



International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report

IR-07-054

**Punish or Perish? Retaliation and Collaboration
Among Humans**

Karl Sigmund (karl.sigmund@univie.ac.at)

Approved by

Ulf Dieckmann
Leader, Evolution and Ecology Program

December 2007

Interim Reports on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.



The Evolution and Ecology Program at IIASA fosters the development of new mathematical and conceptual techniques for understanding the evolution of complex adaptive systems.

Focusing on these long-term implications of adaptive processes in systems of limited growth, the Evolution and Ecology Program brings together scientists and institutions from around the world with IIASA acting as the central node.

Scientific progress within the network is collected in the IIASA Studies in Adaptive Dynamics series.

- No. 1 Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS: *Adaptive Dynamics: A Geometrical Study of the Consequences of Nearly Faithful Reproduction*. IIASA Working Paper WP-95-099 (1995). van Strien SJ, Verduyn Lunel SM (eds): *Stochastic and Spatial Structures of Dynamical Systems*, Proceedings of the Royal Dutch Academy of Science (KNAW Verhandelingen), North Holland, Amsterdam, pp. 183-231 (1996).
- No. 2 Dieckmann U, Law R: *The Dynamical Theory of Coevolution: A Derivation from Stochastic Ecological Processes*. IIASA Working Paper WP-96-001 (1996). *Journal of Mathematical Biology* 34:579-612 (1996).
- No. 3 Dieckmann U, Marrow P, Law R: *Evolutionary Cycling of Predator-Prey Interactions: Population Dynamics and the Red Queen*. IIASA Preprint (1995). *Journal of Theoretical Biology* 176:91-102 (1995).
- No. 4 Marrow P, Dieckmann U, Law R: *Evolutionary Dynamics of Predator-Prey Systems: An Ecological Perspective*. IIASA Working Paper WP-96-002 (1996). *Journal of Mathematical Biology* 34:556-578 (1996).
- No. 5 Law R, Marrow P, Dieckmann U: *On Evolution under Asymmetric Competition*. IIASA Working Paper WP-96-003 (1996). *Evolutionary Ecology* 11:485-501 (1997).
- No. 6 Metz JAJ, Mylius SD, Diekmann O: *When Does Evolution Optimize? On the Relation Between Types of Density Dependence and Evolutionarily Stable Life History Parameters*. IIASA Working Paper WP-96-004 (1996).
- No. 7 Ferrière R, Gatto M: *Lyapunov Exponents and the Mathematics of Invasion in Oscillatory or Chaotic Populations*. *Theoretical Population Biology* 48:126-171 (1995).
- No. 8 Ferrière R, Fox GA: *Chaos and Evolution*. IIASA Preprint (1996). *Trends in Ecology and Evolution* 10:480-485 (1995).
- No. 9 Ferrière R, Michod RE: *The Evolution of Cooperation in Spatially Heterogeneous Populations*. IIASA Working Paper WP-96-029 (1996). *The American Naturalist* 147:692-717 (1996).
- No. 10 van Dooren TJM, Metz JAJ: *Delayed Maturation in Temporally Structured Populations with Non-Equilibrium Dynamics*. IIASA Working Paper WP-96-070 (1996). *Journal of Evolutionary Biology* 11:41-62 (1998).
- No. 11 Geritz SAH, Metz JAJ, Kisdi É, Meszéna G: *The Dynamics of Adaptation and Evolutionary Branching*. IIASA Working Paper WP-96-077 (1996). *Physical Review Letters* 78:2024-2027 (1997).
- No. 12 Geritz SAH, Kisdi É, Meszéna G, Metz JAJ: *Evolutionary Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree*. IIASA Working Paper WP-96-114 (1996). *Evolutionary Ecology* 12:35-57 (1998).
- No. 13 Heino M, Metz JAJ, Kaitala V: *Evolution of Mixed Maturation Strategies in Semelparous Life-Histories: The Crucial Role of Dimensionality of Feedback Environment*. IIASA Working Paper WP-96-126 (1996). *Philosophical Transactions of the Royal Society of London Series B* 352:1647-1655 (1997).
- No. 14 Dieckmann U: *Can Adaptive Dynamics Invade?* IIASA Working Paper WP-96-152 (1996). *Trends in Ecology and Evolution* 12:128-131 (1997).
- No. 15 Meszéna G, Czibula I, Geritz SAH: *Adaptive Dynamics in a 2-Patch Environment: A Simple Model for Allopatric and Parapatric Speciation*. IIASA Interim Report IR-97-001 (1997). *Journal of Biological Systems* 5:265-284 (1997).
- No. 16 Heino M, Metz JAJ, Kaitala V: *The Enigma of Frequency-Dependent Selection*. IIASA Interim Report IR-97-061 (1997). *Trends in Ecology and Evolution* 13:367-370 (1998).
- No. 17 Heino M: *Management of Evolving Fish Stocks*. IIASA Interim Report IR-97-062 (1997). *Canadian Journal of Fisheries and Aquatic Sciences* 55:1971-1982 (1998).
- No. 18 Heino M: *Evolution of Mixed Reproductive Strategies in Simple Life-History Models*. IIASA Interim Report IR-97-063 (1997).
- No. 19 Geritz SAH, van der Meijden E, Metz JAJ: *Evolutionary Dynamics of Seed Size and Seedling Competitive Ability*. IIASA Interim Report IR-97-071 (1997). *Theoretical Population Biology* 55:324-343 (1999).
- No. 20 Galis F, Metz JAJ: *Why Are There So Many Cichlid Species? On the Interplay of Speciation and Adaptive Radiation*. IIASA Interim Report IR-97-072 (1997). *Trends in Ecology and Evolution* 13:1-2 (1998).

- No. 21 Boerlijst MC, Nowak MA, Sigmund K: *Equal Pay for all Prisoners/ The Logic of Contrition*. IIASA Interim Report IR-97-073 (1997). American Mathematical Society Monthly 104:303-307 (1997). Journal of Theoretical Biology 185:281-293 (1997).
- No. 22 Law R, Dieckmann U: *Symbiosis Without Mutualism and the Merger of Lineages in Evolution*. IIASA Interim Report IR-97-074 (1997). Proceedings of the Royal Society of London Series B 265:1245-1253 (1998).
- No. 23 Klinkhamer PGL, de Jong TJ, Metz JAJ: *Sex and Size in Cosexual Plants*. IIASA Interim Report IR-97-078 (1997). Trends in Ecology and Evolution 12:260-265 (1997).
- No. 24 Fontana W, Schuster P: *Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping*. IIASA Interim Report IR-98-004 (1998). Journal of Theoretical Biology 194:491-515 (1998).
- No. 25 Kisdi É, Geritz SAH: *Adaptive Dynamics in Allele Space: Evolution of Genetic Polymorphism by Small Mutations in a Heterogeneous Environment*. IIASA Interim Report IR-98-038 (1998). Evolution 53:993-1008 (1999).
- No. 26 Fontana W, Schuster P: *Continuity in Evolution: On the Nature of Transitions*. IIASA Interim Report IR-98-039 (1998). Science 280:1451-1455 (1998).
- No. 27 Nowak MA, Sigmund K: *Evolution of Indirect Reciprocity by Image Scoring/ The Dynamics of Indirect Reciprocity*. IIASA Interim Report IR-98-040 (1998). Nature 393:573-577 (1998). Journal of Theoretical Biology 194:561-574 (1998).
- No. 28 Kisdi É: *Evolutionary Branching Under Asymmetric Competition*. IIASA Interim Report IR-98-045 (1998). Journal of Theoretical Biology 197:149-162 (1999).
- No. 29 Berger U: *Best Response Adaptation for Role Games*. IIASA Interim Report IR-98-086 (1998).
- No. 30 van Dooren TJM: *The Evolutionary Ecology of Dominance-Recessivity*. IIASA Interim Report IR-98-096 (1998). Journal of Theoretical Biology 198:519-532 (1999).
- No. 31 Dieckmann U, O'Hara B, Weisser W: *The Evolutionary Ecology of Dispersal*. IIASA Interim Report IR-98-108 (1998). Trends in Ecology and Evolution 14:88-90 (1999).
- No. 32 Sigmund K: *Complex Adaptive Systems and the Evolution of Reciprocation*. IIASA Interim Report IR-98-100 (1998). Ecosystems 1:444-448 (1998).
- No. 33 Posch M, Pichler A, Sigmund K: *The Efficiency of Adapting Aspiration Levels*. IIASA Interim Report IR-98-103 (1998). Proceedings of the Royal Society London Series B 266:1427-1435 (1999).
- No. 34 Mathias A, Kisdi É: *Evolutionary Branching and Coexistence of Germination Strategies*. IIASA Interim Report IR-99-014 (1999).
- No. 35 Dieckmann U, Doebeli M: *On the Origin of Species by Sympatric Speciation*. IIASA Interim Report IR-99-013 (1999). Nature 400:354-357 (1999).
- No. 36 Metz JAJ, Gyllenberg M: *How Should We Define Fitness in Structured Metapopulation Models? Including an Application to the Calculation of Evolutionarily Stable Dispersal Strategies*. IIASA Interim Report IR-99-019 (1999). Proceedings of the Royal Society of London Series B 268:499-508 (2001).
- No. 37 Gyllenberg M, Metz JAJ: *On Fitness in Structured Metapopulations*. IIASA Interim Report IR-99-037 (1999). Journal of Mathematical Biology 43:545-560 (2001).
- No. 38 Meszéna G, Metz JAJ: *Species Diversity and Population Regulation: The Importance of Environmental Feedback Dimensionality*. IIASA Interim Report IR-99-045 (1999).
- No. 39 Kisdi É, Geritz SAH: *Evolutionary Branching and Sympatric Speciation in Diploid Populations*. IIASA Interim Report IR-99-048 (1999).
- No. 40 Ylikarjula J, Heino M, Dieckmann U: *Ecology and Adaptation of Stunted Growth in Fish*. IIASA Interim Report IR-99-050 (1999). Evolutionary Ecology 13:433-453 (1999).
- No. 41 Nowak MA, Sigmund K: *Games on Grids*. IIASA Interim Report IR-99-038 (1999). Dieckmann U, Law R, Metz JAJ (eds): The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, Cambridge, UK, pp. 135-150 (2000).
- No. 42 Ferrière R, Michod RE: *Wave Patterns in Spatial Games and the Evolution of Cooperation*. IIASA Interim Report IR-99-041 (1999). Dieckmann U, Law R, Metz JAJ (eds): The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, Cambridge, UK, pp. 318-332 (2000).
- No. 43 Kisdi É, Jacobs FJA, Geritz SAH: *Red Queen Evolution by Cycles of Evolutionary Branching and Extinction*. IIASA Interim Report IR-00-030 (2000). Selection 2:161-176 (2001).
- No. 44 Meszéna G, Kisdi É, Dieckmann U, Geritz SAH, Metz JAJ: *Evolutionary Optimisation Models and Matrix Games in the Unified Perspective of Adaptive Dynamics*. IIASA Interim Report IR-00-039 (2000). Selection 2:193-210 (2001).
- No. 45 Parvinen K, Dieckmann U, Gyllenberg M, Metz JAJ: *Evolution of Dispersal in Metapopulations with Local Density Dependence and Demographic Stochasticity*. IIASA Interim Report IR-00-035 (2000). Journal of Evolutionary Biology 16:143-153 (2003).
- No. 46 Doebeli M, Dieckmann U: *Evolutionary Branching and Sympatric Speciation Caused by Different Types of Ecological Interactions*. IIASA Interim Report IR-00-040 (2000). The American Naturalist 156:S77-S101 (2000).
- No. 47 Heino M, Hanski I: *Evolution of Migration Rate in a Spatially Realistic Metapopulation Model*. IIASA Interim Report IR-00-044 (2000). The American Naturalist 157:495-511 (2001).
- No. 48 Gyllenberg M, Parvinen K, Dieckmann U: *Evolutionary Suicide and Evolution of Dispersal in Structured Metapopulations*. IIASA Interim Report IR-00-056 (2000). Journal of Mathematical Biology 45:79-105 (2002).
- No. 49 van Dooren TJM: *The Evolutionary Dynamics of Direct Phenotypic Overdominance: Emergence Possible, Loss Probable*. IIASA Interim Report IR-00-048 (2000). Evolution 54:1899-1914 (2000).
- No. 50 Nowak MA, Page KM, Sigmund K: *Fairness Versus Reason in the Ultimatum Game*. IIASA Interim Report IR-00-57 (2000). Science 289:1773-1775 (2000).
- No. 51 de Feo O, Ferrière R: *Bifurcation Analysis of Population Invasion: On-Off Intermittency and Basin Riddling*. IIASA Interim Report IR-00-074 (2000). International Journal of Bifurcation and Chaos 10:443-452 (2000).

