**Group Extinction in Iterated Two Person Games with Evolved Group-Level Mixed Strategies**

**Abstract**

The shift to a genetic basis of evolution in the 1960s, and away from group selection, created a problem in regard to the origin of cooperative behaviour in human societies. The resolution essentially involves mutual recognition of individuals, thus permitting the phenomena of reputation, reciprocation and retribution to arise, these being key to stable cooperative societies. The analysis presented, based on evolutionary game theory, serves to emphasise the crucial role of individual recognition by illustrating the consequences of assuming the opposite. It is shown that where tribal membership is apparent, but individuals are not recognisable, evolving mistrust leads to tribal extinction in an evolutionary game theory model. Moreover, a single tribe is also unstable to schism. Subsequently, extinction of one schismatic group occurs. Failure to recognise individuals therefore facilitates a mechanism which leads to increasing conformity.

**Keywords**: individual, iterated games, two person games, evolutionary game theory, group extinction, schism

**Cooperation and Evolution**

Human societies are unique in forming extremely large social groups which are generally harmonious and cooperative, despite no close genetic relationship between most pairs of individuals. The depressing Hobbesian view was that such cooperation arises only because it is enforced by a powerful Government. Before such Government, Hobbes claimed, life was “solitary, poor, nasty, brutish and short” (Hobbes, 1651). A rather more uplifting perspective is that humans are, either by nature or by social evolution, inclined to be moral, acting with consideration towards others even when it is to their own detriment. That view certainly applies, to some degree at least, when the individuals in question are closely related. In such cases the altruism of individuals arises from the selfishness of their genes. It is rather harder to understand altruism emerging between unrelated individuals. However, under favourable conditions, considerate behaviour might emerge and survive where there is a reasonable expectation of reciprocation; where cooperation is a matter of enlightened self-interest. This is reciprocal altruism.

The relevance of game theory to the phenomenon of reciprocal altruism goes back at least to (Robert Trivers, 1971), who credited W.D.Hamilton. The historical neglect, before this time, of the problem of how cooperative behaviour arises has a bearing on the present subject matter. (Axelrod and Hamilton, 1981) express it as follows: “*Before about 1960, accounts of the evolutionary process largely dismissed cooperative phenomena as not requiring special attention. This position followed from a misreading of theory that assigned most adaptation to selection at the level of populations or whole species. As a result of such misreading, cooperation was always considered adaptive. Recent reviews of the evolutionary process, however, have shown no sound basis for a pervasive group-benefit view of selection; at the level of a species or a population, the processes of selection are weak. The original individualistic emphasis of Darwin's theory is more valid.*”

In other words, the shift in evolutionary theory in favour of gene-based selection, rather than group selection, created a new problem: how altruism arises between genetically distant individuals even in the same species. Attention was thus drawn to the relevance of models such as evolutionary game theory, (Maynard Smith and Price, 1973; Maynard Smith, 1974). This history is apposite because it concerns the distinction between the role of the individual and the role of the group, as does the present work.

Since the 1970s, both theoretical and experimental studies of simple games have continued to be a rich source of insight into cooperative or other social behaviours, for example: the evolution of fairness (Yen-Sheng Chiang, 2007); the emergence of conventions (Centola and Baronchelli, 2015); the interplay of trust and social capital (Frey, Buskens & Raub, 2015); how inequality influences cooperative behaviour (Aksoy and Weesie, 2009, 2013); how strong social ties may mitigate against the common good (Flache, 2002); the behaviour of the iterated Prisoners’ Dilemma with incomplete information (Dijkstra and van Assen, 2017); the relevance of game theory in cognitive science (Camerer, 2003a, 2003b; Camerer and Smith, 2012), and the issue of trust in the presence of inequality (Sato, 2002), and many more.

Ever since Axelrod’s computer tournaments of competing Prisoners’ Dilemma strategies it has been appreciated that reciprocal altruism can arise through enlightened self-interest under suitable circumstances, simple strategies like tit-for-tat often proving remarkably effective, (Axelrod and Hamilton, 1981; Axelrod, 1984). This hinges crucially on the prevalence of pairs of individuals who interact repeatedly, who are mutually recognised by each other, and whose past encounters are remembered (or, at least, the last encounter is remembered).

Of course, several problems remain which prevent this alone being a satisfactory explanation of the origin of reciprocal altruism. One problem is how a tit-for-tat cooperative strategy could gain a foothold in an original population of belligerent defectors. There are possible answers to this. The cooperative strategy might arise initially only locally in clusters, perhaps involving the boosting effect of kinship selection. Clusters arise naturally due to geographical proximity, e.g., (Perc, 2008), or the action of social networks, e.g., (Roca, Sanchez and Cuesta, 2012; Rezaei, Kirley, and Pfau, 2009). However, some authors see a difficulty with cooperation arising in large clusters, (Boyd and Richerson, 1988; Manhart, 1989).

There is general recognition that the key features necessary for the evolutionary stability of cooperative behaviours are the ‘three Rs’: reputation, reciprocation and retribution. All three appear to be required. Tit-for-tat illustrates the role of reputation as a condition for reciprocation. Reward and punishment alone appear insufficient to induce prosocial behaviour; reputation is also required, (Sigmund, Hauert & Nowak, 2001).

