12. Loughrin, J. H., Manukian, A., Heath, R. R. & Tumlinson, J. H. Diurnal cycle of emission of induced volatile terpenoids by herbivore-injured cotton plants. *J. Chem. Ecol.* **21,** 1217–1227 (1994).
13. Takabayashi, J., Dicke, M. & Posthumus, M. A. Variation in composition of predator-attracting allelochemicals emitted by herbivore-infested plants: Relative influence of plant and herbivore. *2,* 1–6 (1991).
14. Du, Y.-J., Poppy, G. M. & Powell, W. Relative importance of semiochemicals from first and second trophic levels in host foraging behavior of *Aphidius ervi*. *J. Chem. Ecol.* **22,** 1591–1605 (1996).
15. Lewis, W. J. & Takasu, K. Use of learned odours by a parasitic wasp in accordance with host and food needs. *Nature* **348,** 635–636 (1990).
16. Tumlinson, J. H., Lewis, W. J. & Vet, L. E. M. How parasitic wasps find their hosts. *Sci. Am.* **268,** 100–106 (1993).
17. Bell, W. J., Kipp, L. R. & Collins, R. D. in *Chemical Ecology of Insects 2* (eds Cardé, R. T. & Bell, W. J.) 105–154 (Chapman & Hall, New York, 1995).
18. Strand, M. R. & Obrycki, J. J. Host specificity of insect parasitoids and predators. *BioScience* **46,** 422–429 (1996).
19. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19,** 207–233 (1988).
20. Thompson, J. N. *The Coevolutionary Process* (Univ. of Chicago Press, Chicago, 1994).
21. Röse, U. S. R., Manukian, A., Heath, R. R. & Tumlinson, J. H. Volatile semiochemicals released from undamaged cotton leaves: A systemic response of living plants to caterpillar damage. *Plant Physiol.* **111,** 487–495 (1996).
22. Heath, R. R. & Manukian, A. *J. Chem. Ecol.* **20,** 593–608 (1994).

Acknowledgements. We thank M. C. Mescher for assistance in preparing the manuscript; T. C. Turlings, P. J. Landolt, J. Garcia, B. Benrey, K. Ross, K. Korth and J. Ruberson for comments on the manuscript; A. T. Proveaux for assistance with mass spectrometric analysis; J. H. Loughrin for leaf measurement software; and T. Green for insect rearing. C.D.M. is recipient of fellowship from CAPES (Brazil).

Correspondence and requests for materials should be addressed to W.J.L. (e-mail: WJL@tifton.cpes.peachnet.edu).

Evolution of indirect reciprocity by image scoring

Martin A. Nowak*† **& Karl Sigmund**‡

* *Department of Zoology, University of Oxford, Oxford OX1 3PS, UK*
‡ *Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria*

Darwinian evolution has to provide an explanation for cooperative behaviour. Theories of cooperation are based on kin selection (dependent on genetic relatedness)[1,2], group selection[3–5] and reciprocal altruism[6–9]. The idea of reciprocal altruism usually involves direct reciprocity: repeated encounters between the same individuals allow for the return of an altruistic act by the recipient[10–16]. Here we present a new theoretical framework, which is based on indirect reciprocity[17] and does not require the same two individuals ever to meet again. Individual selection can nevertheless favour cooperative strategies directed towards recipients that have helped others in the past. Cooperation pays because it confers the image of a valuable community member to the cooperating individual. We present computer simulations and analytic models that specify the conditions required for evolutionary stability[18] of indirect reciprocity. We show that the probability of knowing the 'image' of the recipient must exceed the cost-to-benefit ratio of the altruistic act. We propose that the emergence of indirect reciprocity was a decisive step for the evolution of human societies.

Humans have achieved one of the pinnacles of sociality, and the complexity of their cooperative actions is without parallel. In contrast to other examples of ultrasociality[19–22] (for example, clones, bee hives or termite colonies), human cooperation is due less to kin selection than to cultural forces rooted in pervasive moral systems. From hunter tribes and village communities to nation states and global enterprises, the economic effects of nepotism, although certainly present, are minor compared with those of reciprocity. Reciprocity is usually understood to take the form of direct reciprocity: help someone who may later help you. But indirect reciprocity also prevails in human communities. In this case, one does not expect a return from the recipient, but from someone else, according to the pious advice of 'give, and you shall be given'. Cooperation is channelled towards the 'valuable' members of the community. This has been called the 'I won't scratch your back if you won't scratch their backs' principle[23]. A donor provides help if the recipient is likely to help others (which often means, if the recipient has helped others in the past). In this case, it pays to advertise cooperation, as the cost of an altruistic act is offset by an increased chance to become the recipient of an altruistic act later. Animal and human behaviour may be influenced by attempting to increase image (or status) in the group[24,25].