- No. 52 Heino M, Laaka-Lindberg S: *Clonal Dynamics and Evolution of Dormancy in the Leafy Hepatic Lophozia Silvicola*. IIASA Interim Report IR-01-018 (2001). *Oikos* 94:525-532 (2001).
- No. 53 Sigmund K, Hauert C, Nowak MA: *Reward and Punishment in Minigames*. IIASA Interim Report IR-01-031 (2001). *Proceedings of the National Academy of Sciences of the USA* 98:10757-10762 (2001).
- No. 54 Hauert C, De Monte S, Sigmund K, Hofbauer J: *Oscillations in Optional Public Good Games*. IIASA Interim Report IR-01-036 (2001).
- No. 55 Ferrière R, Le Galliard J: *Invasion Fitness and Adaptive Dynamics in Spatial Population Models*. IIASA Interim Report IR-01-043 (2001). Clobert J, Dhondt A, Danchin E, Nichols J (eds): *Dispersal*, Oxford University Press, pp. 57-79 (2001).
- No. 56 de Mazancourt C, Loreau M, Dieckmann U: *Can the Evolution of Plant Defense Lead to Plant-Herbivore Mutualism?* IIASA Interim Report IR-01-053 (2001). *The American Naturalist* 158:109-123 (2001).
- No. 57 Claessen D, Dieckmann U: *Ontogenetic Niche Shifts and Evolutionary Branching in Size-Structured Populations*. IIASA Interim Report IR-01-056 (2001). *Evolutionary Ecology Research* 4:189-217 (2002).
- No. 58 Brandt H: *Correlation Analysis of Fitness Landscapes*. IIASA Interim Report IR-01-058 (2001).
- No. 59 Dieckmann U: *Adaptive Dynamics of Pathogen-Host Interactions*. IIASA Interim Report IR-02-007 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, Cambridge University Press, Cambridge, UK, pp. 39-59 (2002).
- No. 60 Nowak MA, Sigmund K: *Super- and Coinfection: The Two Extremes*. IIASA Interim Report IR-02-008 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, Cambridge University Press, Cambridge, UK, pp. 124-137 (2002).
- No. 61 Sabelis MW, Metz JAJ: *Evolution Management: Taking Stock - Relating Theory to Experiment*. IIASA Interim Report IR-02-009 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, Cambridge University Press, Cambridge, UK, pp. 379-398 (2002).
- No. 62 Cheptou P, Dieckmann U: *The Evolution of Self-Fertilization in Density-Regulated Populations*. IIASA Interim Report IR-02-024 (2002). *Proceedings of the Royal Society of London Series B* 269:1177-1186 (2002).
- No. 63 Bürger R: *Additive Genetic Variation Under Intraspecific Competition and Stabilizing Selection: A Two-Locus Study*. IIASA Interim Report IR-02-013 (2002). *Theoretical Population Biology* 61:197-213 (2002).
- No. 64 Hauert C, De Monte S, Hofbauer J, Sigmund K: *Volunteering as Red Queen Mechanism for Co-operation in Public Goods Games*. IIASA Interim Report IR-02-041 (2002). *Science* 296:1129-1132 (2002).
- No. 65 Dercole F, Ferrière R, Rinaldi S: *Ecological Bistability and Evolutionary Reversals under Asymmetrical Competition*. IIASA Interim Report IR-02-053 (2002). *Evolution* 56:1081-1090 (2002).
- No. 66 Dercole F, Rinaldi S: *Evolution of Cannibalistic Traits: Scenarios Derived from Adaptive Dynamics*. IIASA Interim Report IR-02-054 (2002). *Theoretical Population Biology* 62:365-374 (2002).
- No. 67 Bürger R, Gimelfarb A: *Fluctuating Environments and the Role of Mutation in Maintaining Quantitative Genetic Variation*. IIASA Interim Report IR-02-058 (2002). *Genetical Research* 80:31-46 (2002).
- No. 68 Bürger R: *On a Genetic Model of Intraspecific Competition and Stabilizing Selection*. IIASA Interim Report IR-02-062 (2002). *Amer. Natur.* 160:661-682 (2002).
- No. 69 Doebeli M, Dieckmann U: *Speciation Along Environmental Gradients*. IIASA Interim Report IR-02-079 (2002). *Nature* 421:259-264 (2003).
- No. 70 Dercole F, Irisson J, Rinaldi S: *Bifurcation Analysis of a Prey-Predator Coevolution Model*. IIASA Interim Report IR-02-078 (2002). *SIAM Journal on Applied Mathematics* 63:1378-1391 (2003).
- No. 71 Le Galliard J, Ferrière R, Dieckmann U: *The Adaptive Dynamics of Altruism in Spatially Heterogeneous Populations*. IIASA Interim Report IR-03-006 (2003). *Evolution* 57:1-17 (2003).
- No. 72 Taborsky B, Dieckmann U, Heino M: *Unexpected Discontinuities in Life-History Evolution under Size-Dependent Mortality*. IIASA Interim Report IR-03-004 (2003). *Proceedings of the Royal Society of London Series B* 270:713-721 (2003).
- No. 73 Gardmark A, Dieckmann U, Lundberg P: *Life-History Evolution in Harvested Populations: The Role of Natural Predation*. IIASA Interim Report IR-03-008 (2003). *Evolutionary Ecology Research* 5:239-257 (2003).
- No. 74 Mizera F, Meszéna G: *Spatial Niche Packing, Character Displacement and Adaptive Speciation Along an Environmental Gradient*. IIASA Interim Report IR-03-062 (2003). *Evolutionary Ecology Research* 5:363-382 (2003).
- No. 75 Dercole F: *Remarks on Branching-Extinction Evolutionary Cycles*. IIASA Interim Report IR-03-077 (2003). *Journal of Mathematical Biology* 47:569-580 (2003).
- No. 76 Hofbauer J, Sigmund K: *Evolutionary Game Dynamics*. IIASA Interim Report IR-03-078 (2003). *Bulletin of the American Mathematical Society* 40:479-519 (2003).
- No. 77 Ernande B, Dieckmann U, Heino M: *Adaptive Changes in Harvested Populations: Plasticity and Evolution of Age and Size at Maturation*. IIASA Interim Report IR-03-058 (2003). *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:415-423 (2004).
- No. 78 Hanski I, Heino M: *Metapopulation-Level Adaptation of Insect Host Plant Preference and Extinction-Colonization Dynamics in Heterogeneous Landscapes*. IIASA Interim Report IR-03-028 (2003). *Theoretical Population Biology* 63:309-338 (2003).
- No. 79 van Doorn G, Dieckmann U, Weissing FJ: *Sympatric Speciation by Sexual Selection: A Critical Re-Evaluation*. IIASA Interim Report IR-04-003 (2004). *American Naturalist* 163:709-725 (2004).
- No. 80 Egas M, Dieckmann U, Sabelis MW: *Evolution Restricts the Coexistence of Specialists and Generalists - the Role of Trade-off Structure*. IIASA Interim Report IR-04-004 (2004). *American Naturalist* 163:518-531 (2004).

