Science AAAS

Lab Experiments for the Study of Social-Ecological Systems

Marco A. Janssen, *et al.*

Science **328**, 613 (2010);
DOI: 10.1126/science.1183532

subtropics (~0° to 20°S) during late 2004 and late 2006, which result from enhanced austral spring burning over Indonesia during these years (*24*). However, direct transport to the stratosphere from these episodes appears smaller than the boreal summer sources linked to the Asian monsoon.

The exact causes of the enhanced tropical lower stratospheric HCN during 2005 and 2007 seen in Fig. 3 are difficult to determine from the limited sampling of the satellite observations. We have searched for systematic changes in transport or circulation of the Asian monsoon anticyclone during these years [or links to the stratospheric quasibiennial oscillation (QBO)], but we do not find obvious links to the enhanced HCN anomalies. Rather, it is likely that these patterns reflect variations in tropospheric sources, subsequently transported through the monsoon circulation; we note that the detailed attribution of such tropospheric sources is difficult based on the sparsely sampled ACE-FTS measurements. Recent model simulations of global HCN variability (*25*) suggest enhanced sources linked to the Indonesian fires in late 2004 and 2006, and the persistence into the following years and entrainment into the Asian monsoon circulation is reasonable given the long HCN photochemical lifetime in the free atmosphere.

These HCN observations demonstrate a large discernible chemical influence on the stratosphere from the Asian monsoon circulation. This pathway complements the large-scale troposphere-to-stratosphere transport that occurs in the deep tropics throughout the year (*26*), and there are likely distinct source regions for air within each pathway. Upwelling over the deep tropics primarily transports air with recent contact with the ocean surface and less concentrated anthropogenic influences. In contrast, transport in the monsoon region connects surface air with enhanced pollution (biomass and biofuel burning, plus urban and industrial emissions) to the lower stratosphere. Model calculations (*6*) suggest that surface emissions over a broad region covering India to eastern Asia are entrained into the monsoon circulation and transported to the lower stratosphere. This air will have enhanced black and organic carbon, sulfur dioxide (SO_2), reactive nitrogen species (NOx), and possibly short-lived halogen compounds from Asian industrial emissions, which have the potential to influence stratospheric ozone chemistry, aerosol behavior, and associated radiative balances. For example, a recent increase in background stratospheric aerosol concentrations has been observed, possibly linked to growth in SO_2 emissions over China since 2002 (*27*), and the monsoon is an effective pathway for such transport. The monsoon influence on the stratosphere is expected to become increasingly important given the ongoing growth of Asian emissions (*28*), with large continued increases over the next decades expected for SO_2 and NOx. Furthermore, potential changes in the strength and variability of the Asian monsoon circulation in an evolving climate [linked to increased convection and rainfall (*29*)] could modify this transport pathway, with potential influence on composition and climate of the stratosphere.