According to Alexander[17], indirect reciprocity, which "involves reputation and status, and results in everyone in the group continually being assessed and reassessed", is important in human societies (and possibly in some primates, social canines and other groups). Alexander interprets moral systems as systems of indirect reciprocity. Indirect reciprocity presupposes rather sophisticated players, and therefore is likely to be affected by anticipation, planning, deception and manipulation. The politicking needed to

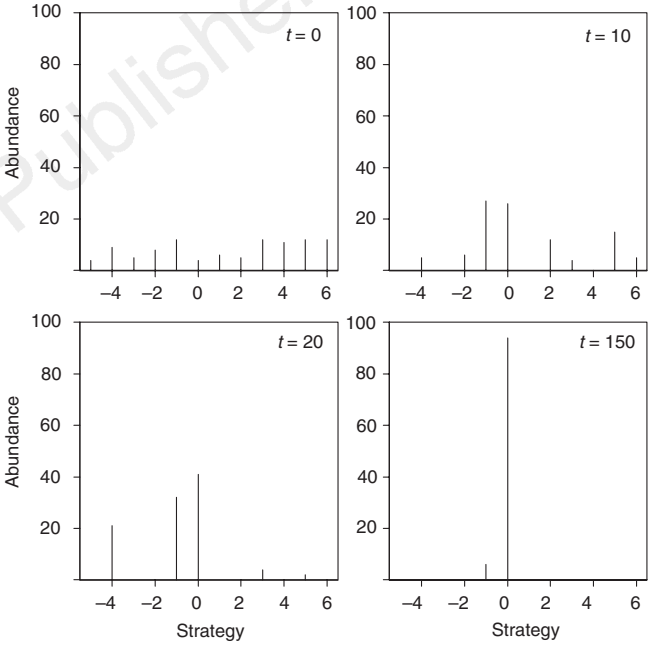

Figure 1 Cooperation wins in a computer simulation of indirect reciprocity. The population consists of $n = 100$ individuals. The image scores range from –5 to +5, the strategy (k) values from –5 to +6. The strategy $k = -5$ represents unconditional cooperators, whereas the strategy $k = +6$ represents defectors. In each round of the game, two individuals are chosen at random, one as donor, the other as recipient. The donor cooperates if the image score of the recipient is greater than or equal to the donor's k value. Cooperation means the donor pays a cost, c, and the recipient obtains a benefit, b. There is no payoff in the absence of cooperation. At the beginning of each generation, all players have image score 0. Hence, strategies with $k \leq 0$ are termed 'cooperative', because individuals with these strategies cooperate with individuals that have not had an interaction. In each generation 125 donor–recipient pairs (n) are chosen; each player has, on average, 2.5 interactions. The chance that a given player meets the same player again, or that a chain of possible altruistic acts ever leads back to the original donor, is negligibly small. Therefore, direct reciprocity cannot work here. At the end of each generation, players produce offspring proportional to their payoff. At generation, $t = 0$, we start with a random distribution of strategies. After $t = 10$ generations, the strategies $k = -1, 0, +2$ and $+5$ have increased in abundance. After $t = 20$ generations, the strategies $k = -4, -1$ and 0 dominate the population. After $t = 150$ generations, the population consists almost entirely of the strategy $k = 0$, which is the most discriminating among all cooperative strategies. Players with this strategy cooperate with everyone who has image score 0 or greater. After $t = 166$ generations, all other strategies have become extinct and $k = 0$ is fixed in the population. Parameter values: $b = 1, c = 0.1$ (to avoid negative payoffs we add 0.1 in each interaction).

† Present address: Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA.

continually assess the status of all members of our community and to bolster our own has probably been a major force for shaping our intelligence. But if we want to understand the basic mechanisms of indirect reciprocity, we must analyse drastically simplified models.

Imagine a population of individuals with the option to help one another or not. Random pairs of players are chosen, of which one is the potential donor of some altruistic act and the other is the recipient. The donor can cooperate and help the recipient at a cost c to himself, in which case the recipient receives a benefit of value b (with $b > c$). If the donor decides not to help, both individuals receive zero pay-off. Each player has an image score, s, which is known to every other player. If a player is chosen as a donor and decides to cooperate then his (or her) image score increases by one unit; if the donor does not cooperate then it decreases by one unit. The image score of a recipient does not change. First we consider strategies where donors decide to help according to the image score of the recipient. A strategy is given by a number k: a player with this strategy provides help if, and only if, the image score of the potential recipient is at least k.