- No. 81 Ernande B, Dieckmann U: *The Evolution of Phenotypic Plasticity in Spatially Structured Environments: Implications of Intraspecific Competition, Plasticity Costs, and Environmental Characteristics*. IIASA Interim Report IR-04-006 (2004). *Journal of Evolutionary Biology* 17:613-628 (2004).
- No. 82 Cressman R, Hofbauer J: *Measure Dynamics on a One-Dimensional Continuous Trait Space: Theoretical Foundations for Adaptive Dynamics*. IIASA Interim Report IR-04-016 (2004).
- No. 83 Cressman R: *Dynamic Stability of the Replicator Equation with Continuous Strategy Space*. IIASA Interim Report IR-04-017 (2004).
- No. 84 Ravigné V, Olivieri I, Dieckmann U: *Implications of Habitat Choice for Protected Polymorphisms*. IIASA Interim Report IR-04-005 (2004). *Evolutionary Ecology Research* 6:125-145 (2004).
- No. 85 Nowak MA, Sigmund K: *Evolutionary Dynamics of Biological Games*. IIASA Interim Report IR-04-013 (2004). *Science* 303:793-799 (2004).
- No. 86 Vukics A, Asbóth J, Meszéna G: *Speciation in Multidimensional Evolutionary Space*. IIASA Interim Report IR-04-028 (2004). *Physical Review* 68:041-903 (2003).
- No. 87 de Mazancourt C, Dieckmann U: *Trade-off Geometries and Frequency-dependent Selection*. IIASA Interim Report IR-04-039 (2004). *American Naturalist* 164:765-778 (2004).
- No. 88 Cadet CR, Metz JAJ, Klinkhamer PGL: *Size and the Not-So-Single Sex: Disentangling the Effects of Size on Sex Allocation*. IIASA Interim Report IR-04-084 (2004). *American Naturalist* 164:779-792 (2004).
- No. 89 Rueffler C, van Dooren TJM, Metz JAJ: *Adaptive Walks on Changing Landscapes: Levins' Approach Extended*. IIASA Interim Report IR-04-083 (2004). *Theoretical Population Biology* 65:165-178 (2004).
- No. 90 de Mazancourt C, Loreau M, Dieckmann U: *Understanding Mutualism When There is Adaptation to the Partner*. IIASA Interim Report IR-05-016 (2005). *Journal of Ecology* 93:305-314 (2005).
- No. 91 Dieckmann U, Doebeli M: *Pluralism in Evolutionary Theory*. IIASA Interim Report IR-05-017 (2005). *Journal of Evolutionary Biology* 18:1209-1213 (2005).
- No. 92 Doebeli M, Dieckmann U, Metz JAJ, Tautz D: *What We Have Also Learned: Adaptive Speciation is Theoretically Plausible*. IIASA Interim Report IR-05-018 (2005). *Evolution* 59:691-695 (2005).
- No. 93 Egas M, Sabelis MW, Dieckmann U: *Evolution of Specialization and Ecological Character Displacement of Herbivores Along a Gradient of Plant Quality*. IIASA Interim Report IR-05-019 (2005). *Evolution* 59:507-520 (2005).
- No. 94 Le Galliard J, Ferrière R, Dieckmann U: *Adaptive Evolution of Social Traits: Origin, Trajectories, and Correlations of Altruism and Mobility*. IIASA Interim Report IR-05-020 (2005). *American Naturalist* 165:206-224 (2005).
- No. 95 Doebeli M, Dieckmann U: *Adaptive Dynamics as a Mathematical Tool for Studying the Ecology of Speciation Processes*. IIASA Interim Report IR-05-022 (2005). *Journal of Evolutionary Biology* 18:1194-1200 (2005).
- No. 96 Brandt H, Sigmund K: *The Logic of Reprobation: Assessment and Action Rules for Indirect Reciprocity*. IIASA Interim Report IR-04-085 (2004). *Journal of Theoretical Biology* 231:475-486 (2004).
- No. 97 Hauert C, Haiden N, Sigmund K: *The Dynamics of Public Goods*. IIASA Interim Report IR-04-086 (2004). *Discrete and Continuous Dynamical Systems - Series B* 4:575-587 (2004).
- No. 98 Meszéna G, Gyllenberg M, Jacobs FJA, Metz JAJ: *Link Between Population Dynamics and Dynamics of Darwinian Evolution*. IIASA Interim Report IR-05-026 (2005). *Physical Review Letters* 95:Article 078105 (2005).
- No. 99 Meszéna G: *Adaptive Dynamics: The Continuity Argument*. IIASA Interim Report IR-05-032 (2005).
- No. 100 Brännström NA, Dieckmann U: *Evolutionary Dynamics of Altruism and Cheating Among Social Amoebas*. IIASA Interim Report IR-05-039 (2005). *Proceedings of the Royal Society London Series B* 272:1609-1616 (2005).
- No. 101 Meszéna G, Gyllenberg M, Pasztor L, Metz JAJ: *Competitive Exclusion and Limiting Similarity: A Unified Theory*. IIASA Interim Report IR-05-040 (2005).
- No. 102 Szabo P, Meszéna G: *Limiting Similarity Revisited*. IIASA Interim Report IR-05-050 (2005).
- No. 103 Krakauer DC, Sasaki A: *The Greater than Two-Fold Cost of Integration for Retroviruses*. IIASA Interim Report IR-05-069 (2005).
- No. 104 Metz JAJ: *Eight Personal Rules for Doing Science*. IIASA Interim Report IR-05-073 (2005). *Journal of Evolutionary Biology* 18:1178-1181 (2005).
- No. 105 Beltman JB, Metz JAJ: *Speciation: More Likely Through a Genetic or Through a Learned Habitat Preference?* IIASA Interim Report IR-05-072 (2005). *Proceedings of the Royal Society of London Series B* 272:1455-1463 (2005).
- No. 106 Durinx M, Metz JAJ: *Multi-type Branching Processes and Adaptive Dynamics of Structured Populations*. IIASA Interim Report IR-05-074 (2005). Haccou P, Jager P, Vatutin V (eds): *Branching Processes: Variation, Growth and Extinction of Populations*, Cambridge University Press, Cambridge, UK, pp. 266-278 (2005).
- No. 107 Brandt H, Sigmund K: *The Good, the Bad and the Discriminator - Errors in Direct and Indirect Reciprocity*. IIASA Interim Report IR-05-070 (2005). *Journal of Theoretical Biology* 239:183-194 (2006).
- No. 108 Brandt H, Hauert C, Sigmund K: *Punishing and Abstaining for Public Goods*. IIASA Interim Report IR-05-071 (2005). *Proceedings of the National Academy of Sciences of the United States of America* 103:495-497 (2006).
- No. 109 Ohtsuki A, Sasaki A: *Epidemiology and Disease-Control Under Gene-for-Gene Plant-Pathogen Interaction*. IIASA Interim Report IR-05-068 (2005).
- No. 110 Brandt H, Sigmund K: *Indirect Reciprocity, Image-Scoring, and Moral Hazard*. IIASA Interim Report IR-05-078 (2005). *Proceedings of the National Academy of Sciences of the United States of America* 102:2666-2670 (2005).
- No. 111 Nowak MA, Sigmund K: *Evolution of Indirect Reciprocity*. IIASA Interim Report IR-05-079 (2005). *Nature* 437:1292-1298 (2005).
- No. 112 Kamo M, Sasaki A: *Evolution Towards Multi-Year Periodicity in Epidemics*. IIASA Interim Report IR-05-080 (2005). *Ecology Letters* 8:378-385 (2005).