References and Notes

1. Q. Li *et al.*, *Geophys. Res. Lett.* **32**, L14826 (2005).
2. W. J. Randel, M. Park, *J. Geophys. Res.* **111**, (D12), D12314 (2006).
3. R. Fu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5664 (2006).
4. M. Park, W. J. Randel, A. Gettelman, S. Massie, J. Jiang, *J. Geophys. Res.* **112**, D16309 (2007).
5. M. Park *et al.*, *Atmos. Chem. Phys.* **8**, 757 (2008).
6. M. Park, W. J. Randel, L. K. Emmons, N. J. Livesey, *J. Geophys. Res.* **114**, D08303 (2009).
7. K. H. Rosenlof, A. F. Tuck, K. K. Kelly, J. M. Russell III, M. P. McCormick, *J. Geophys. Res.* **102**, (D11), 13,213 (1997).
8. A. Gettelman, D. E. Kinnison, T. J. Dunkerton, G. P. Brasseur, *J. Geophys. Res.* **109**, (D22), D22101 (2004).
9. R. N. Bannister, A. O'Neill, A. R. Gregory, K. M. Nissen, *Q. J. R. Meteorol. Soc.* **130**, 1531 (2004).
10. C. P. Rinsland *et al.*, *J. Geophys. Res.* **104**, (D15), 18,667 (1999).
11. H. B. Singh *et al.*, *J. Geophys. Res.* **108**, (D20), 8795 (2003).
12. Q. B. Li *et al.*, *J. Geophys. Res.* **108**, (D21), 8827 (2003).
13. A. Kleinböhl *et al.*, *Geophys. Res. Lett.* **33**, L11806 (2006).
14. A. Lupu *et al.*, *Atmos. Chem. Phys.* **9**, 4301 (2009).
15. P. F. Bernath *et al.*, *Geophys. Res. Lett.* **32**, L15S01 (2005).
16. The ACE-FTS is a solar occultation instrument that provides limited space-time sampling, and the climatological HCN patterns in Figs. 1 and 2, and figs. S1 and S2, are derived by combining data from all years during 2004 to 2009.
17. C. D. Boone *et al.*, *Appl. Opt.* **44**, 7218 (2005).
18. D. P. Edwards *et al.*, *J. Geophys. Res.* **109**, (D24), D24202 (2004).
19. Figure 1B shows results from the National Center for Atmospheric Research 3D Whole Atmosphere Community Climate Model (WACCM). This version of WACCM is relaxed to observed meteorological fields from the Goddard Modeling and Assimilation Office (GMAO) GEOS5.1 data assimilation system. HCN has been added to the standard chemical mechanism with a chemical loss by reactions with OH [with a corresponding lifetime of 4.3 years (*13*)] and with $O(^1D)$. The model also includes wet deposition through washout (which is weak because HCN is insoluble) and parameterized dry deposition over open-ocean [with a corresponding lifetime of 3 months (*12*)]. HCN emissions were determined by scaling CO emissions (using 0.012 HCN/CO molar ratio) for biomass burning and anthropogenic biofuel combustion.
20. R. A. A. Plumb, *J. Geophys. Res.* **101**, (D2), 3957 (1996).
21. H. C. Pumphrey, C. J. Jimenez, J. W. Waters, *Geophys. Res. Lett.* **33**, L08804 (2006).
22. H. C. Pumphrey, C. Boone, K. A. Walker, P. Bernath, N. J. Livesey, *Geophys. Res. Lett.* **35**, L05801 (2008).
23. P. W. Mote *et al.*, *J. Geophys. Res.* **101**, (D2), 3989 (1996).
24. J. A. Logan *et al.*, *Geophys. Res. Lett.* **35**, L03816 (2008).
25. Q. Li, P. I. Palmer, H. C. Pumphrey, P. Bernath, E. Mahieu, *Atmos. Chem. Phys.* **9**, 8531 (2009).
26. S. Fueglistaler, H. Wernli, T. Peter, *J. Geophys. Res.* **109**, (D3), D03108 (2004).
27. D. Hofmann, J. Barnes, M. O'Neill, M. Trudeau, R. Neely, *Geophys. Res. Lett.* **36**, L15808 (2009).
28. T. Ohara *et al.*, *Atmos. Chem. Phys.* **7**, 4419 (2007).
29. H. Ueda, A. Iwai, K. Kuwako, M. E. Hori, *Geophys. Res. Lett.* **33**, L06703 (2006).

30. The ACE mission is funded primarily by the Canadian Space Agency. Some funding was also provided by the UK Natural Environment Research Council (NERC). The U.S. National Center for Atmospheric Research is sponsored by the National Science Foundation. We also acknowledge support from the U.S. National Aeronautics and Space Administration.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1182274/DC1
SOM Text
Figs. S1 to S3

21 September 2009; accepted 3 March 2010
Published online 25 March 2010;
10.1126/science.1182274
Include this information when citing this paper.

Lab Experiments for the Study of Social-Ecological Systems

Marco A. Janssen,[1]* Robert Holahan,[2] Allen Lee,[1] Elinor Ostrom[1,2]

Governance of social-ecological systems is a major policy problem of the contemporary era. Field studies of fisheries, forests, and pastoral and water resources have identified many variables that influence the outcomes of governance efforts. We introduce an experimental environment that involves spatial and temporal resource dynamics in order to capture these two critical variables identified in field research. Previous behavioral experiments of commons dilemmas have found that people are willing to engage in costly punishment, frequently generating increases in gross benefits, contrary to game-theoretical predictions based on a static pay-off function. Results in our experimental environment find that costly punishment is again used but lacks a gross positive effect on resource harvesting unless combined with communication. These findings illustrate the importance of careful generalization from the laboratory to the world of policy.

Designing and conducting laboratory experiments in the social sciences enables the unpacking of complex problems to examine the effects of different components on outcomes and to replicate results with diverse participants (*1*). In this report, we discuss an experimental research program on the study of social-ecological systems, especially the governance of common-pool resources (CPRs). CPRs are resource systems where the harvesting of resource units by one user subtracts units from a pool potentially available to others. Examples in the field include forests, pastures, fisheries, and water systems.

The widely accepted economics textbook model of CPRs (*2*, *3*) is a simple, static production function that is used to conclude that the users of a

CPR would drastically overharvest or exhaust it. In his often-cited article, Hardin (*4*) concluded that overharvesting of a CPR was inevitable unless an external authority imposed rules on the helpless users. Hardin's judgment has been widely accepted because of its consistency with predictions from resource economics (*3*) and noncooperative game theory (*5*) and with well-publicized examples of resource collapses (*6*, *7*).