Reciprocation may be direct, in the sense that A favouring B is reciprocated by B favouring A; or it may be indirect, in the sense that A favours B, who favours C, who favours D, and so on, with A hoping to be favoured in the end, (Alexander, 1987; Boyd and Richerson, 1989),

Retribution is important to supress the rampant spread of free-loaders. It may be accomplished through policing, (Frank, 1995, 1996, 1998), or through punishment. In the latter context, (Gardner & West, 2004) observe that if punishment is sufﬁciently frequent and harsh, it can successfully maintain cooperative behaviour. However, for cooperative behaviour to arise, “*the crucial factor is a positive correlation between the punishment strategy of an individual and the cooperation it receives*”. (Marlowe et al, 2008) have emphasised the relevance of “third party” punishment, defined as punishing someone who defected on someone else, especially in large and complex societies.

However, (Boyd and Richerson, 1992, 2009) observe that sufficient punishment can stabilise any behaviour, not just that which is socially beneficial: “*If everybody agrees that individuals must do X, and punish those who do not do X, then X will be evolutionarily stable as long as the costs of being punished exceed the costs of doing X. It is irrelevant whether X benefits the group or is socially destructive*”. These authors opine that socially cooperative, rather than socially destructive, behaviours arise due to cultural adaption via competition between groups. From this perspective, such cultural adaptation is seen as the origin of “*moral systems enforced by systems of sanctions and rewards*” which act upon the individual through evolved social emotions and behaviours, like shaming and guilt. The relevance of these evolved social proclivities to reputation, reciprocation and retribution is clear.

There is a danger, amongst all this detailed theorising, of failing to see the wood for the trees. Despite the fact that the phenomenon to be explained is widespread socially cooperative behaviour, all the mechanisms by which this may be brought about involve behaviours enacted at the individual level. A cooperative society is the aggregate of individuals inclined to be cooperative. The key factors of reputation and reciprocation relate irreducibly to individual behaviour, and retribution is enacted upon the individual for that individual’s failure to conform to societal norms. A population consisting of N/2 committed co-operators and N/2 committed defectors is profoundly different from a population of N individuals each of whom randomly chooses between cooperation and defection on each occasion. Moreover, even if a population consists of N/2 committed co-operators and N/2 committed defectors, if they are indistinguishable as individuals the situation becomes functionally identical to a population of N individuals who choose randomly between cooperation and defection, and the potential benefit of cooperation by half the population may be lost. Recognition of the individual is essential to the operation of reputation, and hence the subsequent reciprocation or retribution, as appropriate.

The purpose of the present work is to illustrate and emphasise the importance of recognising the individual by considering what transpires when one does not.

In this work a population is assumed to consist of two tribes, the individual members of which are indistinguishable; only tribal membership is apparent. Individuals interact via an iterated two-player game, i.e., a large number of plays is considered so that expectation values of utility can be assumed to arise. The relative prevalence of the two tribes in the overall population is allowed to evolve in accord with their average game payoff (fitness or utility). Stochastic (“mixed”) strategies can produce evolutionarily stable equilibrium populations of the two tribes for suitable choices of the parameters. The stochastic tendency to defect, rather than cooperate, defines “mistrust” parameters which characterise the tribes. A secondary evolution of these mistrust parameters is then considered. It is shown that the result is always the extinction of one of the two tribes.

**Two-Player Games with Group Level Cooperation-Dilemmas**

A conventional two player game is considered. Each individual actor has the choice between binary options, labelled and . The payoff, or utility function, is defined as usual by a matrix,

(1)

For example, the first actor playing and the second actor playing results in a payoff to the first actor of and a payoff to the second actor of . We consider only symmetrical games in which it makes no difference if an actor is regarded as the first or second player. In contrast with some forms of game theory it is assumed that the actors may communicate prior to play, though there is no guarantee that they will not subsequently be treacherous.

Each actor will be assumed to belong to one of two tribes. Individual actors are assumed to have no ability to recognise other individual actors but can identify the tribe to which an individual belongs. It is not necessary to assume that the play strategies adopted by any two members of the same tribe are statistically identical. However, any difference in play strategy by two members of the same tribe is not observable because the two actors cannot be distinguished. An opponent cannot perceive any differences in individual play strategy; only the stochastic behaviour of the tribe as a whole can be perceived, i.e., the probability of cooperation or defection across the whole tribe. This is a crucial distinction. It is not assumed that individual members of a given tribe are (statistically) identical in their play, only that individuals are unrecognisable so that their different strategies would have no impact on others’ behaviour.

A tribe’s play strategy is defined by the probabilities of cooperating and defecting against a given opposing tribe, these probabilities being averages over the whole tribe.

The simple model will assume that all individuals play all other individuals an equal number of times, and that declining to play is not an option. More sophisticated models might restrict which actors can interact with other actors based on spatial proximity or some form of social network, but this is not considered here, nor is declining to play.

The payoff parameters in (1) may be positive or negative. Qualitatively different behaviours result according to the relative magnitudes of these parameters. Some cases are trivial and hence not of interest. For example, if parameter is larger than all of and then both actors will always play and there is nothing of interest to say. Similarly, if is the largest parameter then both actors will always play and the game is trivial. That leaves 12 combinations of possibilities, since the largest parameter must be either or , leaving three ways of choosing the second largest, and two ways of choosing the smallest (2 x 3 x 2 = 12).

However, without loss of generality we can take since the relative size of and merely distinguishes between the two plays, and (i.e., swapping and and also swapping and results in an identical game). Consequently, we can take to be the largest parameter, and hence just 6 possibilities remain,

Case 1:

Case 2:

Case 3:

Case 4:

Case 5:

Case 6:

In the discussion of each case which follows it is important to bear in mind our intended application, namely that, (i) we are interested only in ‘iterated games’, i.e., a large number of plays, (ii) it is assumed that the two tribes can communicate prior to play to coordinate their strategy, though recall that individuals cannot be recognised, only tribal membership is apparent. We shall be particularly interested in situations where there is a mutual benefit to a cooperative strategy, but an actor may gain short-term benefit by reneging on the agreed strategy in a given play. Our discussion will differ from classical game theory, which focuses on single plays without prior communication, and in which dilemmas can arise simply from lack of communication. In our case, dilemmas arise from the balance of trust and mistrust.