- No. 113 Dercole F, Ferrière R, Gragnani A, Rinaldi S: *Co-evolution of Slow-fast Populations: Evolutionary Sliding, Evolutionary Pseudo-equilibria, and Complex Red Queen Dynamics*. IIASA Interim Report IR-06-006 (2006). *Proceedings of the Royal Society B* 273:983-990 (2006).
- No. 114 Dercole F: *Border Collision Bifurcations in the Evolution of Mutualistic Interactions*. IIASA Interim Report IR-05-083 (2005). *International Journal of Bifurcation and Chaos* 15:2179-2190 (2005).
- No. 115 Dieckmann U, Heino M, Parvinen K: *The Adaptive Dynamics of Function-Valued Traits*. IIASA Interim Report IR-06-036 (2006). *Journal of Theoretical Biology* 241:370-389 (2006).
- No. 116 Dieckmann U, Metz JAJ: *Surprising Evolutionary Predictions from Enhanced Ecological Realism*. IIASA Interim Report IR-06-037 (2006). *Theoretical Population Biology* 69:263-281 (2006).
- No. 117 Dieckmann U, Brännström NA, HilleRisLambers R, Ito H: *The Adaptive Dynamics of Community Structure*. IIASA Interim Report IR-06-038 (2006). Takeuchi Y, Iwasa Y, Sato K (eds): *Mathematics for Ecology and Environmental Sciences*, Springer, Berlin Heidelberg, pp. 145-177 (2007).
- No. 118 Gardmark A, Dieckmann U: *Disparate Maturation Adaptations to Size-dependent Mortality*. IIASA Interim Report IR-06-039 (2006). *Proceedings of the Royal Society London Series B* 273:2185-2192 (2006).
- No. 119 van Doorn G, Dieckmann U: *The Long-term Evolution of Multi-locus Traits Under Frequency-dependent Disruptive Selection*. IIASA Interim Report IR-06-041 (2006). *Evolution* 60:2226-2238 (2006).
- No. 120 Doebeli M, Blok HJ, Leimar O, Dieckmann U: *Multimodal Pattern Formation in Phenotype Distributions of Sexual Populations*. IIASA Interim Report IR-06-046 (2006). *Proceedings of the Royal Society London Series B* 274:347-357 (2007).
- No. 121 Dunlop ES, Shuter BJ, Dieckmann U: *The Demographic and Evolutionary Consequences of Selective Mortality: Predictions from an Eco-genetic Model of the Smallmouth Bass*. IIASA Interim Report IR-06-060 (2006). *Transactions of the American Fisheries Society* 136:749-765 (2007).
- No. 122 Metz JAJ: *Fitness*. IIASA Interim Report IR-06-061 (2006).
- No. 123 Brandt H, Ohtsuki H, Iwasa Y, Sigmund K: *A Survey on Indirect Reciprocity*. IIASA Interim Report IR-06-065 (2006). Takeuchi Y, Iwasa Y, Sato K (eds): *Mathematics for Ecology and Environmental Sciences*, Springer, Berlin Heidelberg, pp. 21-51 (2007).
- No. 124 Dercole F, Loiacono D, Rinaldi S: *Synchronization in Ecological Networks: A Byproduct of Darwinian Evolution?* IIASA Interim Report IR-06-068 (2006). *International Journal of Bifurcation and Chaos* 7:2435-2446 (2007).
- No. 125 Dercole F, Dieckmann U, Obersteiner M, Rinaldi S: *Adaptive Dynamics and Technological Change*. IIASA Interim Report IR-06-070 (2006).
- No. 126 Rueffler C, van Dooren TJM, Metz JAJ: *The Evolution of Resource Specialization Through Frequency-Dependent and Frequency-Independent Mechanisms*. IIASA Interim Report IR-06-073 (2006). *American Naturalist* 167:81-93 (2006).
- No. 127 Rueffler C, Egas M, Metz JAJ: *Evolutionary Predictions Should be Based on Individual Traits*. IIASA Interim Report IR-06-074 (2006). *American Naturalist* 168:148-162 (2006).
- No. 128 Kamo M, Sasaki A, Boots M: *The Role of Trade-Off Shapes in the Evolution of Virulence in Spatial Host-Parasite Interactions: An Approximate Analytical Approach*. IIASA Interim Report IR-06-075 (2006).
- No. 129 Boots M, Kamo M, Sasaki A: *The Implications of Spatial Structure Within Populations to the Evolution of Parasites*. IIASA Interim Report IR-06-078 (2006).
- No. 130 Andreasen V, Sasaki A: *Shaping the Phylogenetic Tree of Influenza by Cross-Immunity*. IIASA Interim Report IR-06-079 (2006).
- No. 131 Rueffler C, van Dooren TJM, Metz JAJ: *The Interplay Between Behavior and Morphology in the Evolutionary Dynamics of Resource Specialization*. IIASA Interim Report IR-06-082 (2006). *American Naturalist* 169:E34-E52 (2007).
- No. 132 Rueffler C, van Dooren TJM, Metz JAJ: *The Evolution of Simple Life-Histories: Steps Towards a Classification*. IIASA Interim Report IR-06-083 (2006).
- No. 133 Durinx M, Metz JAJ, Meszéna G: *Adaptive Dynamics for Physiologically Structured Population Models*. IIASA Interim Report IR-07-027 (2007).
- No. 134 Ito H, Dieckmann U: *A New Mechanism for Recurrent Adaptive Radiations*. IIASA Interim Report IR-07-048 (2007). *American Naturalist* 170:E96-E111 (2007).
- No. 135 Troost T, Kooi B, Dieckmann U: *Joint evolution of predator body size and prey-size preference*. IIASA Interim Report IR-07-050 (2007).
- No. 136 Nowak MA, Sigmund K: *How Populations Cohere: Five Rules for Cooperation*. IIASA Interim Report IR-07-052 (2007). May RM, McLean A (eds): *Theoretical Ecology: Principles and Applications*, Oxford UP, Oxford, pp. 7-16 (2007).
- No. 137 Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K: *The Emergence of Altruistic Punishment: Via Freedom to Enforcement*. IIASA Interim Report IR-07-053 (2007). *Science* 613:1905-1907 (2007).
- No. 138 Sigmund K: *Punish or Perish? Retaliation and Collaboration Among Humans*. IIASA Interim Report IR-07-054 (2007). *Trends in Ecology and Evolution* 22:593-600 (2007).

Issues of the IIASA Studies in Adaptive Dynamics series can be obtained at www.iiasa.ac.at/Research/EEP/Series.html or by writing to eeep@iiasa.ac.at.

Contents

Abstract.....	1
Introduction: Reciprocation and social enforcement.....	1
Fining free-riders	1
Sanctions and social dilemmas	2
Ultimate reasons of costly punishment.....	3
Proximate causes of costly punishment.....	5
The limitations of peer-punishment.....	7
Acknowledgement.....	7
References	8
Box 1	12
Box 2	13
Box 3	14
Box 4	15
Figure captions	16
Figures	17

Punish or perish? Retaliation and collaboration among humans

Karl Sigmund

Faculty for Mathematics, University of Vienna, Nordbergstrasse 15, 1090 Vienna, Austria
International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria
Corresponding author: Sigmund, K (karl.sigmund@univie.ac.at)

A spate of recent investigations on reciprocation and social enforcement in humans has brought together (and sometimes divided) economists, psychologists, anthropologists, social scientists and evolutionary biologists, as well as neurologists and students of animal behaviour. Experimental work on public goods and social incentives has addressed a wealth of questions on the emotional and cognitive (proximal) factors, as well as on the genetic and cultural (ultimate) evolutionary mechanisms involved in this essential aspect of human nature. This article surveys recent work, highlighting the role of punishment and reward in joint enterprises.

INTRODUCTION: RECIPROCATION AND SOCIAL ENFORCEMENT

How do humans manage to sustain collective efforts in sizable groups of unrelated individuals? The topic is in fashion, but not new. In 1975, for instance, W.D. Hamilton closed his essay on 'Innate Social Aptitudes of Man' [1] with a section on 'Reciprocation and Social Enforcement'. Humans have a special gift for reciprocation. However, in interactions involving more than two individuals, reciprocation works less well than in pair-wise encounters. Even defining it offers a non-trivial task. If your group contains a cooperator and a cheater, whom do you reciprocate with? Social enforcement, by contrast, works better in groups with more than two members, as Hamilton points out [1], and can offer 'at least a partial cure' for the problems with reciprocation in larger groups. 'There may be reason to be glad that human life is a many-person game and not just a disjointed collection of two-person games' [1].

Whereas pioneers of socio-biology were aware of the importance of public goods and punishment [1,2,3], recent cross-disciplinary contact between experimental economists and evolutionary biologists has greatly stimulated the field. Here, I review work focussing on incentives which promote cooperation in groups of unrelated humans.

FINING FREE-RIDERS

Let us begin with an experimental Public Good game. Six anonymous players are given \$10 each. They must decide whether to invest this in a common pool, knowing that the experimenter will triple the amount in the common pool, and distribute it equally among all six players, no matter whether they contributed.

This game is easy to analyse. If all players contribute, they triple their accounts. However, each player is better off by not contributing, because only half the contribution returns to the account it came from (i.e. it is multiplied by three and then divided among the six players). If, as a consequence, no player contributes, then the initial amounts remain unchanged. This deplorable outcome of selfish motives is known variously as social dilemma, tragedy of the commons, free-rider problem, market failure... The multiplicity of the names points to the ubiquity of the issue.

In modern society, the exploitation of collective efforts (e.g. free-riding on public buses, shirking tax, dodging military service...) is punished by a plethora of institutions. Obviously, the threat of punishment deters would-be defectors. This can be mimicked by another experiment on Public Goods, this time with Punishment. In this two-stage game, the first stage runs exactly as before. In the second stage, players can impose fines upon their co-players. These fines are collected by the experimenter and do not land in the punisher's account. In fact, each punisher must pay a fee for the experimenter to collect the fine.

Again, the analysis is easy. A player bent on maximizing income should not punish, because this is costly. Hence, nothing should happen in the second stage; thus, the first stage will be unaffected. No punishment, no contributions and no gains: the selfishly motivated inertia in both stages of the game leads to economic paralysis.

Gratifyingly, this does not happen in real experiments, which are usually slightly more sophisticated versions in which players can choose between different levels of contribution and sizes of fines. In seminal experiments by Fehr and Gächter [4][5], the average contribution of players, in the Public Good game without Punishment, was slightly >50% of their endowment. In the Public Good game with Punishment, it was higher – close to 60%. Punishment was usually targeted on defectors, and its mere threat had an immediate effect. However, the full size of this effect only shows when the game is repeated for several rounds (Figure 1). In the absence of punishment, contributions decrease; with punishment, they quickly increase to almost 100%. This happens if the groups stay together but, most significantly, even if the groups are newly formed between rounds, and players know that they will never meet a co-player twice. By inflicting punishment, they can conceivably turn a defector into a cooperator. However, punishers know that the future contributions of such a 'reformed' player will exclusively benefit others. Punishment appears as altruistic act.

This is a stunning outcome. Without sanctions, the public good, i.e. the tripling of the endowment, is not realised. With sanctions, it is, although selfish reckoning prescribes that costly punishment should not be delivered. In the absence of institutions, players are willing 'to take the law into their own hands' (also known as 'peer-punishing'). This enforces cooperation in many-player interactions between unrelated individuals, which is a remarkable trait of human societies, and surely an essential factor in our evolutionary history.