Extensive field research has challenged the prediction that it is impossible for users to self-organize (*8*). Many users have crafted their own institutions to overcome the temptations to overharvest, but others have not. Successful efforts reflect the struggle involved to overcome the incentives to overharvest and the costs of self-organization (*9*).

In light of substantial fieldwork on CPRs and growing use of experimental methods related to a variety of social dilemma games, researchers initiated the first CPR experiments in the late 1980s to test the simple model that was accepted as representing the core dilemma faced by harvesters (*10*). The initial CPR experiments focused on testing the accepted economic model of resource harvesting. We present an experimental environment that goes beyond traditional dilemma experiments to include spatial and temporal dynamics as used in laboratory experiments of complex systems (*11*–*13*) but not yet of CPR dilemmas.

When participants in the initial CPR experiments made independent and anonymous decisions, they substantially overharvested as predicted (*10*). Keeping the underlying mathematical structure representing the costs and benefits of harvesting constant, scholars slowly added variables identified as present in successful and unsuccessful field sites to the experimental settings and found that several made a substantial difference. Allowing participants to engage in face-to-face communication, called "cheap talk" by game theorists, in a CPR experiment where contribution levels were still made anonymously made a very substantial difference (*10*), as it had in many other dilemma situations (*14*).

Further, enabling participants to pay a fee in order to fine another participant also improved gross benefits in CPR experiments (*10*), as well as in some public-good experiments (*15*, *16*). In multiple CPR and public-good experiments, however, the costs of punishment outweighed the benefits of increased cooperation (*10*, *17*–*19*). In the CPR experiments, net benefits became positive when a further experimental enhancement was made—allowing the participants to communicate among themselves and to decide whether or not to adopt a sanctioning system and how much the fines and fees should be (*10*, *20*, *21*). In public-good experiments, increasing the number of rounds also led to net benefits (*22*), as well as combining communication with punishment (*19*, *23*).

The experiments reported herein take an important step toward approximating more closely the decisions users face in field settings, where the decision to harvest usually involves spatial and temporal dynamics instead of simple decisions regarding how much to harvest in an unchanging ecology (*24*–*26*). In a fishery, for example, the fish move rapidly from one location to another. Fishers have to figure out where and how many fish to collect without knowing for sure the specific benefits and costs of each harvesting decision (*3*, *26*).

In previous resource dilemma experiments, participants made decisions individually on a round-by-round basis. Each round typically required one decision. In field settings, decision-making does not have this orderly fashion. The study of dynamic decision-making includes decisions made in context and over time (*11*). Computerized microworlds are used to study dynamic decision-making such as fighting forest fires or leading an organization (*12*). We used methods of dynamic decision-making in order to perform controlled experiments that examine the relevant complexity of social-ecological systems.

Our experimental environment is built as follows (*27*). In each experiment, a group of five participants harvests tokens from a shared renewable resource on a 29-by-29 computer-simulated grid of cells (Fig. 1A). The resource's renewal rate is density dependent (Fig. 1B) to reflect simple ecological dynamics. The participants collect tokens in real time by pressing the arrow keys (left, right, up, and down) to move their avatars around the screen and pressing the space bar to collect a token from the cell on which the avatar is located. Each token is worth 2 cents. Participants make multiple decisions within each 4-min decision period. When participants behave like short-term, entirely self-interested actors, the resource will rapidly be exhausted, the group will collect only a few more than the initial number of 210 tokens, and the participants will face an empty screen for the remaining time in the decision period. On the other hand, if participants restrain themselves and deliberately think about where and when to harvest, the group earnings can increase to 665 tokens in a 4-min decision period [the average expected maximum is derived in the supporting

[1]Arizona State University, Post Office Box 872402, Tempe, AZ 85287–2402, USA. [2]Indiana University, 513 North Park Avenue, Bloomington, IN 47408–3895, USA.