Case 1 () is the Prisoners’ Dilemma. Whatever one actor plays, the other actor is better off playing . Hence both actors might rationally be led to play if there were no cooperation between them. But both would be better off if both played . However, this requires trust because a player treacherously going against an agreement to play will benefit, at least in the immediate term, since he thereby achieves the greatest payoff, . This is their dilemma: whether it is better to cooperate in order to increase benefit but to leave oneself vulnerable to treachery - or not? Note that in this case we can describe as cooperation and as reneging or defection.

In the case there is another cooperative strategy which individually recognisable actors could adopt which would be more lucrative than both playing on every occasion. This is for the first actor to play while the other plays . Both players score per play on average. This strategy is also prone to treachery. The actor whose turn it is to play can do better in that round by reneging and playing instead, because . (Axelrod, 1981) vetoed this case, defining the Prisoners’ Dilemma as requiring , but this is just semantics. However, this strategy requires coordinated play between two individual actors, which cannot arise when actors are selected at random and individuals are not recognisable. This type of cooperation cannot arise when only tribal identity is recognisable.

Case 2 () and Case 3 () are, like the Prisoners’ Dilemma, such that whatever one actor plays, the other actor is better off playing . Hence both actors might rationally be led to play if there were no cooperation between them. Unlike the Prisoners’ Dilemma, there is no advantage in both playing because , so this is not a beneficial cooperative strategy. However, in the case there would be a mutually beneficial cooperative strategy if individuals were recognisable, namely for the first actor to play while the other plays , thus scoring per play on average. This strategy is again prone to treachery. The actor whose turn it is to play can do better in that round by reneging and playing instead, because . But, as noted above, this type of cooperative play cannot arise when only tribal identity is recognisable.

Case 4 () and Case 5 () are often regarded as leading to dilemmas also. But this is only true if there is no prior communication between the actors. If players may agree a strategy in advance, the optimal coordinated strategy for individually recognisable actors is for one to play while the other plays . There is no temptation to deviate from this agreement since it would be immediately disbeneficial to the actor who does so. Thus, there is no dilemma, and, in any case, this type of cooperative play cannot arise when only tribal identity is recognisable.

Case 6 () is the Game of Chicken: Whatever one actor plays, the other actor is better off playing the opposite. With prior agreement, and if individual actors were recognisable, the alternating strategy where one actor plays while the other plays would be most beneficial if . Moreover, this strategy is not prone to treachery because there is no gain for an actor who deviates from the agreed alternating play, so there is no dilemma. But again, this type of cooperation cannot arise when only tribal identity is recognisable. However, if a more beneficial strategy for both actors is that they both play , always – and this strategy is equally applicable to unrecognisable individuals. This strategy is prone to treachery because an actor playing will gain an immediate reward because . Note that in this case we can again describe as cooperation and as reneging or defection.

In summary, where only tribal identities but not individuals are recognisable, the cases of greatest interest - where there is a cooperative strategy which is subject to potential treachery (i.e., there is a cooperation dilemma) - are Case 1 (Prisoners’ Dilemma) and Case 6 (Game of Chicken).

**Formulation of the Stochastic Strategy for Two Tribes**

Because individuals are assumed unrecognisable, strategies based on past experience with an individual (e.g., tit-for-tat) are not available. Allowing for different plays by individual actors in a given tribe requires tribal-level strategy to be stochastic (or a ‘mixed strategy’ in the usual game-theory terminology). Recall that we are chiefly interested in the Prisoners’ Dilemma and the Game of Chicken for which we have argued above that both actors playing may be regarded as cooperation, whilst one actor playing instead can be regarded as reneging on this cooperation. Hence, we shall refer to playing as reneging, or treachery. Playing is engendered by mutual trust, whereas playing tends to suggest (or to give rise to) mistrust.

Thus, the probability of a randomly chosen individual in tribe 1 ‘reneging’ by playing when playing against an individual in tribe 2 is defined as , whereas the probability of a randomly chosen individual in tribe 2 ‘reneging’ by playing when playing against an individual in tribe 1 is . The corresponding probabilities of playing (cooperating) are, of course, and . Individuals may also cooperate or renege against members of their own tribe. Thus, the probability of an individual in tribe 1 ‘reneging’ by playing when playing against an individual in the same tribe is , whereas the probability of an individual in tribe 2 ‘reneging’ by playing when playing against an individual in the same tribe is .

This is written succinctly as a “mistrust matrix”, i.e., the probabilities of playing ,

(2)

Suppose that, at some given time, a fraction of the population consists of tribe 1 (and so a faction of tribe 2). Writing the average payoff, or fitness, across tribes 1 and 2 as and we have, from (1) and (2),

(3a)

(3b)

where, (4)

(5)

(6)

(7)

Hence, in terms of the parameters which define the ‘game’, and the parameters which define the play strategy (the mistrust parameters), , the average payoff to each tribe can be found from (3a,b) if the current make-up of the population is known, i.e., the fraction, , which is tribe 1.

**Evolution of a Stable Population**

In an evolutionary scenario, it is assumed that population change of a tribe is related to the tribe’s average game payoff (fitness). By this means an equilibrium population fraction for the two tribes arises for suitable parameter values, assuming play strategy defined by the “mistrust” matrix, (2), is fixed.