SANCTIONS AND SOCIAL DILEMMAS

The investigation of the interplay between mutual assistance and social enforcement is a booming enterprise. Economists use experimental games to study the effects of positive and negative incentives (i.e. reward and punishment) on our propensity to collaborate [6,7]; anthropologists visit small-scale societies to measure the culture-dependence and universality of norms that enforce cooperation [8]; psychologists study the often sub-conscious cues eliciting emotions that lead to helping behaviour or moralistic aggression [9-11]; neurologists use magnetic resonance techniques to correlate social dilemmas with brain activities [12,13]; game theorists modify their utility functions to take account of non-monetary concerns [14,15]; biologists look for signs of policing and sanctions in bees or bacteria [16,17]; and political scientists attempt to improve governance of institutions promoting collective actions [18,19]. Trans-disciplinary dialogues are in full swing, although communication sometimes needs improving [20].

The underlying questions concerning political creatures and human nature go back at least to Aristotle. The formal framework for discussing social dilemmas that arise with public goods

was provided by game theory [21]. As Olson stated in 1964 in his '*Logic of Collective Action*' [22], self-interested individuals will not act to achieve their group interest, except when prodded by incentives directed selectively towards individuals in the group, (i.e. punishing exploiters or rewarding contributors). In 1965, the biologist Hardin addressed the same issue in a highly influential *Science* paper on the Tragedy of the Commons [23], and offered as solution of the social dilemma: 'mutual coercion, mutually agreed upon'. This is also advocated in Hobbes' *Leviathan*. But how can the agreement be enforced? The role of sanctioning institutions, in our civilisation and in other societies, to uphold social norms and protect public goods has become an object of intense scrutiny [24].

Providing selective incentives for collective action is itself a public good, of course, so that the prevalence of punishing (or rewarding) institutions seems to present a chicken-and-egg problem. "The provision of a sanctioning system as a public good" was the title of an experimental paper by Yamagishi (1986) studying the effect of costly punishment on contributions towards a collective benefit [25]. In the 'nineties, a wide range of papers studied punishing and rewarding, usually in two-person games (Box 1).

The basic finding in all those economic experiments is that most humans are not self-centred, but are other-regarding [27,28]. Their aim is not uniquely the maximization of their income. They are strongly motivated by emotions ("moral sentiments", in the words of Adam Smith).

Witness for example the Ultimatum game (Box 1): a Responder uniquely interested in income maximization should accept any positive offer, even the smallest. A similarly disposed Proposer should, therefore, make the minimal offer, and keep the rest of the sum. But the real outcome is vastly different. Most offers are close to a fair split; the rare unfair offers are mostly rejected. Dozens of experiments verified the robustness of this outcome. In particular, in all 'modern' societies, some two-thirds of the offers are between 40% and 50% of the total sum; those <20% are few, and get usually rejected [26]. A vast collective effort of anthropological studies [8] was able to document cultural variation in small-scale societies: but even among the Machiguenga, an Amazonian population of hunter-gatherers, the mean offer was 26%; this record of unfairness is still a long way from the minimal offer predicted for selfish agents.

The Ultimatum game is played in a group of two players only and looks, at first glance, distinct from the Public Goods game with Punishment: the players are in different roles, and each one has only one decision to make. But the rejection of an offer by the Responder is a costly punishment (less costly to the punisher if the offer is small, and hurting the Proposer all the more).

The economic experiments displaying 'Human Nature' motivate both social psychologists and evolutionary biologists (as well as evolutionary psychologists and sociobiologists) to study the proximal or ultimate causes (i.e. the how and the why of human cooperation). This review can only give pointers to a literature drawing, in each field, on a rich tradition.

ULTIMATE REASONS OF COSTLY PUNISHMENT

Altruistic behaviour and selfish genes provide a favourite playground for theories on the evolution of cooperation, and have led to a rich tool-box (Box 2). Does this tool-box offer an explanation for our propensity to punish cheaters in public good interactions? How can the trait emerge, and how can it be maintained (Box 3)?

Two evolutionary approaches to these questions are based on group selection, and invoke selective group extinction [34,35]. It seems likely that intergroup conflict was frequent in early human history, and had a large impact on shaping human instincts. If a group is threatened with extinction, solidarity soars, as people can tell who have experienced bombing raids. Courage, comradeship, bonding and the readiness to risk, or sacrifice, one's own life for the group must have been shaped by such recurrent episodes from the past. It seems less compelling that this also moulded the behaviour observed in economic games, which is based on common concerns for fairness or reciprocation.

Several other models [36-38] exploit the fact that, if populations are not well-mixed, but interact only locally, benefits through punishment are easier to achieve. According to one view, which defines relatedness by statistical correlation rather than common descent, both group selection and localized interaction can be translated into a kin selection framework [39]. But as stressed in ref. [40], the relatedness of social partners counts less than the effect of the punishment dealt out by an individual on the cooperation received by that individual, and this can be due to facultative adjustment, especially in small groups.

In a different approach [41], costly punishment is explained by two additional factors: (i) a tendency to copy not only the most successful strategy, but also the most frequent; and (ii) 'second-order punishment' directed at those who contribute, but fail to punish. Whether 'higher-order punishment' really occurs is a moot point, as experiments have failed to show evidence for it [42]. However, we surely are conformists. But this tendency to swim with the majority, which helps to stabilize a widespread trait, works against its gaining a foothold in a population dominated by another strategy.

A similar problem besets another model [43], which assumes that, with some small probability, players defect when they learn that their co-players do not punish. This requires information about the others, and a dose of opportunism. Again, a population of punishers cannot be invaded by exploiters, whether first or second order. But conversely, punishers are unable to invade a population of defectors. Thus, the emergence of punishers remains an open issue.

Fowler [44] suggests a possible solution, which exploits the fact that, in many public good interactions, players are not obliged to participate. Imagine that players randomly sampled from the population are offered to participate in a Public Good game, or to stand aside. Those who participate, and find themselves in a group of cooperators, will increase their payoff; those who participate, but land in a group of defectors, will lose. The collective effort is thus a speculation whose success depends on the co-players. This model leads to the emergence of costly punishment, provided a group of two or more cooperators does better than the non-participants [45,46]. If second-order exploiters manage to spread, defectors quickly take over and make the joint effort unattractive. But such episodes are rare and short; when only few are willing to participate in the public good game, cooperation and costly punishment re-appear, and dominate most of the time. Other things being equal, a voluntary Public Good game with Punishment is more likely to prosper than is a compulsory one.

Once punishers have invaded and taken over, all factors mentioned before (conformism, reputation, etc.) can join in and stabilize the propensity to punish. Individual adaptation, in this model, is based on imitation rather than on inheritance. Similarly, the spread from one group to another is easier to conceive as cultural, rather than genetic. But once punishment is established, genetic selection will favour suitable cognitive or emotional adaptations [47].

These optimistic lines do not imply that all problems with costly punishment are solved. There are public good situations that require other models, for instance when participation is compulsory; the substantial percentage of unconditional defectors (or co-operators) remains unexplained; and so on. But group selection is not the only alternative.

However, this alternative is favoured by some experimentalists [48,49], because games with anonymous subjects eliminate all possible effects of relatedness, reputation, future interactions, or signalling. Hence, neither direct nor indirect reciprocity nor kin selection nor costly signalling is at work and thus, the reason for costly punishment has to be the only alternative left, namely group selection. This argument is in the venerable tradition of Sherlock Holmes: “When you have eliminated the impossible, whatever remains, however improbable, must be the truth”. But the experiments also eliminate group benefits [31]. This would be different if the players were told, for instance, that their group was one of several, and that the members of the group with smallest total payoff would lose all their earnings (a setup that is likely to promote cooperation even without punishment). But players know that they are an anonymous sample from a large population, and will disperse after one round of the game. On the other hand, such anonymity is a highly artificial condition, and many doubt that humans have evolved suitable adaptations.

As Burnham and Johnson write [31]: ‘People may behave as if they were from the same evolutionarily relevant group, but in that case we may just as well assume that anonymous subjects behave as if they were related, or as if they were destined to meet again, or as if they are observed by others.’ Thus, kin selection, direct or indirect reciprocity, costly signalling as well as group selection, might all work. Nothing can be excluded out of hand. It could even be that, as in Agatha Christie’s *Murder in the Orient Express*, all suspects collude. Theoreticians tend to look for the most parsimonious explanation, but this principle need not be appropriate to the historical contingencies of human evolution.

PROXIMATE CAUSES OF COSTLY PUNISHMENT

It makes no sense to assume that Ultimatum games or Public Good games, in their clinical sterility, have shaped our evolution. But human behaviour in these games is based on evolved traits. The stark artificiality of economic experiments helps (as in physics or physiology) to reveal the mechanisms underlying these traits.

It appears from cross-cultural studies that the readiness to inflict costly punishment on cheaters is a human universal [8]. It varies across societies, but is strongly correlated with altruistic behaviour, such as the readiness to help others in Dictator games. Brain imaging techniques show that, when players inflict costly punishment, special satisfaction-related zones in the dorsal striatum are activated, indicating physiological adaptations [13]. Both punishing and rewarding appear to be facets of the deep-seated human propensity to reciprocate good and bad, a propensity guided by reasoning and emotions, and based on heuristics and cues.

Two recurrent findings of experimental economics are, on one hand, the diversity within populations, and on the other hand, the flexibility of individuals. All populations appear to be polymorphic, with a substantial percentage displaying little reciprocation. If groups of ‘high trusters’ or ‘low trusters’ are assorted according to simple test questions, they achieve different levels of cooperation in Public Good games [25]. However, many humans can adapt quickly and fine-tune their actions to their social environment. Players do not merely respond

to the threat of punishment or the promise of a reward, but they update constantly, taking account of their experience [5]. If players are told that they will be re-matched with the same co-players, or that their decisions will be made known, they often change their behaviour, obviously motivated by concerns for longer-lasting interactions or for reputation [50-53]. Similarly, if they can opt out of the Public Good game, or back into it, they base their decisions on the current state of the population and adapt rapidly [54]. Voluntary participation elicits a greater readiness to cooperate [55].

