Legal and Criminological Psychology (1997), **2**, 23–34 *Printed in Great Britain*

Antisocial personality disorder: An evolutionary game theory analysis

Andrew M. Colman* and **J. Clare Wilson**
Department of Psychology, University of Leicester, Leicester LE1 7RH, UK

Purpose. To develop a multi-person evolutionary game, with population replicator dynamics based on the payoffs of the Chicken (Hawk–Dove) game, to model Antisocial Personality Disorder (APD), and to offer an explanation for the relatively stable prevalence of APD in widely diverse societies despite increasing resources devoted to reducing antisocial behaviour.

Methods. Beginning with a basic two-person game, a multi-person evolutionary model is developed. According to the model, changes in the frequency of APD in the population depend on frequency-dependent Darwinian selection or a form of social evolution that mimics it.

Results. The population evolves to a stable equilibrium with a fixed proportion of individuals habitually behaving antisocially, and with suitable payoffs the proportion of antisocial individuals corresponds to the known prevalence of APD. An unexpected result of the analysis is the finding that the prevalence is necessarily low when the relative gain from behaving antisocially towards a cooperator is very much smaller than the relative loss to the cooperator.

Conclusion. The model provides an evolutionary game-theoretic explanation for the low but stable prevalence of APD. If the evolutionary mechanism is social rather than biological, then removing increasing numbers of antisocial individuals from society will result in others taking their places, and the population will return to the equilibrium point.

The American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) describes Antisocial Personality Disorder (APD) as 'a pervasive pattern of disregard for, and violation of, the rights of others that begins in childhood or early adolescence and continues into adulthood' (American Psychiatric Association, 1994, p. 645). The World Health Organization's *International Classification of Diseases* (ICD-10) calls it Dissocial Personality Disorder and describes it as a 'personality disorder characterized by disregard for social obligations and callous unconcern for the feelings of others' (Cooper, 1994, p. 226). There is an unresolved debate in the literature about whether APD shares a common referent with psychopathy and sociopathy, but in any event their operational measures are rather different. For example, factor analytic studies of scores from the Psychopathy Checklist-Revised (PCL-R), constructed by Robert D. Hare (1985, 1991) on the basis of an earlier prototype (Hare, 1980), have revealed two stable, oblique factors. Factor 1 consists of

* Requests for reprints.

affective and interpersonal traits (such as superficial charm, pathological lying, egocentricity, lack of remorse, and callousness) that have traditionally been regarded by clinicians as characteristic of psychopaths; Factor 2 reflects social deviance aspects of psychopathy (such as need for stimulation, parasitic lifestyle, poor behavioural controls, impulsivity, and irresponsibility) that are highly correlated with diagnoses of APD (Hare, Hart & Harpur, 1991).

The pattern of behaviour associated with APD is characteristically repetitive and persistent: people with APD tend to engage in innumerable antisocial or criminal activities such as assault, robbery, threatening behaviour, criminal damage, theft, harassment, substance abuse, and fraud, and to be habitually deceitful and manipulative in their pursuit of pleasure or personal gain. The course of the disorder is chronic, although its symptoms may become less evident as a person grows older. To fulfil the definition of APD given in DSM-IV, a person must be at least 18 years of age and must meet three or more of the following seven diagnostic criteria (paraphrased from American Psychiatric Association, 1994, pp. 649–650):

(1) Failure to conform to social norms, as indicated by repeated unlawful behaviour;
(2) deceitfulness, as indicated by repeated lying or swindling for pleasure or personal gain;
(3) impulsivity or failure to plan ahead;
(4) irritability and aggressiveness involving frequent assaults or fights;
(5) reckless disregard for the safety of self or others;
(6) consistent irresponsibility involving failure to hold down jobs or honour financial obligations;
(7) lack of remorse for the mistreatment of others, as indicated by indifference and rationalization.