The evolution of the population defines how changes with time and how may achieve an equilibrium value. This evolution can be formulated in a number of different ways. A discrete dynamic may be defined using a discrete (integer) time variable, and assuming that the entire population at time dies and is completely replaced by an offspring population at time . The number of individuals in a given tribe in the ‘offspring’ generation may be taken as proportional to the average payoff of that tribe in the ‘parental’ generation. Here we opt instead for a continuum time formulation and assume that the reproduction *rate* is proportional to the average payoff. In this model, the parental generation does not die off suddenly, to be completely replaced by the next generation. Instead there is a smooth differential evolution. Thus, if is the number of individuals in tribe 1 in generation , and is the average payoff to tribe 1 in generation , then the model evolves the population of tribe 1 in a small increment of time,, as,

(8)

and a similar expression for tribe 2. The continuum time formulation has some technical advantages over the discrete time formulation, for example equilibrium and stability are more easily analysed (below). Another advantage of (8) is that it remains valid when the average payoffs, and , are negative, so that we can allow the game parameters, , to be negative without worrying if this results in negative average payoffs. Indeed, a negative payoff is required to cause a population to reduce.

The prevalence of tribe 1 is expressed as a fraction, , of the total population, , i.e., . The evolution equation, (8), can be expressed in terms of this population fraction, thus,

(9)

In the differential limit, and , this gives,

(10)

The evolution of from some arbitrary initial value, say , is obtained by time-stepping integration of equ.(10), substituting equs.(3a,b) for and .

An observation is appropriate here regarding the range of values that may be taken by and the relevance of the magnitude, and sign, of the expected payoff. The absolute magnitude of the payoff has no meaning, and the absolute changes in a population given by equ.(8) can be scaled as desired (if necessary by scaling the arbitrary time variable). Consequently, because it has been shown that can be taken to be the largest of the parameters, we can set in all numerical calculations. In conventional game theory, a positive affine transformation can be defined which transforms all the parameters to the range [0,1] and leaves the game invariant. However, such a transformation does not leave the payoff itself invariant, and hence it will alter the rate of change of the populations of the tribes according to equ.(8). In particular, if all parameters are in the range [0,1], the payoff can only be positive and hence the populations will monotonically increase, never decrease. Consequently, equ.(8) breaks affine invariance. However, equ.(10) depends only upon the difference of the payoffs to the two tribes, and hence can be negative even if the payoffs themselves are positive. Combined with the degree of freedom implicit in scaling the time variable, if we confine attention to the evolution of the population *fraction*, , then positive affine invariance is restored. Nevertheless, we shall not adopt the transformed parameter range [0,1] here, preferring to retain the possibility of negative parameters, and possibly negative payoffs, in deference to the lack of affine invariance of equ.(8).

Evolutionary equilibrium is obtained when the population fraction, , reaches some value which no longer changes, i.e., when (10) is zero. Hence an equilibrium occurring with both tribes still extant, and hence with , must have,

Equilibrium: (11)

We note in passing that the discrete time formulation produces the same equilibrium condition, equ.(11), although the stability condition is different. Inserting equs.(3a,b) into equ.(11) and solving for provides the equilibrium population fraction of tribe 1,

(12)

if this lies within the meaningful range with both tribes extant, i.e., . Otherwise evolution of the population results in one or other of the tribes becoming extinct, i.e., if (12) is less than zero then the final population has (tribe 1 becomes extinct), or, if (12) exceeds unity then the final population has (tribe 2 becomes extinct).

When both tribes are present at equilibrium, the requirement that together with equ.(12) leads to,

Either: and (13a)

Or: and (13b)

The requirement that the equilibrium be stable allows option (13a) to be rejected, and (13b) is found to be necessary and sufficient for stability. This is established as follows. Equ.(9) can be written in the form,

(14)

Suppose we consider a perturbation from the equilibrium distribution defined by . Now allow the population to evolve. Denoting the population fraction after the iteration to be , and writing its difference from the equilibrium value as , we have, to first order in the Taylor series,

(15)

noting that by definition of the equilibrium condition. Hence, (16)

Convergence back to the equilibrium distribution requires that as , so that we conclude that the necessary and sufficient condition for stability is,

(17)

But using equ.(9) for we get,

(18)

In the differential limit the left-hand inequality of (17) is automatic whilst the right-hand inequality becomes,

(19)

But equs.(3a,b) give and so that (19) gives . Hence we see that (13a) is ruled out and (13b) becomes the necessary and sufficient condition for a stable equilibrium with both tribes extant. Solutions of equ.(11) with and , i.e., unstable equilibria, with initial , will evolve so as to make one tribe or the other extinct.

In passing we note that the basin of attraction for stable equilibria is the whole open interval, . This follows because, as a result of (3a,b) and (13b), the time derivative (10) will be positive if and negative if , so for the whole parameter range the evolution will be in the direction towards the equilibrium point. This is illustrated by Figure 1.

We shall be interested here only in combinations of game parameters and mistrust parameters such that there is a stable equilibrium with both tribes extant.

**Figure 1:** Illustrating the evolution of the population, converging on an equilibrium population-fraction of tribe 1 of (for )



**Evolution of Mistrust**

Thus far the parameters in the payoff matrix, , and the parameters in the mistrust matrix, , have been assumed fixed and given by fiat. Only the population ratio, , has been subject to evolution, whilst these other parameters were held constant.

We will continue to assume that the parameters in the payoff matrix, , are fixed constants.

But what if the mistrust parameters could also evolve? It is natural to assume they would. The trust or mistrust an individual feels for another individual is something that can change according to experience. This is the basis of strategies like tit-for-tat when individuals are recognisable and prior outcomes remembered. The analogue in this case, where individuals are not recognisable, is that tribes as a whole will be prepared to increase or decrease their mistrust if doing so leads to an increased payoff.