*To whom correspondence should be addressed. E-mail: Marco.Janssen@asu.edu

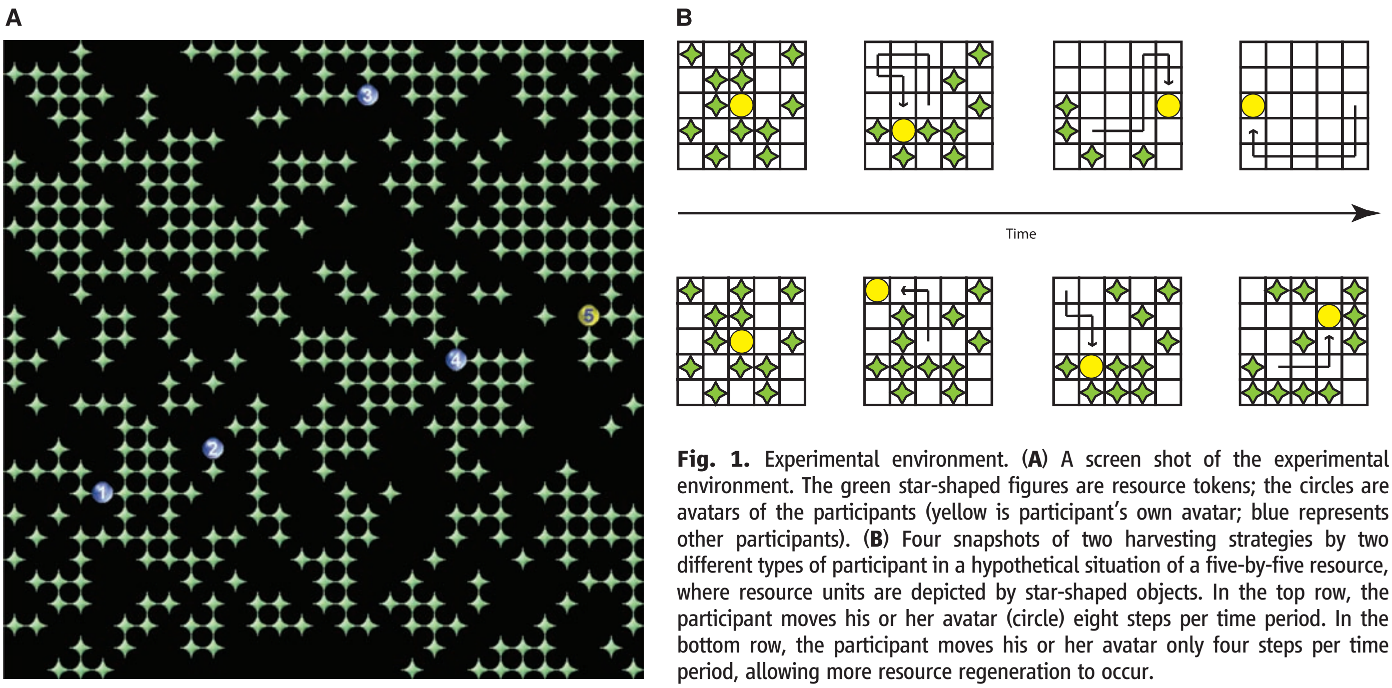

Fig. 1. Experimental environment. (**A**) A screen shot of the experimental environment. The green star-shaped figures are resource tokens; the circles are avatars of the participants (yellow is participant's own avatar; blue represents other participants). (**B**) Four snapshots of two harvesting strategies by two different types of participant in a hypothetical situation of a five-by-five resource, where resource units are depicted by star-shaped objects. In the top row, the participant moves his or her avatar (circle) eight steps per time period. In the bottom row, the participant moves his or her avatar only four steps per time period, allowing more resource regeneration to occur.

online material (SOM)]. In the optimal strategy, the group maintains a 50% density of tokens for most of the decision period and then rapidly harvests the remaining tokens at the end of the decision period. In periods where punishment is allowed, participants can subtract two tokens from another participant at the cost of one of their own tokens. Written communication, when implemented, takes place via text messages in a "chat room" before a decision period.

We conducted a series of experiments to test the impact of communication and punishment. The findings that participants use costly punishment, contrary to theoretical predictions, has stimulated a large number of CPR and public-good experiments during the past decade to test the generality of the earlier findings (*16*). In previous experiments, all participants first made an investment decision and then were given a special decision moment when they could decide to punish one of the other participants. In the current experimental environment, all decisions are real time (within the 4-min period).When the period includes costly punishment, participants can pay to punish whenever they see reason to as long as they have funds in their account and tokens remain on the screen to be harvested.

To test the effect of costly punishment versus communication, we performed a series of experiments using six different treatments. Each treatment consisted of three consecutive 4-min periods of costly punishment (P), communication (C), or a combination of both (CP) and three consecutive 4-min periods when neither communication nor punishment (NCP) is allowed. All treatments thus consisted of six decision periods, each lasting 4 min. Half the treatments started with NCP and the others finished with NCP. Each treatment was run five or six times. In total, 165 persons participated in 33 groups.

When participants started with three periods of NCP, the resource was consistently depleted within about 90 s, confirming that without communication or punishment, the "tragedy of the commons" prediction of Hardin is supported (Fig. 2). In this treatment, we found that the depletion of the resource occurred even faster in periods 2 and 3 than it did in period 1.