This alertness can misfire: it has been shown that by merely seeing the image of an eye, players can be motivated to increase donations substantially [56-58]. Such obvious maladaptions strongly support the hypothesis that our evolutionary legacy shapes our economic behaviour. Similarly, cooperation can be increased by cues of reciprocity or kinship [59]: a face with a family resemblance elicits more help. It is also well known that seemingly unimportant factors (for instance, a preference for Klee rather than Kandinsky) can establish a group identity among complete strangers and boost solidarity [60]. Even in the absence of cues, players could be influenced by the relevant concerns, at least in the sense of hedging their bets. (The tendency to invest roughly half in the first round of a Public good game could be such an insurance policy.)

Our understanding of when and why subliminal factors can affect decision making is far from complete. Players can strongly react to an appropriate cue even when knowing that reality does not back it up. (An example often mentioned in discussions is the sexual arousal produced by centrefolds.) In particular, the fear of punishment can be easily evoked. Under normal circumstances, the donations in the Dictator game (see Box 1) are smaller than in the Ultimatum game, because Proposers understand that low offers cannot be rejected. But if Proposers know that they will be informed of what the Responder thinks of their offer, they offer as much as with the Ultimatum [61]. Such purely symbolic punishment is not very costly any longer. It has been argued that a strong motive for cooperation and moral behaviour is the fear of punishment by supernatural spirits [62]. Superstitious maladaptions are widespread, possibly because they strongly promote conformism and obedience.

Fear, shame, guilt and their converse, the elation and inner glow after a generous action, work to keep humans from cheating. Being cheated arouses anger, indignation and moral outrage, and often causes individuals to inflict costly punishment on defectors.

However, reducing punishment to retaliatory motives and anger at norm-breakers might be premature. Recently, the Public Good experiment with Punishment was repeated, with the difference that the first stage (public good) was replaced by a lottery [63]. Players received randomly assigned sums (distributed as in the Fehr-Gächter experiment [4]) and then could inflict costly 'punishment' just as before, except that they were fully aware that their co-players had done nothing wrong. Many chose to reduce the top earners' income, producing an effect statistically undistinguishable from the reduction of the income of below-average contributors in the Public Good game with Punishment. Inequality arouses negative emotions.

If the Public Good game is repeated, but this time with a fee:fine ratio of 1:1 (the punisher has to pay as much as the punished), then the difference between the two players' payoffs is not altered by punishment (although payoff variance can be reduced). Nevertheless, contributors punish defectors vigorously, and the threat of punishment boosts contributions [64]. Significantly, whereas in the 1:3 treatment defectors sometimes impose sanctions (on defectors and co-operators alike), this rarely happens in the 1:1 treatment. Defectors appear to

be little affected by fairness norms. By contrast, cooperators (who usually punish only defectors) more than double their efforts on imposing fines, obviously willing to incur higher costs to inflict the ‘just’ retribution on wrong-doers.

THE LIMITATIONS OF PEER-PUNISHMENT

Although punishment works to boost cooperation, it can also be counterproductive. It often lowers the average income in Public Good games, despite raising the average level of contributions. In games of trust, or games involving rewards, adding the threat of punishment can decrease the menaced player’s willingness to cooperate [65]. In a particularly elegant set of experiments, it has been shown that, if players of a Public Good game are offered before each round the choice between the versions with or without Punishment, many tend first to shun negative incentives. They need a few rounds to learn to switch to the version with sanctions [66]. Together with the theoretical model of a Public Good game with Punishment, based on voluntary participation [46], this provides a neat application of Hardin’s principle ‘Mutual coercion, mutually agreed upon’ [23].

Punishment is not the only way to enforce cooperation; harassing those having access to a resource [67], chasing shirkers [68] or sabotaging the attempts of cheaters [16], are different examples, and can also be found in other animals, such as mammals, fishes or insects. But humans, with their cognitive capacities for individual recognition, temporal discounting, memory, empathy and language, are uniquely gifted to develop the proximal mechanisms needed for reciprocation and in particular for punishment.

On the other hand, it needs to be stressed that peer-punishment seems relatively rare in real life (in contrast to experiments under anonymity). It can be costly indeed. In small scale societies, or village life, reputation may play a more pervasive role. It is easier to gossip behind the back of a bully than to confront him. Undermining a good reputation is an inexpensive but ominous form of sanctioning, which may eventually lead to ostracism, i.e. exclusion from the market for trustworthy partners. In large societies, peer-punishment is also rare, and repressed by the institutions upholding law and order. Both the pervasive market economy for reliable partners, and the step from peer-punishment to the establishment of sanctioning institutions, deserve closer future investigation.

Acknowledgement: I want to thank Hannelore Brandt, Robert Boyd, Daniel Fessler, Simon Gächter, Mayuko Nakamaru, Martin Nowak, and Bettina Rockenbach for helpful discussions.

References

- 1 Hamilton, W.D. (1996) *Narrow Roads of Geneland, Collected Papers I*, Freeman
- 2 Trivers, R.L. (2002) *Natural Selection and Social Theory: Selected Papers of Robert Trivers*, Oxford University Press
- 3 Clutton-Brock, T.H. and Parker G.A. (1995) Punishment in animal societies, *Nature* 373, 209-216
- 4 Fehr, E. and Gächter, S. (2002) Altruistic punishment in humans, *Nature* 415, 137-140
- 5 Fehr, E and Gächter, S (2000) Cooperation and punishment in public goods experiments, *American Economic Review* 90, 980-994
- 6 Charness G. and Haruvy, E. (2002) Altruism, fairness and reciprocity in a gift-exchange experiment: an encompassing approach, *Games Econ. Behav.* 40, 203-231
- 7 Masclet, D. et al (2003) Monetary and non-monetary punishment in the voluntary contributions mechanism, *Am. Econ. Rev.* 93, 366-380
- 8 Henrich, J. (2006) Costly punishment across human societies, *Science* 312, 176—177
- 9 Price, M.E. et al. (2002), Punitive sentiment as an anti-free rider psychological device *Evol. Hum. Behav.* 23, 203--231
- 10 Baumeister, R.F. et al. (2001) Bad is stronger than good, *Rev. Gen. Psych.* 5, 323-370
- 11 Kurzban, R. and Houser D. (2005) An experimental investigation of cooperative types in human groups: a complement to evolutionary theory and simulations, *Proc. Natl. Acad. Sci. U.S.A.* 102, 1803-1807
- 12 Sanfey, A.G. et al. (2003) The neural basis of economic decision making in the ultimatum game, *Science* 300, 1755-1758
- 13 de~Quervain, D. J.-F. et al. (2004), The neural basis of altruistic punishment, *Science* 305, 1254--1258
- 14 Bolton, G. and Ockenfels, A. (2000) ERC – A theory of equity, reciprocity and competition, *Am. Econ. Rev.* 90, 166-193
- 15 Fehr, E. and Fischbacher, U. (2005) Human altruism – proximate patterns and evolutionary origins, *Anal. Kritik* 27, 6-47
- 16 Wenseleers, T., and Ratnieks, F.L.W. (2006) Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. *American Naturalist* 168:E163-E179.
- 17 Kiers, E. et al. (2003) Host sanctions and the legume-rhizobium mutualism, *Nature* 425, 78-81
- 18 Gneezy, U. and Rustichini, A. (2000) A fine is a price, *J. Legal Stud.* 29, 1-17
- 19 Ostrom, E. and Walker, J. (2003) *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, Russel Sage Funds
- 20 West, S.A. et al., Social semantics: altruism, cooperation, mutualism and strong reciprocity, *J. Evol. Biol.* (in press)

- 21 Samuelson, P.A. (1954) The pure Theory of Public Expenditure, *Rev. Econ. Stat.* 36, 387-389
- 22 Olson M. (1965) *The Logic of Collective Action*, Harvard University Press
- 23 Hardin, G. (1968), The tragedy of the commons, *Science* 162, 1243--1248
- 24 Ostrom, E. (1990) *Governing the Commons*, Cambridge University Press
- 25 Yamagishi, T. (1986) The provision of a sanctioning system as a public good, *J. Personal. Soc. Psychol.* 51, 110-116
- 26 Camerer, C. (2003) *Behavioural game theory: experiments in strategic interactions*, Princeton University Press
- 27 Hammerstein, P. ed. (2003) *Genetic and cultural evolution of cooperation* MIT Press
- 28 Gintis, H., Bowles, S., Boyd, R., and Fehr, E., eds. (2005), *Moral sentiments and material interests: the foundations of cooperation in economic life*, MIT Press
- 29 Wedekind, C. and Milinski, M. (2000) Cooperation through image scoring in humans, *Science* 288, 850-852
- 30 Fehr, E. and Fischbacher, U. (2004) Third-party punishment and social norms, *Evol. Hum. Behav.* 25, 63-87
- 31 Burnham, T. and Johnson, D.D.P. (2005) The biological and evolutionary logic of human cooperation, *Anal. Kritik* 27, 113-135
- 32 Frank, S. (2003) Repression of competition and the evolution of cooperation, *Evolution* 57, 693-705
- 33 Lehmann, L. and Keller, L. (2006) The evolution of cooperation and altruism – a general framework and a classification of models, *J. Evol. Biol.* 19, 1365-1376
- 34 Boyd, R. et al. (2003) The evolution of altruistic punishment, *Proc. Natl. Acad. Sci. U.S.A.* 100, 3531--3535
- 35 Bowles, S. and Gintis, H. (2002) Homo reciprocans, *Nature* 415, 125-128
- 36 Brandt, H. et al. (2003) Punishment and reputation in spatial public goods games, *Proc. R. Soc. B* 270, 1099-1104
- 37 Nakamaru, M. and Iwasa Y. (2005) The evolution of altruism by costly punishment in lattice-structured populations: score-dependent viability vs score-dependent fertility, *Evol. Ecol. Res.* 7, 853-870
- 38 Nakamaru, M. and Iwasa, Y. (2006), The evolution of altruism and punishment: role of the selfish punisher, *J. Theor. Biol.* 240, 475-488
- 39 Lehmann, L Rousset, F, Roze, D and Keller, L (2007) Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment, *American Naturalist*, 170 (in press)
- 40 Gardner A. and West S.A. (2004) Cooperation and punishment, especially in humans, *Am. Nat.* 164, 753-764
- 41 Henrich, J. and Boyd, R. (2001), Why people punish defectors, *J. Theor. Biol.* 208, 79—89
- 42 Kiyonari, T. et al (2004). Second-order punishment in one-shot social dilemma. *Int. J. Psychol.* 39, 329
- 43 Sigmund, K et al. (2001) Reward and punishment, *Proc Nat. Acad. Sci. U.S.A.* 98, 10757-10762
- 44 Fowler, J.H. (2005), Altruistic punishment and the origin of cooperation. *Proc. Natl. Acad. Sci. U.S.A.* 102(19), 7047--7049