Increasing the average payoff to a tribe, other things being equal, will increase the proportion of that tribe in the population. Consequently we can use the population fraction of the tribe as a surrogate for the payoff itself as a guide to the propensity to change the level of mistrust.

The formulation deployed here can be interpreted as arising when the evolution of mistrust is sufficiently slow that the population fractions are always in equilibrium. Thus, for a two-tribe population, it will always be the case that despite these quantities changing as the mistrust evolves. At every instant, the population fraction is given by the equilibrium value, equ.(12), although this value will be evolving as the mistrust parameters evolve, by virtue of the dependences of on the mistrust parameters, as given by equs.(4-7).

However it is important to appreciate that a slow evolution of mistrust is not essential. The results apply equally to the case where there is a finite jump in mistrust parameters followed by a ‘relaxation’ to the corresponding equilibrium population. The evolution may progress by a sequence of jumps in mistrust parameters and “punctuated equilibria”, as illustrated by Figure 2.

**Figure 2:** Comparing the algorithm for mistrust evolution based on following the equilibrium population curve via its tangents with finite mistrust increments and punctuated equilibria (illustrated for a single mistrust parameter)



It is the “desire” of tribe 1 to increase which drives the evolution of mistrust. However, it is equally true that the population evolves in response to the “desire” of tribe 2 to increase , and hence to decrease .

What matters, therefore, is the dependence of on the mistrust parameters, . Recall that is the intra-tribal mistrust of tribe 1, whilst is the mistrust of tribe 1 for tribe 2. These are the two parameters under the control of tribe 1. Tribe 2, however, has no control over these parameters – they are defined by the behaviour of tribe 1, namely by the relative frequency with which members of tribe 1 play against a given opponent. Since tribe 1 wants to increase their population fraction, tribe 1 will wish to vary and so as to increase .

Noting that the derivatives of and with respect to are readily found from equs.(4,5), it follows from equ.(12) that,

(20)

where . Similarly we find,

(21)

If (20) is a positive quantity, tribe 1 will tend to increase their mistrust of tribe 2 () so as to increase their prevalence in the population, . If it were negative, however, then tribe 1 would tend to decrease their mistrust of tribe 2 so as to increase . Similar remarks apply to (21) in respect of changes in tribe 1’s intra-tribe mistrust, .

In contrast, tribe 2 has control only over their own mistrust, i.e., parameters and . The corresponding derivatives are,

(22)

(23)

Tribe 2 wishes to maximise its population fraction, which is . So tribe 2 will tend to act so as to *decrease* . Thus, if (22) is negative, tribe 2 will tend to increase its mistrust of tribe 1, i.e., . But if (22) is positive, tribe 2 will tend to decrease . Similar remarks apply to (23) in the context of tribe 2’s intra-tribe mistrust, .

Consider small changes in the mistrust parameters: . The population changes by,

(24)

Is it possible for the population to reach a stable equilibrium with respect to evolution of the mistrust parameters? This requires (24) to vanish for arbitrary , and hence for mistrust parameters to exist such that all four derivatives, equs.(20-23), are zero. Moreover, to be a stable equilibrium, the second derivatives of with respect to the mistrust parameters would need to satisfy certain inequalities.

Consequently it can be concluded immediately that there is no equilibrium with respect to the evolution of mistrust in the case of the Prisoners’ Dilemma, which has , because, by virtue also of (13b), the derivative (20) is positive definite whilst the derivative (22) is negative definite. The mutual mistrusts of the two tribes, and , will therefore relentlessly increase: there is no equilibrium. The evolution of mistrust can terminate only when mutual mistrust becomes maximal, , or when one or other of the tribes becomes extinct ( or ). We shall see that the latter condition is invariably what occurs.

Which tribe becomes extinct is not determined by the analysis so far. If tribe 1 decides upon a much more rapid escalation in their mistrust of tribe 2 than the reverse, then this belligerent action will drive eradication of tribe 2 (and vice-versa).

This ambiguity is because the above analysis establishes only the *direction* of change in the mistrust parameters , but does not establish the relative magnitude of the increments in these parameters. At this point it is necessary to introduce some assumption for the relative magnitudes of these mistrust-increments in order to be able to track the trajectory through mistrust parameter space, and ultimately to conclude which tribe becomes extinct. However, it is emphasised that the conclusion regarding tribal extinction is inevitable, whatever path through mistrust parameter space is taken. It is only *which* tribe becomes extinct which may be varied. The only way in which tribal extinction could be avoided would be for the two tribes to coordinate their mistrust increments so as to ensure that (24) summed to zero, but such a thing would be against their interests.

The most reasonable hypothesis is that the magnitude of the increment in a given parameter is proportional to the derivative of with respect to that parameter. Hence, if is particularly sensitive to a given parameter, that parameter will tend to change faster than others. In mathematical language, the rate of change of is proportional to its vector gradient in space, and the same is true in space but in the reversed direction. Hence, in some small time increment we shall assume an evolution increment in the mistrust parameters given by,

(25)

By time-stepping integration of (25) the trajectory of the evolution of the mistrust parameters, , and the corresponding evolution of is obtained. If any of the mistrust parameters reaches 0 or 1 then they are subsequently held fixed. One need only be careful that the time steps, , are sufficiently small to obtain converged behaviour.