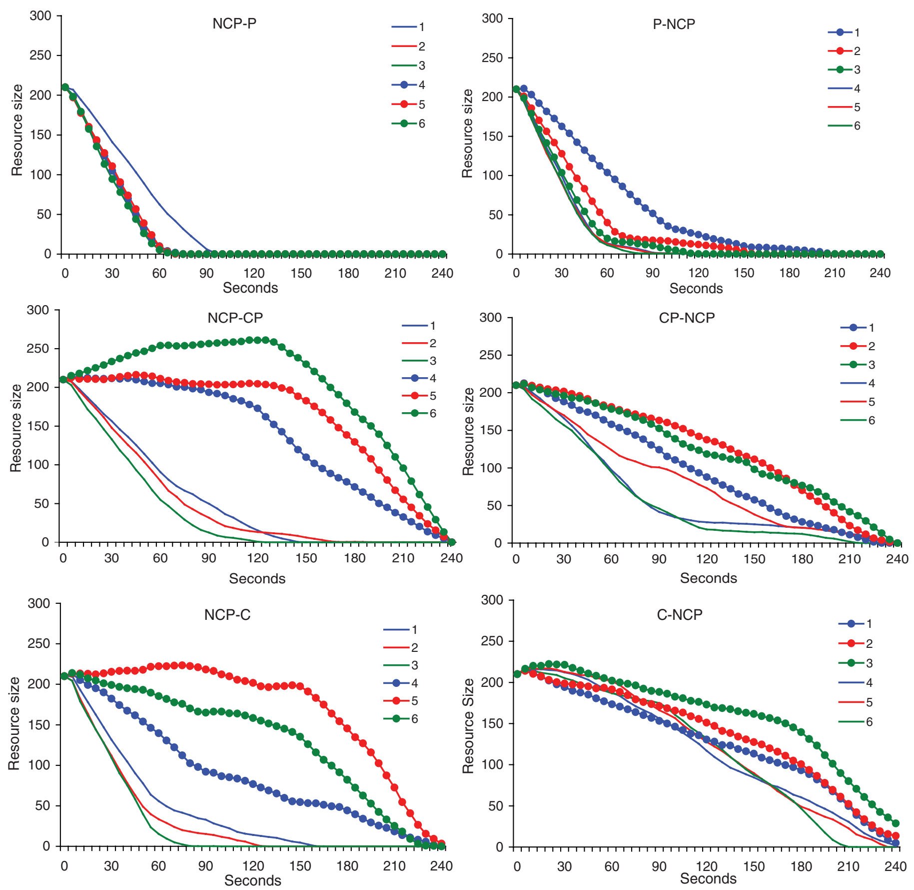

Fig. 2. Resource availability at given times. The diagrams show the average remaining level of the resource for the five or six groups of each treatment. Each diagram shows a treatment condition, and each line represents a particular period. The treatment is a combination of two sets of three periods of a specific condition. The names for these conditions are noted in the upper left of each display: NCP for neither communication nor costly punishment, C for communication, P for costly punishment, CP for communication and costly punishment. A treatment A-B refers to condition A for the first three periods and B for the last three periods. The colors and shapes referring to data of each period are noted in the upper right.

Periods with communication lead on average to a slower harvesting rate and more resource regeneration. Thus, earnings in periods 4 to 6 of C were significantly higher than in periods 1 to 3 of NCP. The strategies discussed by the participants in the C periods focused on the timing and location of harvesting. A common strategy worked out during the communication phase was to refrain from any harvesting for a set length of time, thereby allowing the resource to regenerate (see examples in SOM). When participants started with C periods, the amount of earnings did not drop significantly when participants were no longer able to communicate after period 3 (Fig. 3).

When P was introduced in period 4 after the first three periods were NCP, significant reductions in gross earnings occurred (Fig. 3), with an average of three punishment events per period. When participants started with CP periods and ended with NCP periods, significant reductions in the gross levels of tokens collected from the resource also occurred in the last three periods. Reduction in earnings in the last NCP periods were not found, however, when C is used in the first three periods. A puzzling finding is why communication with punishment does not lead to as long-lasting cooperative behavior as communication without punishment. Additional analysis (see SOM) shows that when participants used punishment in CP periods, we observed a reduction of the earnings in the last three NCP periods. This did not happen when participants did not use punishment in CP periods. It appears that the use of punishment erodes any cooperative agreements that were made, which are then less likely to persist when punishment and communication are not available anymore.

Punishment was not used by the participants in half of the periods when it was allowed. On average, there were 2.03 punishment events if communication was possible before the periods and 3.09 punishment events if communication was not allowed before the periods. In experiments with and without communication, a participant who was punished punished back 9 and 13% of the time, respectively. This appears to be a form of retaliation. Public-good studies that allow counterpunishment also found low levels of punishment and cooperation (*28*, *29*).

In the postexperiment survey, participants indicated that an important reason for the reluctance to punish others when communication was not allowed was their fear of retaliation (table S6). This reason was not mentioned for the treatments in which participants can communicate. If participants can communicate, the main reason they gave for punishing others, when they did, was for not following the agreements. Going too fast or having too many tokens were given as other reasons for punishment when participants cannot communicate, which is confirmed by the analysis of punishment events. In addition to using costly punishment, participants will scold others whom they consider to be free riders if communication is enabled (see examples in the SOM).