Figure 1 shows computer simulations of a population consisting of n players. The strategies are given by k_i and the image levels by s_i. At the beginning of each generation, the image levels of all players are zero (assuming that children do not inherit the image of their parents). In succession, m donor–recipient pairs are chosen. A donor, i, cooperates with a recipient, j, if $k_i \leq s_j$. The fitness of a player is given by the total number of points received during the m interactions. Some players may never be chosen, in which case their payoff from the game will be zero. On average, a player will be chosen $2m/n$ times, either as donor or as recipient. At the end of each generation, players leave offspring in proportion to the their fitness. We find that if the game is played for many generations, then eventually all players will adopt the same strategy. If the k value of this strategy is 0 or less then cooperation is established; if the value is 1 or more then defection has won. Cooperation is more likely to win if the number of interactions, m, per generation is large. (A different model of indirect reciprocity has been studied by Boyd and Richerson[26], who assumed that individuals interact in loops such that a cooperative action can be returned, after several steps, to the original donor. According to Boyd and Richerson, their model is unlikely to lead to a cooperative outcome, as it requires the loops to

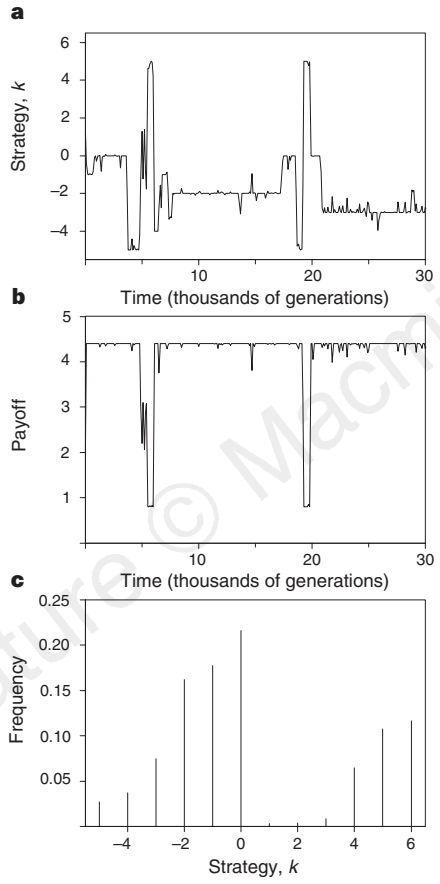

Figure 2 Long-term evolution of indirect reciprocity under mutation and selection. We used the same computer simulation as in Fig. 1, but included mutation: there is a probability of 0.001 that an offspring does not act like its parent and uses another randomly chosen strategy instead. We observe endless cycles of cooperation and defection. Cooperative populations are relatively stable if they consist of discriminating players with strategies such as $k = 0$ or –1. But after some time these populations are undermined (through random drift) by players with strategies such as $k = -4$ or –5, which are too cooperative. Then defectors, with strategies $k = 4$ or 5, can invade. These defectors can, in turn, be overcome by stern discriminators again. In the long run, cooperation is harmed by unconditional cooperators, because they enable defectors to invade. In the absence of unconditional cooperators, cooperative populations persist for much longer. **a**, The average k value of the population. **b**, The average payoff per individual, per generation. **c**, Frequency distribution of strategies sampled over many generations ($t = 10^7$). Parameter values are as for Fig. 1, but $m = 300$ rounds per generation.

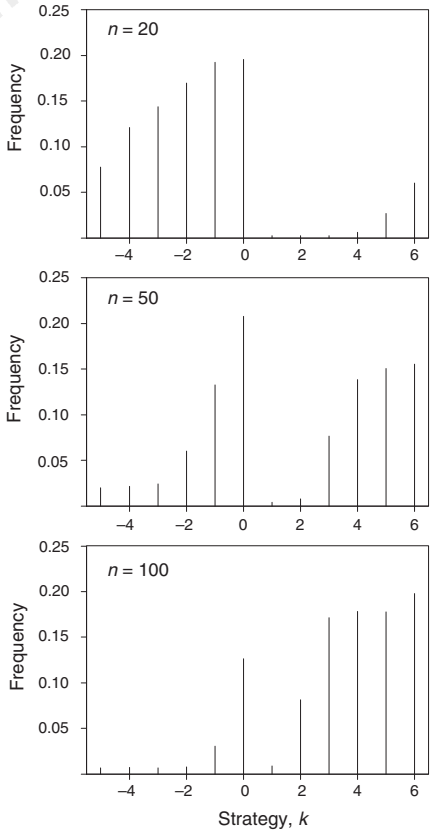

Figure 3 Indirect reciprocity with incomplete information about the image score of other players. We performed the same simulation as in Fig. 2, but updated the image score of a donor only for the recipient and for the observers of an interaction. Each interaction is observed, on average, by ten randomly chosen players. The figure shows the frequency distribution of strategies for three different population sizes, $n = 20$, $n = 50$ and $n = 100$, sampled over many generations ($t = 10^7$) in order to obtain representative results. There is a clear effect of group size: cooperation predominates for $n = 20$, but is rare for $n = 100$. For $n = 50$ we find cooperative and defective strategies at roughly equal frequencies. The averages over time of the frequency of cooperative strategies (defined by $k \leq 0$) are 90%, 47% and 18% for, respectively, $n = 20$, 50 and 100. Parameter values are as for Fig. 2, but the number of donor–recipient interactions, m, is $10n$ rounds per generation.

be relatively small, closed and long lasting. We think that this is because their model does not include image scores.)