**Tribal Extinction**

Numerical integrations to determine the evolution of mistrust and the evolution of the population, , were carried out using the prescription (25). The parameter was set to unity in all cases, without loss of generality since the absolute magnitude of the payoff has no meaning. Figure 3 shows the results for an example of the Prisoners’ Dilemma, defined by , for the specific case , and with initial mistrust parameters . Figure 3 shows how the mistrust parameters increase and the corresponding population fraction of tribe 1, , declines until it reaches zero. Hence this case results in the extinction of tribe 1.

Figure 4 is another Prisoners’ Dilemma case, this time resulting in the extinction of tribe 2.

In total some 605 different combinations of parameters were explored for the Prisoners’ Dilemma, i.e., for , as specified in Figure 5. For each one of these combinations, 10,000 possible combinations of initial mistrust parameters were explored, each of spanning the range 0 to 0.92 in steps of 0.1 (avoiding parameters being exactly equal). Hence, 6,050,000 Prisoners’ Dilemma cases were analysed. Many of these combinations did not correspond to initially stable equilibria (with respect to population evolution), i.e., either the resulting from (12) did not lie in the physical range, , or (13b) was not respected. These cases were ignored. All cases for which there was a valid stable equilibrium initially (namely 619,394 cases, or 10.2%) were subject to mistrust evolution according to prescription (25). Every such case terminated in the extinction of either tribe 1 or tribe 2, qualitatively similar to Figure 3 or Figure 4.

**Figure 3:** Illustrating the evolution of mistrust, resulting in the extinction of tribe 1 (for a Prisoners’ Dilemma case with )



**Figure 4:** Illustrating the evolution of mistrust, resulting in the extinction of tribe 2 (for a Prisoners’ Dilemma case with )



**Figure 5:** Combinations of parameters explored numerically for Prisoners’ Dilemma cases (with



The exercise was repeated for the Game of Chicken, i.e., for cases with , analysing 363 such cases as defined by Figure 6. Again 10,000 combinations of mistrust parameters, , were used for each, making 3,630,000 cases in all. The number of cases which started with parameters producing a stable equilibrium with respect to population evolution was 1,288,821 (35.5%). Under subsequent evolution of mistrust, all such cases terminated in the extinction of one of the two tribes. Figure 7 shows the evolution of the parameters in one example case.

Finally, Figure 8 defines 320 combinations of with , which were analysed together with 10,000 combinations of mistrust parameters. Once again, all cases which started with a stable equilibrium terminated in the extinction of one of the two tribes.

**Figure 6:** Combinations of parameters explored numerically for Game of Chicken cases (with



**Figure 7:** Illustrating the evolution of mistrust, resulting in the extinction of tribe 1 (for a Game of Chicken case with )



**Figure 8:** Combinations of parameters explored numerically for cases with and



**After Extinction, Schism?**

We have seen that allowing evolution of the population commonly gives rise to a stable equilibrium population, but that allowing evolution in the game strategy (i.e., evolution of the mistrust parameters) inevitably leads to the extinction of one tribe. What happens after one tribe becomes extinct and the population consists of a single tribe of indistinguishable individuals? A further stability question arises: is a single tribe stable under the possibility of schism, i.e., splitting in two – assuming membership of the two new schismatic groups is distinguishable?

One schismatic group is envisaged as having an infinitesimally perturbed mistrust. Additional analysis is not necessary. The preceding analysis already tells us that one or other of these schismatic groups will ultimately be driven out of the population.

In fact, tribes are vulnerable to schism even before the extinction of the other tribe. Consider tribe 1 undergoing a schism into two new tribes 1a and 1b. The mistrust matrix is defined as,

(26)

Denoting the average payoffs to the three tribes as , the equilibrium condition is . Denoting the equilibrium population fractions as where , the *relative* prevalence of the two schismatic tribes is found to be,

(27a)

(27b)

where the parameters depend only on mistrust parameters but not upon , thus,

(28a)

(28b)

(28c)

(28d)

Consequently, the relative prevalence of the two schismatic tribes can be evolved by considering variations in only at fixed . Noting the formal similarity between (28a-d) and equs.(4,5) it follows that, considering evolution of the mistrust parameters only, at fixed , will lead to extinction of one of the schism tribes, i.e., either tribe 1a or tribe 1b. Thus, schism of a tribe, followed by extinction of one of the schismatic groups, can occur prior to extinction of tribe 1 or 2, depending simply upon which set of parameters is considered to evolve first (or fastest).

**Correlated Mixed Strategies**

Conventional game theory was given a new twist by (Aumann, 1974) who observed that the stochastic play associated with a mixed strategy need not be uncorrelated. This gave rise to the concept of a “correlated equilibrium”, in which the utility might exceed that in the conventional (uncorrelated) equilibrium. It is difficult to imagine how such correlations could come about in a stochastic strategy without the involvement of a third party (whether animate or inanimate). Here we make the point that extension of our evolutionary scenario to such a correlated stochastic strategy does not provide an escape from the conclusion of tribal extinction when mistrust parameters are evolved.

Consider an actor from tribe 1 playing another actor from tribe 1. Instead of the single mistrust parameter which defines the uncorrelated mixed strategy we introduce the parameters where . These represent the probabilities of the two actors playing respectively. Hence, in the expression (3a) for the expected payoff to a tribe 1 actor, the term is replaced by,

(29)

Introducing similar correlated stochastic strategy parameters for each of the four combinations of two actors from tribes 1 and 2 results in 12 independent strategy parameters in place of the four mistrust parameters, . Equ.(12) for the equilibrium population fraction still holds, but the terms are replaced by expressions in terms of the twelve new parameters, of which (29) is one. We now have a 12 parameter space in which to evolve the new correlated stochastic strategy. But the original 4-dimensional parameter space is a curved 4-dimensional “surface” within the 12-dimensional parameter space, as obtained by setting and together with similar requirements obtained from . Consequently, we may still confine attention to evolution on this “surface” if we wish, and hence reproduce the foregoing results. The extra degrees of freedom available within the 12-dimensional parameter space cannot restrict the scope for increasing (or decreasing) the population fraction, only enhance the scope. Hence we conclude that tribal extinction, and schism, remain the fate of a population which is allowed to evolve within this enhanced correlated strategy.