- 45 Brandt, H. et al.(2006), Punishing and abstaining for public goods, *Proc. Natl. Acad. Sci. U.S.A.* 103, 495-497
- 46 Hauert, C. et al. (2007) Between freedom and coercion: the emergence of altruistic punishment, *Science* (in Press)
- 47 Richerson, P. and Boyd, R. (2005) *Not by genes alone*, University of Chicago Press
- 48 Fehr, E. and Fischbacher, U. (2005) Human altruism – proximate patterns and evolutionary origins, *Anal. Kritik* 27, 6-47
- 49 Fehr, E. and Henrich, J. (2003) Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism, in P. Hammerstein (ed) *The Genetical and Cultural Evolution of Cooperation*, Cambridge University Press, 55-82
- 50 Sell, J. and Wilson R. K. (1999) The maintenance of cooperation: expectations of future interactions and the trigger of group punishment, *Social Forces* 77, 1551-1570
- 51 Small, D. A. and Loewenstein G. (2005) The devil you know: the effects of identifiability on punishment, *J. Behav. Dec. Mak.* 18, 311-318
- 52 Rege, M. and Telle, K. (2004) The impact of social approval and framing on cooperation in public good situations, *J. Pub. Econ.* 88, 1625-644
- 53 O’Gorman, R. et al. (2005) Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions, *Evol. Hum. Behav.* 26, 375-387
- 54 Semmann, D. et al. (2003), Volunteering leads to rock-paper-scissors dynamics in a public goods game, *Nature* 425, 390--393
- 55 Orbell, J.H. and Dawes, R.M. (1993) Social welfare, cooperator’s advantage, and the option of not playing the game, *Amer. Soc. Rev.* 58, 787-800
- 56 Haley, K and Fessler D. (2005) Nobody’s watching? Subtle cues affect generosity in an anonymous economic game, *Evol. Hum. Behav.* 26, 245-256
- 57 Burnham, T and Hare, B (2007), Engineering cooperation: does involuntary neural activation increase public goods contributions?, *Human Nature* (in press)
- 58 Bateson, M, Nettle, and Roberts, G (2006) Cues of being watched enhance cooperation in a real-world setting, *Biol. Lett.* 2, 412-414
- 59 DeBruine L. (2005) Facial resemblance enhances trust, *Proc. R. Soc. Lond. B* 269, 1307-1312
- 60 Yamagishi, T. et al.(1999) Bounded generalised reciprocity: ingroup boasting and ingroup favouritism, *Adv. Group Proc.* 16, 161-197
- 61 Xiao, E. and Houser, D. (2005) Emotion expression in human punishment behaviour, *Proc. Natl. Acad. Sci. U.S.A.* 102, 7398-7401
- 62 Johnson D. and Bering, J. (2006) Hand of God, mind of man: punishment and cognition in the evolution of cooperation, *Evol. Psychol.* 4, 219-233
- 63 Dawes, C.T. et al. (2007) Egalitarian motives in humans, *Nature* 446, 794-796
- 64 Falk, A, Fehr, E and Fischbacher, U (2005) Driving Forces behind Informal Sanctions, *Econometrica* 73, 2017-2030
- 65 Fehr, E. and Rockenbach, B. (2003) Detrimental effects of sanctions on human altruism, *Nature* 422, 137--140
- 66 Güererk, O. et al. (2006) The competitive advantage of sanctioning institutions, *Science* 312, 108--111

- 67 Stevens, J.R. et al.(2005) Evolving the psychological mechanisms for cooperation, *Ann Rev of Ecology, Evol.Syst.*, 36, 499-518
- 68 Bshary, R. and Grutter, A.S. (2005) Punishment and partner switching causes cooperative behaviour in a cleaning mutualism, *Biology Letters* 1, 396-399
- 69 Dickinson, D.L. (2001) The carrot vs. the stick in work team motivation, *Exp. Econ.* 4, 107-124
- 70 McCabe K. et al. (2003) Positive reciprocity and intentions in trust games, *J. Econ. Behav. Org.* 52, 267-275
- 71 Sefton, M, Shupp R and Walker J (2007) The effects of rewards and sanctions in provision of public goods, *Economic Inquiry* (in Press)
- 72 Andreoni, J. et al. (2003) The carrot or the stick: rewards, punishments, and cooperation, *Am. Econ. Rev.* 93, 893-902
- 73 Milinski, M. et al.(2002) Reputation helps solve the ‘Tragedy of the Commons’, *Nature* 415, 424-426
- 74 Panchanathan, K. and Boyd, R. (2006), Indirect reciprocity can stabilize cooperation without the second-order free rider problem, *Nature* 432, 499--502
- 75 Rockenbach, B. and Milinski, M. (2006) The efficient interaction of indirect reciprocity and costly punishment, *Nature* 444, 718-723

Box 1: Game Zoo: a brief lexicon of two-person games

Many experimental two-person games are related to the issues of public goods [26]. Typically, the players are anonymous, and are endowed with a certain amount of money beforehand (e.g. a show-up fee). They are asked to make their decision after having understood the rules of the game and being assigned to the role of Proposer and Responder (or Donor and Recipient).

Gift Giving: in some sense, an atom of social interaction. The Donor decides whether to pay one dollar to confer a benefit of three dollars on the Recipient.

Prisoner's Dilemma: the mother of all cooperation games is played in many variations. In one particularly transparent set-up, both players engage in a Gift Giving Game with each other. When players decide simultaneously, this is similar to a two-player Public Good game. If both cooperate by sending a gift to the other, both gain two dollars. But sending a gift costs one dollar, so that the best reply to whatever the co-player decides is not to cooperate (i.e. to defect). If both players defect, however, they gain nothing.

Ultimatum: the experimenter assigns a certain sum, and the Proposer can offer a share of it to the Responder. If the Responder (who knows the sum) accepts, the sum is split accordingly between the two players, and the game is over. If the Responder declines, the experimenter withdraws the money. Again, the game is over: but this time, none of the two players gets anything.

Dictator: same as Ultimatum, except that the Responder cannot reject the offer.

Trust: in a first stage, the Proposer can confer a certain benefit on the Responder, as in the Gift Giving Game. In the second stage, the Responder can decide how much of it to return to the Proposer. This is similar to the sequential Prisoner's Dilemma Game (when first one player acts as Donor and then the other).

Repeated Prisoner's Dilemma: the two players interact for several rounds of the Prisoner's Dilemma. Usually, they are not told beforehand when the interaction will be over, so as to avoid 'last round effects' (defection motivated by the fact that the co-player cannot retaliate).

Indirect Reciprocity: in a large population of players, two players are sampled at random and play the Gift Giving game or the (non-repeated) Prisoner's Dilemma game. This is repeated again and again. The players know that they interact only once, so that retaliation is impossible.

Box 2: Tools for Fitness: a semantic guide to the evolution of cooperation

Punishing and rewarding are responses to previous actions. In principle, **reciprocation** can only occur as the second stage of an interaction; however, the first stage of the interaction (the contribution to the public good, or the offer in the Ultimatum Game) is often affected by the expectation of a return. The challenge is to explain that return (i.e. the second stage).

Reciprocity operates even with third parties. This is called **indirect reciprocation** and comes in two flavours. A bystander watching Joe harm Bill (or help Bill) is likely to harm (respectively help) Joe, in turn: that is vicarious reciprocity. Conversely, an individual who has been harmed, or helped, by some agent Ted, can vent his anger, or his gratitude, on some passerby, Bob: that is misdirected reciprocity. These effects have been documented in experiments [29, 30]. For instance, the propensity of a bystander to punish defectors in a Prisoner's Dilemma game is high, except if both players defected and, thus, in a sense, performed the punishment themselves.

The emotionally driven disposition to return good with good, and bad with bad, is called **strong reciprocity** by a group of researchers who co-operated on a vast effort to study economic games in small-scale societies. Despite substantial differences in their views, they are often perceived as an in-group around the banner of strong reciprocity [31].

Reciprocation is usually costly, whether it is **altruistic** (rewarding) or **spiteful** (punishing). Evolutionary biologists have to understand its adaptive value, and show that such behaviour is (on average) not fitness-reducing after all. The classical approaches invoke interactions between kin, or mutual benefits to cooperating individuals [20]. Both approaches rely on positive assortment between individuals conferring help. The explanation is genetic in one case, and economic in the other.

Kin selection operates if a loss in **direct fitness** (the own reproductive success) is compensated by a gain in **indirect fitness** (the reproductive success of related individuals)[32,33].

Mutual benefits to cooperating individuals can accrue, for instance, if the same two players engage in a long chain of give-and-take, as with **direct reciprocity**, or if reputation effects enable cooperators to channel benefits towards those individuals who are benefiting others, as with **indirect reciprocity**. Costly **signalling** can be subsumed under this heading, if players who are able to signal higher value (because they can afford to contribute) are preferentially chosen as partners or mates and, thus, obtain benefits in return.