The diagnostic criteria of Dissocial Personality Disorder in ICD-10 are broadly similar, the main differences being the inclusion of criteria referring to 'incapacity to maintain enduring relationships, though with no difficulty in establishing them' and 'incapacity to experience guilt, or to profit from adverse experience, particularly punishment' (Cooper, 1994, pp. 227–228).

The forms of behaviour characteristic of APD have been reported in societies with widely different social and economic systems and in all eras, which suggests that they are not the results of a recent pathology of modern industrial culture (Robins, Tipp & Przybeck, 1991, p. 259). The prevalence of APD in the United States is relatively stable at about 2 per cent (about 3 per cent among men and 1 or 2 per cent among women), or perhaps marginally higher, and broadly similar prevalence rates have been reported in other societies (American Psychiatric Association, 1994; Cloninger & Gottesman, 1987; Davison & Neale, 1994; Robins *et al.*, 1991). There are no marked differences in prevalence rates between different ethnic groups in the United States (Robins *et al.*, 1991, pp. 271–276), although prevalence is generally higher in inner-city areas than in rural villages (Robins *et al.*, 1984, 1991, pp. 280–283). It is generally agreed that people with APD account for about 50–60 per cent of the federal prison population in the United States (Hare, 1980, 1993; Hare *et al.*, 1991; Harpending & Sobus, 1987; Mealey, 1995), and according to one study about 75 per cent of the male inmates in Canadian federal prisons met the criteria for APD (Correctional Service of Canada, 1990), which implies that they are responsible for a large proportion of reported and probably also unreported crime.

Although changes in operational definitions and diagnostic measures make precise intertemporal comparisons somewhat hazardous, there is no evidence to suggest that the prevalence of APD is lower today than it was half a century ago (cf. Page, 1947, p. 397). If the prevalence rate has indeed been relatively stable over time, then this seems surprising in the light of public policy over the past several decades, which has involved devoting large and escalating resources of the criminal justice system and the social and psychiatric services to controlling antisocial behaviour and removing increasing numbers of people with APD from society.

What accounts for the persistence of a small but disproportionately disruptive minority of people with APD? One possible approach to this problem is to construct a mathematical model of the evolution of antisocial behaviour in an effort to simulate and thereby clarify the dynamic processes that control it. A mathematical model is an abstract idealization of a phenomenon—in this case of the sociobiology of APD. The phenomenon itself is invariably too nebulous and complex to understand clearly, so it is replaced by a deliberately simplified mathematical structure whose rules and basic elements are clearly specified and from which other properties can be deduced by formal reasoning. Such deductions are necessarily true, but they apply to the mathematical model rather than the phenomenon itself; how well the model corresponds to the original phenomenon is always a matter of judgment and empirical evidence. To be useful, the model must not only capture the essential features of the phenomenon but also generate insights that transcend a merely common-sense understanding of it.

Frank (1987, 1988, pp. 43–70, 158–162, 261–269; see also Frank, Gilovich & Regan, 1993), Harpending & Sobus (1987), and Mealey (1995) have suggested game-theoretic sociobiological models of competitive and cheating behaviour according to which APD may be interpreted as the expression of a frequency-dependent life history strategy arising from natural selection. The basic premise is that such behaviour results from an evolutionary process that, through an interaction of genetic and environmental factors, leads to an equilibrium in which a small but stable minority of people habitually use predatory and manipulative strategies in their interactions with others. The evolutionary models suggested by Frank, Harpending & Sobus, and Mealey are based on the well-known Prisoner's Dilemma game, in which two players each face a choice between a cooperative and a non-cooperative strategy. In this game, the highest possible payoff (which means simply the player's most preferred outcome) results from choosing non-cooperatively while the other player cooperates, and the lowest payoff from cooperating while the other player chooses non-cooperatively, and both players are better off if both cooperate than if both choose non-cooperatively.