**Discussion and Conclusion**

Previous work has concentrated almost exclusively, and very reasonably, on the conditions required for social cooperation to arise and be stable. However, it is salutary also to reflect on conditions which would make this impossible, that is those aspects of society which are *necessary* for social cooperation. Recognition of individuals is such a necessary condition, without which reputation, reciprocation and retribution would not be possible. The present work serves to underwrite and emphasise this elementary, but crucial, point. It has been noted before that the highly developed ability of humans to identify others by their faces, and the ‘labelling’ of humans by unique names, are key aspects of individual recognition, of central importance for cooperative behaviour, (Axelrod and Hamilton, 1981).

When a population consists of individuals who are indistinguishable apart from their membership of two tribes, and when individuals interact via a two-player game, the most general strategy which can be adopted is a stochastic (“mixed”) strategy based only on tribal membership. This is the case even if individuals within a given tribe deviate from statistically equivalent play. Even then, at the tribal level, the strategy is as modelled herein because the inability of actors to identify individuals prevents any strategy based on individuals’ play preferences.

Equilibrium populations of the two tribes exist, for suitable parameter values, which are stable in the usual sense that a perturbation of the population, even a large perturbation, will result in reversion back to the equilibrium population. Allowing the stochastic strategy itself to evolve, by permitting changes in degrees of mistrust, reveals that extinction of one of the tribes is the universal characteristic.

In practice, if a group finds itself threatened in this way, it is likely to decline to play, and social segregation would result instead of extinction.

At first sight this is hardly surprising. By removing the possibility of cooperation, by preventing the identification of individuals, the Prisoners’ Dilemma might be expected to revert to its harsh Nash logic in which both players always renege: whatever the other actor plays, one is better off reneging (playing ). But there are a couple of features in the results of the model presented which show this is not quite all that happens. Firstly, the mistrust (the frequency of reneging) never reaches 1. As illustrated by Figures 3 and 4, one tribe becomes extinct when all mistrust values still lie between 0 and 1. Secondly, the behaviour of the model for the Game of Chicken, i.e., when , is the same: one tribe invariably becomes extinct while all mistrust values still lie between 0 and 1. Yet the characteristic of the Game of Chicken is that, whatever the opponent plays, one is better off playing the *opposite*. This does not obviously bias parameter changes to be *increases* in mistrust, but could go either way. This is illustrated by Figure 7 which shows that the dominant effect is the *reduction* in the intra-tribe mistrust.

The tribal markers which make tribal membership apparent need not be related to the nature of the game itself. For example, a tribal marker may be a visible racial characteristic, whereas the ‘game’ may be some commercial exchange. Alternatively, tribes may be defined by belief systems or political opinion, providing that these are made apparent. The growth or reduction of a tribe in the population would then represent the growth or decline in the popularity of that belief or opinion. The analysis provides a cautionary tale against tendencies to assume that all members of an identifiable group are “the same”. By relegating the importance of the individual in favour of promoting the significance of group membership, there is a risk that the conditions of the foregoing analysis become a credible approximation. If so, one group may be destined to disappear. In the case of groups defined by opinions or beliefs, this may provide a mechanism by which conformity and group-think arises, whilst view point diversity is suppressed.

**References**

Aksoy, O., & Weesie, J. (2009). Inequality and Procedural Justice in Social Dilemmas. *The Journal of Mathematical Sociology*, 33(4), 303-322. doi:[10.1080/00222500903069624](https://doi.org/10.1080/00222500903069624)

Aksoy, O., & Weesie, J. (2013). Social Motives and Expectations in One-Shot Asymmetric Prisoner's Dilemmas. *The Journal of Mathematical Sociology*, 37(1), 24-58. doi:[10.1080/0022250X.2011.556764](https://doi.org/10.1080/0022250X.2011.556764)

Alexander, R.D. (1987). *The Biologv of Moral Systems*. New York: Aldine De Gruyter.

Aumann, R.J. (1974). Subjectivity and Correlation in Randomized Strategies. *J.Math.Econ*., 1, 67-96. doi <https://doi.org/10.1016/0304-4068(74)90037-8>

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books Inc, NY.

Axelrod, R., Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, 211(4489), 1390-1396. doi [10.1126/science.7466396](http://dx.doi.org/10.1126/science.7466396)

Boyd, R., & Richerson, P. J. (1988). The Evolution of Reciprocity in Sizable Groups. *J. theorl Biol*., 132(7), 337-356. doi:[10.1016/S0022-5193(88)80219-4](https://doi.org/10.1016/S0022-5193(88)80219-4)

Boyd, R., & Richerson, P. J. (1989). The Evolution of Indirect Reciprocity. *Social Networks*, 11(3), 213-236. doi:10.[1016/0378-8733(89)90003-8](https://doi.org/10.1016/0378-8733(89)90003-8)

Boyd, R., & Richerson, P. J. (1992). Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups. *Ethology and Sociobiology*, 13(3), 171-195. doi:[10.1016/0162-3095(92)90032-Y](http://dx.doi.org/10.1016/0162-3095(92)90032-Y)

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Phil. Trans. R. Soc. B*, 364(1533), 3281–3288. doi: [10.1098/rstb.2009.0134](http://dx.doi.org/10.1098/rstb.2009.0134)

Camerer, C. F. (2003a). Behavioural studies of strategic thinking in games. *Trends in Cognitive Sciences*, 7(5), 225-231. doi:[10.1016/S1364-6613(03)00094-9](https://doi.org/10.1016/S1364-6613(03)00094-9)

Camerer, C. F. (2003b ). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton NJ.