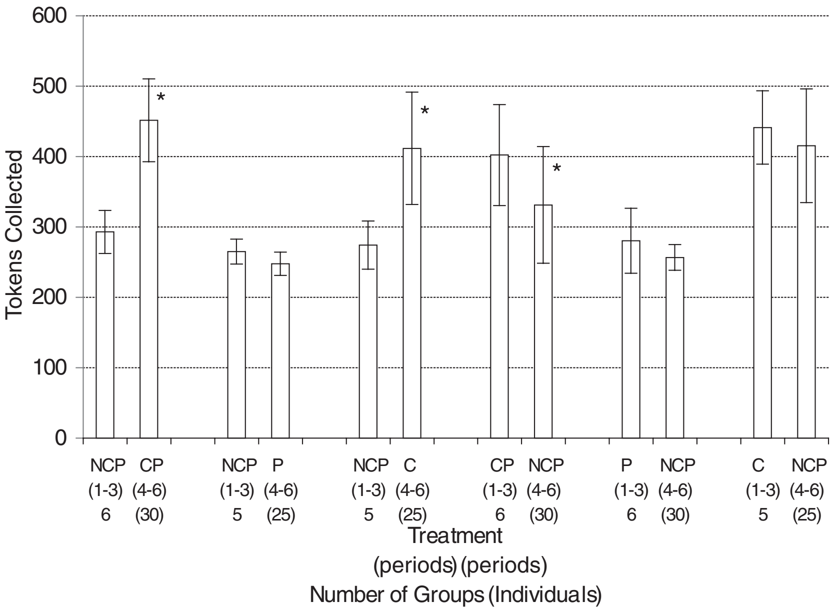

Fig. 3. Average net number of tokens collected by groups per period. The tokens lost due to punishment are subtracted from the total tokens harvested. Six different treatments are distinguished with combinations of neither communication nor costly punishment (NCP), communication (C), costly punishment (P), or communication and costly punishment (CP). Asterisks (*) refer to statistically significant differences between the first three and last three periods ($P < 0.01$) using a pairwise, two-tailed Mann-Whitney test.

Table 1. Punishment, communication, and harvesting levels. A multilevel mixed-effects linear regression is performed with the gross number of tokens that groups collected for each period. The independent variables are a set of dummy variables: whether participants could communicate and/or punish during the period and whether participants could have communicated and/or punished during the first three periods. Learning is tested by the effect of experiencing the same condition during multiple periods by including a dummy variable that indicates whether it is the first, second, or third time in this condition. LearnNCP is zero when it is not in the NCP condition, 1 for the first time in a NCP condition, 2 for the second time, and 3 for the third time. LearnCP, LearnC, and LearnP are defined in the same way.

Independent variables	Dependent variable: tokens harvested by group (SE)
Constant	298.147** (13.494)
Communication in current period (0 = no, 1 = yes)	92.130** (23.157)
Punishment in current period (0 = no, 1 = yes)	3.862 (22.475)
Communication and punishment in current period (0 = no, 1 = yes)	13.015 (26.944)
Communication in first three periods (0 = no, 1 = yes)	121.260** (16.744)
Punishment in first three periods (0 = no, 1 = yes)	–31.554* (15.786)
Communication and punishment in first three periods	–17.230 (26.476)
LearnNCP	–9.576* (4.818)
LearnCP	4.314 (3.505)
LearnC	17.111* (8.649)
LearnP	–13.127 (8.255)
–Log likelihood	1038.726
Number of decision periods	198
Variance contributions	
Group	38.567 (5.709)
Individual	39.138 (2.164)
χ^2	66.55 (P < 0.001)

*P < 0.05, **P < 0.01

A statistical analysis in Table 1 summarizes our findings. Communication leads to a significant increase in the number of tokens groups collected. In each C period, the number of tokens collected increases. When communication is not allowed in subsequent periods, previous communication still has a positive effect on the level of cooperation. The number of tokens collected remains significantly higher than without any communication.

Why does costly punishment (without communication) in a dynamic spatial environment lack a positive effect on resource use? In a modestly complex dynamic and spatial environment where participants can punish back but cannot discuss why they are punished, receiving a sanction does not carry a clear message. Does the sanction relate to the amount harvested, the location, the spatial pattern of harvesting, the speed in which the avatar moves over the screen, etc.? Communication can answer these questions and becomes an important attribute of the experiment for raising payoffs. When communication is possible, punished participants correct their harvesting rate by slowing down. On the other hand, punished participants in periods without communication do not behave differently from other participants (table S6).