We can also include mutations in the simulation, by assuming that there is a small probability that a strategy does not reproduce accurately but instead gives rise to an offspring adopting a different strategy (Fig. 2). In this case, several strategies can persist. We have studied the frequency distribution of various strategies and analysed how often a cooperative regime is achieved. A minimum number of rounds per generation is needed for cooperation to prevail. This number can be very small: each player needs to be chosen for only about two interactions per lifetime. (In this case, there is a probability of only 1/4 that a defector can be punished, if he is chosen first as a donor and then as a recipient.) Below we present an analytical model for evaluating the minimum number of interactions that is compatible with cooperation.

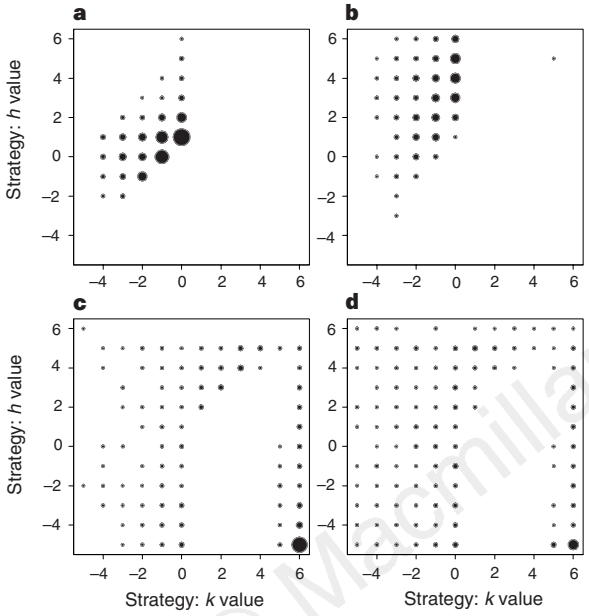

Figure 4 A further dimension is added to the game if donors base their decision to cooperate not only on the image score of the recipient but also on their own score. **a**, **b**, We consider strategists that cooperate if the image score of the opponent is at least k *and* if their own image score is less than h. If the image score of an individual is already high, it makes no sense to invest in a still higher image. The figures show the frequency distribution of strategies that are defined by their k and h values sampled over many generations. **a**, We assume perfect information about the image of all players. The most frequent strategy is $k = 0, h = 1$. Players with this strategy cooperate if the image score of the opponent is at least 0 and their own image score is less than 1. If the whole population adopts this strategy, it does not pay to aim for an image exceeding 0. For the same reason, other strategies with $h = k + 1$ are successful in this simulation. Strategies with $k > 0$ are unsuccessful, because they are too uncooperative. **b**, We assume imperfect information about the other players' image. Here it pays to invest in a higher image than strictly necessary, because a given altruistic act is only seen by a subset of other players. The most frequent strategy is $k = 0, h = 4$. **c**, **d**, We study players that cooperate if the image score of the recipient is at least k *or* if their own image score is less than h. Players with a low image may want to increase their image by helping others indiscriminately. Such a scenario also leads to cooperative societies (dominated by strategies with $k \leq 0$), but unconditional defectors (with strategies $k = 6, h = -5$) benefit from the reduced level of discrimination and represent the most frequent single strategy. **c**, Results are based on perfect information whereas **d** assumes imperfect information about the co-players' image score. In **a**, **b**, **c** and **d**, respectively, the frequency of cooperative interactions is 55%, 57%, 70% and 80%. This must be compared with <0.1% cooperation in simulations where strategies only consider their own image score and do not discriminate according to the image score of the recipient. Parameter values: **a**, **c**, as in Fig. 2, but $m = 500$ rounds per generation; **b**, **d**, as in Fig. 3 with $n = 20$. The frequency of a strategy is proportional to the area of the shaded circles. Strategies with a frequency of <0.5% are not shown.

Long-term simulations that include mutation usually do not converge to a simple equilibrium distribution of strategies, but show endless cycles. In simple terms, what happens is that defectors are invaded by discriminators, who only help players whose score exceeds some threshold. Next, discriminators are undermined by unconditional cooperators. The prevalence of these indiscriminate altruists subsequently allows the return of defectors. In a population consisting only of discriminators and unconditional cooperators, there is no selection against the latter, who can spread by random drift. In simulations without unconditional cooperators, cooperative populations persist for much longer.