Camerer, C. F., & Smith, A. (2012). Cognitive hierarchies and emotions in behavioral game theory. In K. J. Holyoak & R. G. Morrison (Eds), *Oxford Handbook of Thinking and Reasoning* (pp 346-363), Oxford University Press, USA. doi:[10.1093/oxfordhb/9780199734689.013.0018](http://dx.doi.org/10.1093/oxfordhb/9780199734689.013.0018)

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *PNAS*, 112(7), 1989-1994. doi:[10.1073/pnas.1418838112](https://doi.org/10.1073/pnas.1418838112)

Chiang, Y-S. (2007). The Evolution of Fairness in the Ultimatum Game. *The Journal of Mathematical Sociology*, 31(2), 175-186. doi:[10.1080/00222500601188494](https://doi.org/10.1080/00222500601188494)

Dijkstra, J., & van Assen, M. A L. M. (2017). Explaining cooperation in the finitely repeated simultaneous and sequential prisoner’s dilemma game under incomplete and complete information. The Journal of Mathematical Sociology, 41(1), 1-25. doi:[10.1080/0022250X.2016.1226301](https://doi.org/10.1080/0022250X.2016.1226301)

Flache, A. (2002). The rational weakness of strong ties: Failure of group solidarity in a highly cohesive group of rational agents. The Journal of Mathematical Sociology, 26(3), 189-216. doi:[10.1080/00222500212988](https://doi.org/10.1080/00222500212988)

Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, 377:520–522. doi:[10.1038/377520a0](http://dx.doi.org/10.1038/377520a0)

Frank, S. A. (1996). Policing and group cohesion when resources vary. *Animal Behaviour*, 52:1163–1169. doi:<https://doi.org/10.1006/anbe.1996.0263>

Frank, S. A. (1998). *Foundations of Social Evolution*. Princeton University Press, Princeton, N.J.

Frey, V., Buskens, V., & Raub, W. (2015). Embedding Trust: A Game-Theoretic Model for Investments in and Returns on Network Embeddedness. The Journal of Mathematical Sociology, 39(1) 39-72. doi:[10.1080/0022250X.2014.897947](https://doi.org/10.1080/0022250X.2014.897947)

Gardner, A., & West, S. A. (2004). Cooperation and Punishment, Especially in Humans. *The American Naturalist*, 164(6), 753-764. doi:<https://doi.org/10.1086/425623>

Hobbes, T. (1651). *Leviathan*. First published 1651, Wordsworth UK Edition 2014.

Marlowe, F. W., Berbesque, C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., … Tracer, D. (2008). More ‘altruistic’ punishment in larger societies. *Proc. R. Soc. B*, 275(1634), 587–590. doi:[10.1098/rspb.2007.1517](http://dx.doi.org/10.1098/rspb.2007.1517)

Manhart, K., & Diekmann, A. (1989). Kooperative Strategien im Gefangenendilemma. Computersimulation eines N-Personen-Spiels (Cooperation in 2- and N-Person Prisoner's Dilemma Games: A Simulation Study). *Analyse & Kritik.* 11(2), 134-153. doi:[10.1515/auk-1989-0202](https://philpapers.org/go.pl?id=DIEKSI&proxyId=&u=http%3A%2F%2Fdx.doi.org%2F10.1515%2Fauk-1989-0202)

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, 246, 15-18. doi:[10.1038/246015a0](http://dx.doi.org/10.1038/246015a0)

Maynard Smith, J. (1974). The Theory of Games and the Evolution of Animal Conflicts. *J. theor. Biol.*, 47, 209-221. doi:[10.1016/0022-5193(74)90110-6](https://doi.org/10.1016/0022-5193(74)90110-6)

Perc, M., & Szolnoki, A. (2008). Social diversity and promotion of cooperation in the spatial prisoner’s dilemma game. *Phys. Rev. E*, 77, 011904(1-5). doi:[10.1103/PhysRevE.77.011904](http://dx.doi.org/10.1103/PhysRevE.77.011904)

Rezaei, G., Kirley, M. & Pfau, J. (2009). *Evolving Cooperation in the N-player Prisoner's Dilemma: A Social Network Model*. Australian Conference on Artificial Life, ACAL 2009: Artificial Life: Borrowing from Biology, 43-52.

Roca, C. P., Sanchez, A., & Cuesta, J. A. (2012). Individual Strategy Update and Emergence of Cooperation in Social Networks. [*The Journal of Mathematical Sociology*,](https://www.tandfonline.com/toc/gmas20/current) 36(1), 2012. doi:[10.1080/0022250X.2010.520830](http://dx.doi.org/10.1080/0022250X.2010.520830)

Sato, Y. (2002). Trust, Assurance and Inequality: A Rational Choice Model of Mutual Trust. *Journal of Mathematical Sociology*, 26(1/2), 1-16. doi:[10.1080/00222500210525](http://dx.doi.org/10.1080/00222500210525)

Sigmund, K., Hauert, C., & Nowak, M. A. (2001). Reward and punishment. *Proceedings of the National Academy of Sciences of the USA*, 98(19), 10757–10762. doi:[10.1073/pnas.161155698](http://dx.doi.org/10.1073/pnas.161155698)

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 46(1), 35-57. doi:[10.1086/406755](http://dx.doi.org/10.1086/406755)