What makes communication effective and able to affect decisions even after it is no longer feasible? Why is costly punishment ineffective in the short run with negative effects in the long run? Although the effectiveness of communication for small groups has been known for a long time, scholars propose diverse motivations (*30*). Communication can affect the understanding participants have of the resource system (*31*), change the expectations of others' behavior (*32*), coordinate strategies (*31*), or create the feeling of peer pressure (*31*, *32*) or a "group feeling" (*31*). When groups in the field are dependent on the resources, can meet from time to time to discuss the problems they face, and can make their own agreements, they are more likely to self-organize to govern the commons (*24*).

We have presented an experimental environment that begins to capture the spatial and temporal complexity of field settings. Our experiments confirm that participants will use costly punishment. The use of punishment without communication, however, does not increase gross payoffs. When communication is allowed, the performance of the group increases significantly. The performance is not sustained when punishment is used and communication is no longer possible. These results stress the importance of communication in commons dilemmas. This new experimental environment will enable scholars to test the generalizability of these results for different contexts such as mobile versus stationary resource units, visibility of resource extraction activities, and predictability of resource dynamics (*24*).

In order to translate findings from experimental research to policy analysis for social-ecological systems, it is important to understand the structures of both the social systems and the resource systems. Field studies have established the importance of including this complexity, and we have demonstrated how experimental research can begin to introduce more of the spatial and temporal processes found in many social-ecological systems.

References and Notes

1. A. Falk, J. J. Heckman, *Science* **326**, 535 (2009).
2. H. S. Gordon, *J. Polit. Econ.* **62**, 124 (1954).
3. C. W. Clark, *The Worldwide Crisis in Fisheries* (Cambridge Univ. Press, Cambridge, 2006).
4. G. Hardin, *Science* **162**, 1243 (1968).
5. P. S. Dasgupta, G. M. Heal, *Economic Theory and Exhaustible Resources* (Cambridge Univ. Press, Cambridge, 1979).
6. J. Radovich, in *Resource Management and Environmental Uncertainty*, M. Glantz, J. D. Thompson, Eds. (Wiley, New York, 1981), pp. 107–136.
7. R. A. Myers, J. A. Hutchings, N. J. Barrowman, *Ecol. Appl.* **7**, 91 (1997).
8. National Research Council, *The Drama of the Commons* (National Academies Press, Washington, DC, 2002).
9. T. Dietz, E. Ostrom, P. C. Stern, *Science* **302**, 1907 (2003).
10. E. Ostrom, J. M. Walker, R. Gardner, *Am. Polit. Sci. Rev.* **86**, 404 (1992).
11. B. Brehmer, *Acta Psychol. (Amst.)* **81**, 211 (1992).
12. D. Dörner, *The Logic of Failure* (Perseus Books, Reading, MA, 1996).
13. R. L. Goldstone, M. E. Roberts, T. M. Gureckis, *Curr. Dir. Psychol. Sci.* **17**, 10 (2008).
14. D. Sally, *Rationality Soc.* **7**, 416 (1995).
15. E. Fehr, S. Gächter, *Am. Econ. Rev.* **90**, 980 (2000).
16. S. Gächter, B. Herrmann, *Philos. Trans. R. Soc. London Ser. B* **364**, 791 (2009).
17. B. Herrmann, C. Thöni, S. Gächter, *Science* **319**, 1362 (2008).
18. M. R. Sefton, R. Shupp, J. Walker, *Econ. Inq.* **45**, 671 (2007).
19. O. Bochet, T. Page, L. Putterman, *J. Econ. Behav. Organ.* **60**, 11 (2006).
20. Ö. Gürerk, B. Irlenbusch, B. Rockenbach, *Science* **312**, 108 (2006).
21. M. Casari, L. Luini, *J. Econ. Behav. Organ.* **71**, 273 (2009).
22. S. Gächter, E. Renner, M. Sefton, *Science* **322**, 1510 (2008).
23. O. Bochet, L. Putterman, *Eur. Econ. Rev.* **53**, 309 (2009).
24. E. Ostrom, *Science* **325**, 419 (2009).
25. E. Schlager, W. Blomquist, S. Y. Tang, *Land Econ.* **70**, 294 (1994).
26. J. A. Wilson, J. M. Acheson, M. Metcalfe, P. Kleban, *Mar. Policy* **18**, 291 (1994).
27. Materials and methods, experimental design and protocol, and additional analysis are available as supporting material on *Science* Online.
28. L. Denant-Boemont, D. Masclet, C. N. Noussair, *Econ. Theory* **33**, 145 (2007).
29. N. Nikiforakis, *J. Public Econ.* **92**, 91 (2008).
30. N. R. Buchan, E. J. Johnson, R. T. A. Croson, *J. Econ. Behav. Organ.* **60**, 373 (2006).
31. G. Bornstein, in *Social Dilemmas*, W. Leibrand, D. Messick, H. Wilke, Eds. (Pergamon Press, Oxford, UK, 1992), pp. 247–263.
32. D. M. Messick, M. B. Brewer, in *Review of Personality and Social Psychology*, L. Wheeler, P. Shaver, Eds. (Sage, Beverly Hills, CA, 1983), pp. 11–44.