Cooperation based on indirect reciprocity depends crucially on the ability of a player to estimate the image score of the opponent. In the above model, we assume that the image score of each individual is known to every other member of the population. This should be seen as only an idealized scenario. It is more realistic to assume that an interaction between two individuals is observed by a (possibly small) subset of the population. Only these 'onlookers' (and, of course, the recipient) have the possibility of updating their perception of the donor's image score. The onlookers are chosen at random for each particular interaction. Therefore each player has a specific perception of the image score of the other players. The same player can have different image scores in the eyes of different individuals. The information is contained in a matrix whose elements s_{ij} denote the image score of player i as seen by player j. In a donor–recipient interaction between j and i, player j will cooperate if $s_{ij} > k_j$. If j has no information on i, then $s_{ij} = 0$.

The model now depends on the probability that a given individual observes an interaction between two other individuals. Figure 3 shows computer simulations of this extended model. Again, cooperation can easily be established and dominate the population, but a larger number of interactions per generation is needed. There is also an effect of group size. For larger groups, it is more difficult to establish cooperation, because the fraction of individuals that obtain information about any particular interaction will be smaller. Therefore, more interactions are required (relative to group size) in order to discriminate against defectors.

Another interesting expansion of the basic model is to include strategies that consider both the recipient's and the donor's image score. We explored two types of strategies. 'And' strategies involve cooperation if the image score of the recipient is larger than a certain value *and* the image score of the donor is less than a certain value. The idea is that if an individual has already a high image score, it is not necessary to aim for a still higher image score (by helping others). On the other hand, 'or' strategies result in cooperation if the image score of the recipient is larger than a certain value *or* the image score of the donor is less than a certain value. Here the idea is that if an individual has a low image score it may be advantageous to increase the score by helping others regardless of how low their image score is. In both cases, highly cooperative societies form (Fig 4). If, in contrast, we simulate strategies that only consider their own image and do not take into account the image of the recipient, cooperation does not emerge.

The models above are based on computer simulations, but we can derive analytical insights from a simplified model. Suppose that there are only two image levels, 0 (for bad) and 1 (for good). The image of a player depends on his or her last action as a donor: players who defected have score 0, and players who cooperated have score 1. Let us only consider two types of player: first, defectors, who never provide assistance; and second, discriminators who help players having image 1, but not players having image 0. A given player knows the score of only a fraction, q, of the population. A discriminator who has no information on potential recipients will assume, with a certain probability, p, that they have image 1. In each round of the game all individuals of the population are chosen, each

with the same probability of being a donor or a recipient. If $w < 1$ denotes the probability of another round, there are on average $1/(1-w)$ rounds per generation. We have derived the equations (see Methods) that describe how the frequencies of discriminators and defectors change from one generation to the next. It should be stressed that discriminators are not 'tit-for-tat' players; tit-for-tat strategists base their decisions on their own previous experience with the co-player, whereas discriminators use the experience of others. This is an essential advantage for a player who interacts with many co-players but only a few times with each. (Such discriminators are also different from strategies based on 'standing'[27], which is an internal switch distinguishing between defection in response to a co-player's cooperation or defection.)

We observe a frequency threshold: a minimum amount, $x_{\min}$, of discriminators is necessary to ensure the establishment of cooperation. We also obtain the minimum number of rounds per generation that is needed for the evolutionary stability of discriminators. In particular, cooperation through indirect reciprocity can only be stable if $q > c/b$. The probability of knowing the image of another player has to exceed the cost-to-benefit ratio of the altruistic act. This is remarkably similar to Hamilton's rule, which states that cooperation through kin selection works whenever the coefficient of relatedness between two individuals exceeds the cost-to-benefit ratio[1,2]. In our case, relatedness is replaced by acquaintanceship.

Cooperation based on indirect reciprocity works in the following way, therefore: a potential donor can choose whether to accept a certain cost in order to help another individual, or to avoid this cost. In the short term, of course, avoiding the cost yields the higher payoff. In the long term, however, performing the altruistic act increases the image score of the donor and may therefore increase the chance of obtaining a benefit in a future encounter as a recipient. On the other hand, a discriminator who punishes low-score players by refusing them help pays for this by having his own score reduced. The overriding idea, relevant to human societies, is that information about another player does not require a direct interaction, but can be obtained indirectly either by observing the player or by talking to others. The evolution of human language as a means of such information transfer has certainly helped in the emergence of cooperation based on indirect reciprocity. □

