonn. Our goal was to see whether they would contribute their own money, which they had received as the initial endowment, to prevent simulated dangerous climate change in a collective-risk social dilemma. The subjects knew they would obtain their money after the game in a way that did not disclose their identity. They were separated by opaque partitions and each had a laptop computer, on which they received the instructions for the experiment and with which they communicated their decisions to invest €0, €2, or €4 in the ''climate account'' to the main computer. After each of 10 rounds, the decisions of the six subjects were displayed on all computers simultaneously without any clue to a subject's identity. However, each subject's decision was listed in the same position after each round to allow individual strategies to be observed; thus, individual players were not anonymous within the game. Yet, they could not build a reputation from which to profit in another situation (e.g., ref. 28). Note that the cumulative sum of contributions was not displayed on the computer screen. Instead, the students were given pen and paper, and they were encouraged to take notes during the game.

The students knew that the total sum of money in the climate account, accumulated from all groups, would be used to publish a press advertisement on climate protection in a daily German newspaper simultaneously with the publication of the present study. However, they received the ''little information'' version from ref. 28 to explain the climate account, so that we could expect very weak motivation to invest in publishing the advertisement *per se*.

ACKNOWLEDGMENTS. We thank the students at the University of Bonn and the University of Cologne for participation; H. Arndt, T. Bakker, U. Ruschewitz, K. Schnetz, and T. Willecke for support; and O. Gaffney and two anonymous reviewers for comments on the manuscript. F.A.R. is supported by National Institutes of Health Grant F32HG03801.

1. Hardin G (1968) *Science* 162:1243–1248.
2. Pfeiffer T, Nowak MA (2006) *Nature* 441:583–584.
3. Schneider SH (2001) *Nature* 411:17–19.
4. Schellnhuber HJ, Cramer W, Nakicenovic N, Wigley T, Yohe G, eds (2006) *Avoiding Dangerous Climate Change* (Cambridge Univ Press, Cambridge, UK).
5. Intergovernmental Panel on Climate Change (2007) *Climate Change 2007: Impacts, Adaptation and Vulnerability. Summary for Policymakers. Contribution of Working Group II to the Fourth Assessment*. Accessed at www.ipcc-wg2.org/index.html.
6. Alley RB, Marotzke J, Nordhaus WD, Overpeck JT, Peteet DM, Pielke RA, Jr, Pierrehumbert RT, Rhines PB, Stocker TF, Talley LD, Wallace JM (2003) *Science* 299:2005–2010.
7. Intergovernmental Panel on Climate Change (2007) *Climate Change 2007: Mitigation of Climate Change, Summary for Policymakers. Contribution of Working Group III to the Fourth Assessment*. Accessed at www.ipcc-wg2.org/index.html.
8. Palfrey TR, Rosenthal H (1984) *J Pub Econ* 24:171–193.
9. Bagnoli M, Lipman BL (1989) *Rev Econ Stud* 56:583–601.
10. van de Kragt AJC, Orbell JM, Dawes RM (1983) *Am Pol Sci Rev* 77:112–122.
11. Dawes RM, Orbell JM, Simmons RT, Vandekragt AJC (1986) *Am Pol Sci Rev* 80:1171–1185.
12. Erev I, Rapoport A (1990) *J Conflict Res* 34:401–425.
13. Bagnoli M, McKee M (1991) *Econ Inquiry* 29:351–366.
14. Suleiman R, Rapoport A (1992) *J Behav Dec Making* 5:133–153.
15. Cadsby CB, Maynes E (1999) *J Pub Econ* 71:53–73.
16. Croson RTA, Marks MB (2000) *Exp Econ* 2:239–259.
17. Marks MB, Croson RTA (1999) *Public Choice* 99:103–118.
18. Ledyard, J. O (1995) in *Handbook of Experimental Economics*, eds Kagel JH, Roth AE (Princeton Univ Press, Princeton, NJ), pp 111–194.
19. Colman A (1998) *Game Theory and Its Applications*, (Garland Press, New York).
20. Camerer C (2003) *Behavioral Game Theory* (Princeton Univ Press, Princeton, NJ).
21. Thaler RH, Holt CA, Roth AE (2004) *Proc Natl Acad Sci USA* 101:3999–4002.
22. Doebeli M, Hauert C (2005) *Ecol Lett* 8:748–766.
23. Thaler RH (1988) *J Econ Perspectives* 2:195–206.
24. Fehr E, Schmidt KM (1999) *Q J Econ* 114:817–868.
25. Nowak MA, Page KM, Sigmund K (2000) *Science* 289:1773–1775.
26. Goeree JK, Holt CA, Palfrey TR (2003) *Games Econ Behav* 45:97–113.
27. Boyd R, Richerson PJ (1988) *J Theor Biol* 132:337–356.
28. Milinski M, Semmann D, Krambeck H-J, Marotzke J (2006) *Proc Natl Acad Sci USA* 103:3994–3998.
29. Rockenbach B, Milinski M (2006) *Nature* 444:718–723.