33. This research was funded by the U.S. NSF (grants BCS 0625354 and SES 0748632). We thank R. Goldstone, R. Tobias, J. Walker, and the Workshop's Experimental Reading Group for helpful comments and the Interdisciplinary Experimental Laboratory at Indiana University for providing use of its facility.

Supporting Online Material
www.sciencemag.org/cgi/content/full/328/5978/613/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S7

19 October 2009; accepted 22 March 2010
10.1126/science.1183532

Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare

Robert Boyd,[1,2]* Herbert Gintis,[2,3,4]* Samuel Bowles[2,5]*

Because mutually beneficial cooperation may unravel unless most members of a group contribute, people often gang up on free-riders, punishing them when this is cost-effective in sustaining cooperation. In contrast, current models of the evolution of cooperation assume that punishment is uncoordinated and unconditional. These models have difficulty explaining the evolutionary emergence of punishment because rare unconditional punishers bear substantial costs and hence are eliminated. Moreover, in human behavioral experiments in which punishment is uncoordinated, the sum of costs to punishers and their targets often exceeds the benefits of the increased cooperation that results from the punishment of free-riders. As a result, cooperation sustained by punishment may actually reduce the average payoffs of group members in comparison with groups in which punishment of free-riders is not an option. Here, we present a model of coordinated punishment that is calibrated for ancestral human conditions and captures a further aspect of reality missing from both models and experiments: The total cost of punishing a free-rider declines as the number of punishers increases. We show that punishment can proliferate when rare, and when it does, it enhances group-average payoffs.

Humans are a uniquely cooperative species. In even the simplest societies, people cooperate in large groups of genealogically distant individuals (*1*–*3*). In the laboratory, subjects routinely cooperate in situations in which selfish agents would free-ride on the cooperation of others (*4*, *5*). Recent theoretical studies provide an evolutionary explanation for such cooperative behavior: Punishment reduces gain to free-riding, so groups with more punishers can sustain more cooperation (*6*–*9*). Punishment is costly, but unlike unconditional altruism its costs are greatly reduced when punishers are common because punishment then occurs at very low frequency, is effective, and its costs can be shared. As a result, a modest advantage of groups in which cooperation is sustained by the presence of punishers is sufficient to compensate them for the cost of punishment.

There are two important problems with this explanation of human cooperation. First, punishment can reduce the average payoffs of group members because the costs of punishment may exceed the gains from cooperation (*5*). This problem is exacerbated when punishers target cooperative group members, as sometimes occurs in experiments (*10*–*12*). Second, the initial emergence of punishment remains a puzzle. In order to survive, punishers must engage in enough punishment of defectors so that the induced cooperation more than offsets the cost of punishing. Rare punishers do not have the benefit of outnumbering their targets, so the cost of punishing a free-rider is substantial. Moreover, they usually bear this cost alone rather than sharing it with other punishers (*13*–*16*).

These problems are an artifact of the unrealistic way that punishment is implemented in existing models and in most experiments. In these models, punishment is an unconditional and uncoordinated individual action automatically triggered by defection. Similarly and with few exceptions (*17*), in experiments individuals cannot coordinate their punishment. In contrast, ethnographic evidence indicates that punishment is coordinated by means of gossip and other communication among punishers, is contingent on the expected effectiveness of punishment in inducing cooperation, and is not undertaken unless it is judged as legitimate by most group members (*18*–*20*). When it occurs, punishment is usually collective and conveys a message of peer condemnation. Consistent with the anthropological evidence, in behavioral experiments with communication or with the option of coordinating behavior punishment is often highly effective in raising group average payoffs (*21*).

We analyzed a model of the evolution of punishment that incorporates two empirically based features absent from previous work. First, punishment is coordinated among group members so that it is contingent on the number of others predisposed to participate in the punishment. This means that when individuals willing to punish are rare, they demur and so bear only the cost of signaling their willingness to punish. They thus avoid the cost of punishing when it

[1]Department of Anthropology, University of California, Los Angeles, CA 90064, USA. [2]Behavioral Sciences Program, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA. [3]Department of Economics, Central European University, Nádor utca 9, 1051 Budapest, Hungary. [4]Collegium Budapest, Szentharomsag utca 2, 1014 Budapest, Hungary. [5]Department of Economics, University of Siena, Piazza San Francesco 7, Siena 53100, Italy.

*To whom correspondence should be addressed. E-mail: rboyd@anthro.ucla.edu (R.B.); hgintis@comcast.net (H.G.); samuel.bowles@gmail.com (S.B.)