Methods

Defectors and discriminators. Here we develop a simplified model for indirect reciprocity which can be fully understood in analytical terms. Consider two image scores, 0 for someone who defected last round and 1 for someone who cooperated last round. Thus the image score depends only on the last move of a player as a donor. Consider two types of players: discriminators, who help only players with image score 1, and defectors, who never help. Let us suppose that there is a probability, q, that discriminators have information about the image score of the recipient. In the absence of information, they assume an image score of 1 with probability p. (One can show that if indirect reciprocity works at all, then discriminators with larger p always outcompete the others[28]. Therefore we shall restrict ourselves in the following to the limiting value $p = 1$. The discriminator strategy in this case coincides with a variant of 'tit-for-tat', which begins with defection if the future co-player has been seen defecting in his last interaction[29]—a confirmation of Alexander's view that "indirect reciprocity is a consequence of direct reciprocity occurring in the presence of others"[17].) For a defector, information about the image score does not matter. We denote by x_0, x_1, y_0 and y_1, respectively, the frequencies of discriminators (x) with images 0 and 1, and the frequencies of defectors (y) with images 0 and 1. The total frequency of discriminators is $x = x_0 + x_1$ and that of defectors is $y = y_0 + y_1$. We have $x + y = 1$. A generation consists of several rounds of the game, during which x and y do not change. In each round all players are paired up, half of the players being donors, the other half recipients. The frequencies of players of image 0 or 1 change from round to round according to the difference equations $x'_0 = [x_0 + x(1-\psi)q]/2$, $x'_1 = [x_1 + x(1-q+q\psi)]/2$, $y'_0 = [y_0 + y]/2$, and $y'_1 = y_1/2$. Here $\psi = x_1 + y_1$ is the frequency of players with score 1. In each round, the payoff to the individual types is $P(x_0) = [-c(1-q+q\psi) + bx(1-q)]/2$, $P(x_1) = [-c(1-q+q\psi) + bx]/2$, $P(y_0) = bx(1-q)/2$, $P(y_1) = bx/2$. The difference equation yields the expected payoff values $D_e(k)$ and $D_i(k)$ to defectors and discriminators in the kth round: $D_e(k) = bx(1-q+q2^{-(k-1)})/2$, and $D_i(k) = D_e(k) + \{(1-q)(bqx-c)(1-qx)^{-1} - bq2^{-(k-1)} + q(b-c)(1-x)(1-qx)^{-1}[(1+qx)/2]^{k-1}\}/2$. We can assume either that the number of rounds per generation is constant, or that there exists a fixed probability w for a further round. In the latter case, the total payoff to defectors is $D_e = \Sigma_{k=1}^{\infty} w^{k-1} D_e(k)$, and similarly for discriminators. We find that

$$2(D_i - D_e) = (1-q)\frac{bqx - c}{(1-w)(1-qx)} + 2q\left[\frac{(b-c)(1-x)}{(1-qx)(2-w-wqx)} - \frac{b}{2-w}\right]$$

Modelling the change in frequency of discriminators and defectors from one generation to the next by the standard replicator equation[30], we find that defectors win if x is below a threshold value $x_{\min}$ given by $D_i = D_e$, whereas discriminators win if x is above this threshold. Discriminators are evolutionarily stable if and only if $x_{\min} < 1$, that is, if $D_i > D_e$ for $x = 1$. This can happen only for $q > c/b$, that is, if the probability of knowing the image of the co-player exceeds the cost-to-benefit ratio, and if the average number of rounds, that is, $1/(1-w)$, exceeds $(bq + c)/(bq - c)$. Note that for our numerical example of Figs 1 and 2, where $b = 1$ $c = 0.1$ and $q = 1$, we need only about 1.2 rounds per generation for cooperation to be stable against invasion by defectors.

The good, the bad and the discriminating. Indirect reciprocity only works when donors discriminate between individuals that have or have not helped others in the past. To understand the role of indiscriminate altruists, we add to the population of defectors and discriminators a fraction z of cooperators, who always give help irrespective of the their co-player's score. We can calculate the payoffs in each round as before. The cooperators' total expected payoff, D_c, differs from that of the defectors, D_e, by $[-c + (bwqx)/(2-w)]/[2(1-w)]$, whereas

$$D_i - D_e = \frac{(bqx - c)(1 - q + qz)}{2(1-w)(1-qx)} - \frac{bq(x+y)}{2-w} + \frac{qy(b-c)}{(1-qx)(2-w-wqx)}$$

The population is in equilibrium whenever $y = 0$ (no defectors) or $x = c(2-w)/bwq$. If x lies below the latter value, defectors win; if x exceeds it, then a mixture of discriminating and indiscriminating altruists gets established, depending on the initial value. This mixed state is proof against invasion by unconditional defectors, but in such a population both discriminating and indiscriminating altruists do equally well. Their frequencies will be altered only by random drift, not by selection. If the frequency of discriminators falls below $c(2-w)/bwq$ then defectors can invade and take over. Defectors in turn can be overcome by discriminators if their frequency fluctuates above $x_{\min}$.