Box 3: Punishing Logic: the evolutionary problems raised by costly punishment

Punishers raise two evolutionary riddles. They cannot invade; and they can be invaded.

A minority of punishers invading a population of defectors would have to punish left and right. Because each act of punishment is costly, punishers would suffer heavily, whereas the defectors would barely be affected. Hence, punishers would be at a disadvantage, and soon eliminated from the population.

Conversely, suppose that a population is dominated by punishers. Defectors, in that case, cannot invade: a minority of defectors would have to bear the full brunt of punishment from the majority, which more than offsets their gain from not contributing. This is a bi-stable situation: defectors cannot invade punishers, punishers cannot invade defectors. But suppose that a new type enters the population, one who contributes, but does not punish. Such a type can easily arise through recombining traits. The newcomers do just as well as the resident punishers and, thus, can slowly spread by neutral drift. In fact, if occasionally some defectors entered the population (to be promptly eliminated by the punishers), the new type would do better than the punishers, by economizing on the cost of punishment. This new type is a second-order exploiter, free-riding on the sanctions provided by the punishers. Hence, it will spread: and this means that eventually, there will be too few punishers to keep the defectors at bay. Thus, second-order exploiters sabotage the enforcement of contributions to the public good game in the long run and, therefore, sap the basis for the punishers, and for themselves: both contributing types will be displaced by defectors.

A remedy coming to mind is 'second-order punishment' (i.e. in punishing not only the 'first order exploiters' who fail to contribute, but also the 'second order exploiters' who contribute, but fail to punish). However, this could give rise to 'third order exploiters' and so on. If punishers of a sufficiently high order dominate the population, there will be few defectors and, hence, few occasions, for lower-order punishers to reveal their limitations to their fiercer brethren. Thus, they would rarely be selected against, and could spread by neutral drift, sapping the system. Clearly, higher-order punishers cannot gain a foothold in a population of defectors.

Box 4: The carrot: the role of rewards as incentives for cooperation

Investigations comparing negative with positive incentives (i.e. the carrot with the stick), show that rewards are considerably less efficient than punishment, at least for the games considered here, where the public good is a linear function of the number of contributors. Positive incentives become costly, but negative cheap, if success is fully achieved, i.e. all cooperate [10,69-71].

Andreoni et al. [72] studied four treatments of a Proposer-Responder game in which the Proposer had to choose how much to share of a given sum. Depending on the treatment, the Responder had subsequently the possibility, (i) to reward or to punish the Proposer; (ii) only to reward; (iii) only to punish; or (iv) neither/nor. Treatment (iv) reduces to the Dictator game (Box 1); treatment (iii) with the punishing option differs from the Ultimatum Game (Box 1) because the Responder has more freedom in choosing the proper sanction.

Rewards alone prove ineffective. Punishment, by contrast, often induces offers close to 50%. Adding the possibility of rewarding yields a remarkable outcome: half the offers are >50%, and more than a quarter of the Proposers offer 100%. The corresponding reward is half of that. Punishment is hardly affected by the availability of rewards, but rewarding is considerably more pronounced if there is no possibility of punishment.

A particularly interesting system of incentives is considered in [73] and [74]. Between rounds of the Public Good game without Punishment, the population engages in pair-wise interactions of indirect reciprocity (Box 1). Since players tend preferentially to help those who contributed to the public good, this effectively provides rewards which are not costly because those who reward earn a good reputation and thus benefit in later rounds of the indirect reciprocity game. If in addition to indirect reciprocity, the players have an opportunity to directly peer-punish those who do not contribute, they use this opportunity less often, but in a more focussed way, see Figure 1 [75]. Costly punishment and rewarding through indirect reciprocity combine very efficiently to boost cooperation in group interactions.

Fig. 1 Learning to exploit vs. teaching to cooperate. In [4], groups of players engage in six rounds of Public Good without Punishment, followed by six rounds with Punishment. Shown are the average contributions per round. The groups are newly formed between rounds, so that players never interact with a co-player twice. It should be stressed that in the rounds with punishment, the average income is usually below that without punishment: punishment is costly. But in later rounds, when most players cooperate, punishment should be rare. Ideally, it is no longer needed except as a threat. This should yield a stable and economically efficient collaboration. (After Fehr and Gächter [4]).

Fig.2 Peer punishment vs. reputation-building (after Rockenbach and Milinski [75]). In the PUN treatment, player can punish their co-players after every round of a Public Good game. In the PUN&IR treatment, they engage, in addition, in several rounds of indirect reciprocity (see Box 1) after each round of the Public Good game. Such rounds permit to reward players who have contributed to the public good. By withholding a possible donation in the indirect reciprocity game, players can effectively sanction free riders without paying a cost. This reduces the amount of direct, costly punishment (a), but does not eliminate it. Rather, the direct punishment is now more focussed towards free-riders (b), and considerably more efficient in boosting contributions (c).

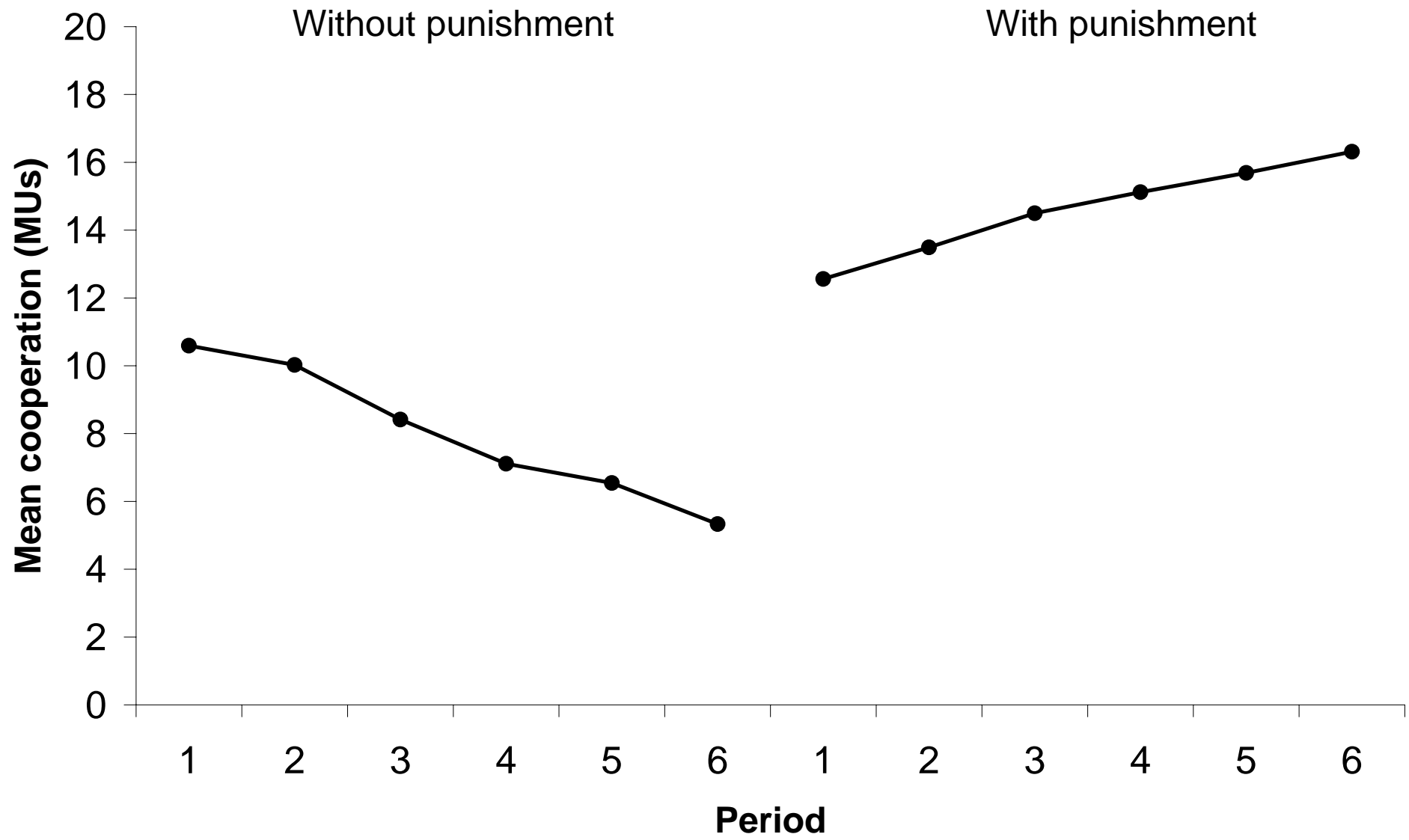


Figure 1

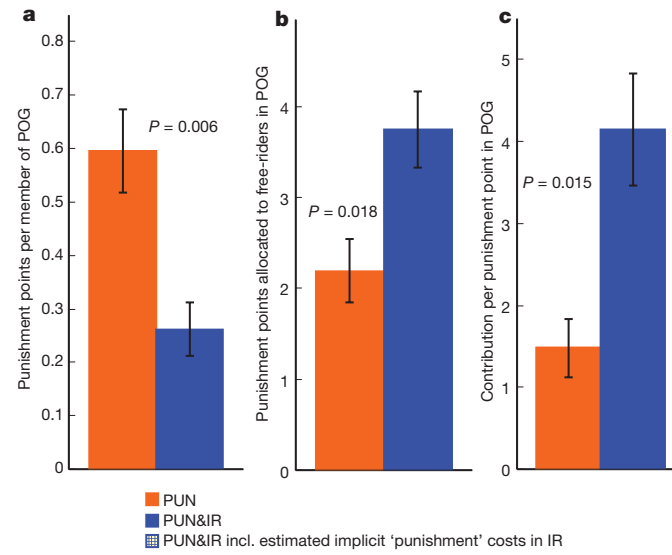


Figure 2