A universal constant of nature. Let us now consider a situation in which the image score can be any integer number between $-\infty$ and $+\infty$, but in which all players adopt the same strategy, $k = 0$. Denote by x_i the frequency of players with image score i. In the next round it is $x'_i = [x_i + x_{i-1}\phi + x_{i+1}(1-\phi)]/2$ where $\phi = \Sigma_{i=0}^{\infty} x_i$. If all players start with an image score of greater than or equal to 0, then all players will cooperate in the first and all subsequent rounds. If all players start with an image score of less than 0, then all players will defect in the first and all subsequent rounds. The situation becomes interesting if there is an initial distribution of image scores above and below 0. The question of whether the system will ultimately converge to cooperation or defection is non-trivial. We find that there is a maximum fraction of players with an initial image score below 0, such that the system ultimately converges to all-out cooperation. Numerical simulations show that this fraction is 0.7380294688360....

Received 11 November 1997; accepted 31 March 1998.

1. Hamilton, W. D. The evolution of altruistic behaviour. *Am. Nat.* **97,** 354–356 (1963).
2. Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7,** 1–16 (1964).
3. Williams, G. C. *Group Selection* (Aldine–Atherton, Chicago, 1971).
4. Eshel, I. On the neighbourhood effect and evolution of altruistic traits. *Theor. Popul. Biol.* **3,** 258–277 (1972).
5. Wilson, D. S. & Sober, E. Reintroducing group selection to the human behavioural sciences. *Behav. Brain Sci.* **17,** 585–654 (1994).
6. Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46,** 35–57 (1971).
7. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211,** 1390 (1981).
8. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
9. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359,** 826–829 (1992).

10. Michod, R. E. & Sanderson, M. J. in *Evolution: Essays in Honor of John Maynard Smith* (eds Greenwood, P. J., Harvey, P. & Slatkin, M.) 95–106 (Cambridge Univ. Press, Cambridge, 1985).
11. Peck, J. & Feldman, M. The evolution of helping in large, randomly mixed populations. *Am. Nat.* **127,** 209–221 (1985).
12. Milinski, M. Tit for tat in sticklebacks and the evolution of cooperation. *Nature* **325,** 433–435 (1987).
13. May, R. M. More evolution of cooperation. *Nature* **327,** 15–17 (1987).
14. Dugatkin, L. A. Mesterton-Gibbons, M. & Houston, A. I. Beyond the prisoner's dilemma: towards models to discriminate among mechanism of cooperation in nature. *Trends Ecol. Evol.* **7,** 202–205 (1992).
15. Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355,** 250–253 (1992).
16. Nowak, M. A. & Sigmund, K. Win–stay, lose–shift outperforms tit for tat. *Nature* **364,** 56–58 (1993).
17. Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
18. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
19. Wilson, E. O. *Sociobiology* (Harvard Univ. Press, Cambridge, MA, 1975).
20. Krebs, J. R. & Davies, N. B. *An Introduction to Behavioural Ecology* (Blackwell, Oxford, 1987).
21. Buss, L. *The Evolution of Individuality* (Princeton Univ. Press, NJ, 1987).
22. Frank, S. A. The origin of synergistic symbiosis. *J. Theor. Biol.* **176,** 403–410 (1995).
23. Binmore, K. G. *Fun and Games: a Text on Game Theory* (Heath, Lexington, MA, 1992).
24. Marler, P. & Evans, C. Bird calls: just emotional displays or something more? *Ibis* **138,** 26–33 (1996).
25. Zahavi, A. & Zahavi, A. *The Handicap Principle: A Missing Piece of Darwin's Puzzle* (Oxford Univ. Press, Oxford, 1997).
26. Boyd, R. & Richerson, P. J. The evolution of indirect reciprocity. *Social Networks* **11,** 213–236 (1989).
27. Sudgen, R. *The Economics of Rights, Cooperation and Welfare* (Blackwell, Oxford, 1986).
28. Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* (submitted).
29. Pollock, G. B. & Dugatkin, L. A. Reciprocity and the evolution of reputation. *J. Theor. Biol.* **159,** 25–37 (1992).
30. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).

Acknowledgements. We thank M. Dawkins, A. Kacelnik, J. Krebs and R. May for discussion. Support from the Wellcome Trust is gratefully acknowledged. Part of this work was done at IIASA (Laxenburg).

Correspondence and requests for materials should be addressed to M.A.N. (e-mail: martin.nowak@zoo.ox.ac.uk).

Selective representation of relevant information by neurons in the primate prefrontal cortex

Gregor Rainer, Wael F. Asaad & Earl K. Miller

Department of Brain and Cognitive Sciences and The Center for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The severe limitation of the capacity of working memory, the ability to store temporarily and manipulate information[1], necessitates mechanisms that restrict access to it. Here we report tests to discover whether the activity of neurons in the prefrontal (PF) cortex, the putative neural correlate of working memory[2–8], might reflect these mechanisms and preferentially represent behaviourally relevant information. Monkeys performed a 'delayed-matching-to-sample' task with an array of three objects. Only one of the objects in the array was relevant for task performance and the monkeys needed to find that object (the target) and remember its location. For many PF neurons, activity to physically identical arrays varied with the target location; the location of the non-target objects had little or no influence on activity. Information about the target location was present in activity as early as 140 ms after array onset. Also, information about which object was the target was reflected in the sustained activity of many PF neurons. These results suggest that the prefrontal cortex is involved in selecting and maintaining behaviourally relevant information.

In the 'array trials', a sample array of three objects was briefly presented while the monkeys maintained central gaze (Fig. 1a). Monkeys needed to find the target object in the array and remember its location. After a brief delay, a test array appeared and the monkeys had to release a lever if the target object appeared in the same location as it had in the sample array. Although each of the three objects was a target, in turn, for a block of trials, its location in the sample array was chosen randomly on each trial. Monkeys were cued to the target object with 'cue trials' (Fig. 1a) in which the target object appeared alone.

Table 1 Summary of neuronal selectivity in different task periods

	Sample period	Delay period	Both periods
$n = 97$ cells			
Number of cells selective for:			
Target object only	20	16	7
Target location only	22	30	13
Target object and location	24	15	8
Total selective for object	44	31	15
Total selective for location	46	45	21
Selectivity depth:			
Object	48%	44%	–
Location	48%	53%	–
Selectivity index:			
Object	0.24	0.24	–
Location	0.24	0.28	–

Cell counts are based on ANOVA (see Methods), evaluated at $P < 0.01$. Mean selectivity depths and selectivity indices were computed from delay activity on array trials for cells showing a significant ANOVA. For cells not showing significant effects, mean selectivity depths ranged from 12 to 15% and mean selectivity indices ranged from 0.08 to 0.09.

We recorded the activity of 97 neurons from the lateral prefrontal cortex of two monkeys (Fig. 1b). Based on analysis of variance (ANOVAs) (evaluated at $P < 0.01$), many PF neurons showed activity to physically identical sample arrays that varied depending on which of three array positions contained the target (46/97 or 47% during sample presentation, 45/97 or 46% during the delay, Fig. 2 and Table 1). Information about the target location appeared very early in neural activity, starting about 140 ms after array onset (Fig. 3a). The activity of these neurons after this time largely reflected the target location alone; information about the location of the irrelevant, non-target, objects had little or no influence. Although almost half of PF cells showed activity that varied with the location of the target object, only a few cells (sample period: 10/97 or 10%; delay period: 5/97 or 5%) showed activity that varied with the location of non-target objects (*t*-tests, evaluated at $P < 0.01$). In fact, many cells showed similar activity on array trials and on cue trials in which the target object appeared alone (Fig. 2).

The task also required monkeys to remember which object was currently the target. This was also reflected in PF activity. On array trials, many PF neurons showed activity during the sample period

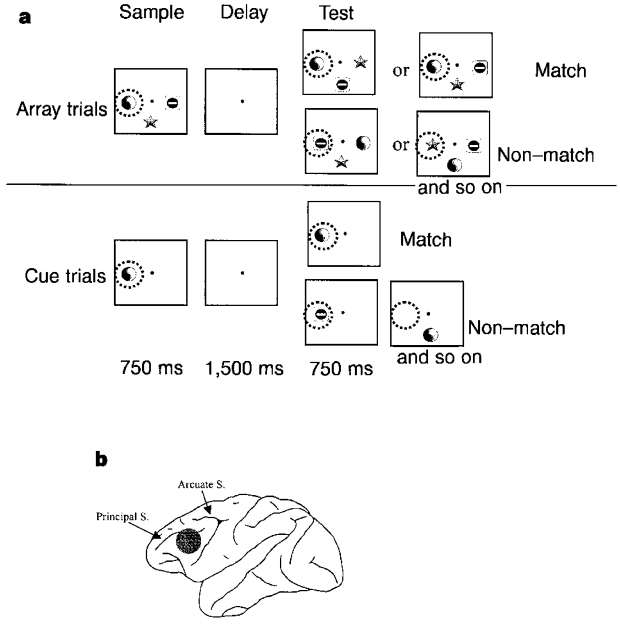

Figure 1 The behavioural task and recording sites. **a**, Sequence of trial events. Each trial began when the monkey grasped a lever and fixated a small fixation target at the centre of a computer screen. The location of the target object is indicated by the dotted circle on this figure. Examples of array trials (top) and cue trials (bottom) are illustrated. **b**, Location of recording sites: Arcuate S, arcuate sulcus; Principal S, principal sulcus.