The Prisoner's Dilemma game is an inappropriate model for the sociobiology of APD for a number of reasons. First, a two-person game can never adequately model an inherently multi-person phenomenon such as the evolution of a mixed population in which a stable minority exhibits APD—what is clearly required is a multi-person game. Second, if many people interact wth one another in pairs according to the payoffs of the Prisoner's Dilemma game, it can be shown that the evolutionarily stable equilibrium that results is not one in which a small minority consistently choose the non-cooperative strategy, but rather a uniform population in which *all* members of the population do so (Colman, 1995*b*). Furthermore, if the encounters involve iterated (rather than one-off) Prisoner's Dilemma games, the population still does not evolve to the type of strategy mixture that is seen in the epidemiology of APD (e.g. Axelrod, 1984; Bendor, 1993; Nowak & Sigmund, 1993; Selten & Hammerstein, 1984). Third, in the Prisoner's Dilemma game the non-cooperative strategy does not provide a convincing model of the predatory, antisocial behaviour that is characteristic of APD. Studies of the social value orientations of players have shown that people with competitive or individualistic social value orientations are more likely than cooperatively motivated people to choose the non-cooperative strategy (Kramer, McClintock & Messick, 1986; Kuhlman & Marshello, 1975; Liebrand & van Run, 1985; McClintock & Liebrand, 1988), but there is no evidence that non-cooperative choices are associated with predatory or antisocial motivations or with disregard for social obligations and callous unconcern for the feelings of others.

The purpose of this article is to suggest a different mathematical model, worked out in detail, that provides a more satisfactory sociobiological explanation of APD. The section immediately below describes the basic structure of the model, and that is followed by a section summarizing the main results that can be deduced from it. The final section contains a discussion of the model and its implications.

Method

Although a multi-person model is required, we begin with a simple two-person game, more appropriate than the Prisoner's Dilemma game, which functions as a basic building block for a multi-person evolutionary model that is later constructed from it via the theory of compound games (Colman, 1995*a*, pp. 209–212).

Basic game

The basic two-person game is usually referred to in the literature of game theory as the game of Chicken (e.g. Hamburger, 1979, pp. 83–87; Rasmusen, 1989, pp. 73–74; Shubik, 1991, pp. 394–395) and in the literature of evolutionary biology as the Hawk–Dove game (e.g. Lazarus, 1995, pp. 37–38; Maynard Smith, 1976, 1978; Maynard Smith & Price, 1973). The game is specified in Table 1.

Player 1 chooses between row C (cooperative or cautious) or row D (dangerous or antisocial), and at the same time Player 2 chooses between column C or column D. The pair of numbers in each cell represent the players' payoffs: the first number always represents Player 1's payoff and the second number Player 2's. Thus, for example, if Player 1 chooses C and Player 2 D, the payoffs in the top-right cell show that Player 1's payoff is 2 and Player 2's is 4. What defines this game as Chicken is the relative rather than the absolute values of the payoffs: the game is Chicken by definition provided that the best payoff results from choosing D when the co-player chooses C, the second-best payoff from choosing C when the co-player also chooses C, the third-best payoff from choosing C when the co-player chooses D, and the worst payoff from choosing D when the co-player also chooses D.

Table 1. The game of Chicken

		II	
		C	D
I	C	3, 3	2, 4
	D	4, 2	1, 1

Chicken is the prototypic dangerous game and is often associated with antisocial, predatory, and manipulative social interactions (Swingle, 1970). Its name derives from a hazardous game involving displays of *machismo* that became notorious after the release of Nicholas Ray's film, *Rebel Without a Cause*, starring James Dean, in 1955. In its most familiar version, two drivers speed towards each other in an open area. Each can choose either the cooperative or cautious strategy (*C*) of swerving to avoid a head-on collision (and thereby being 'chicken') or the dangerous and antisocial strategy (*D*) of driving straight ahead. If both drivers are cautious, the outcome is a draw with second-best payoffs to each (3, 3), and if both drive dangerously at each other, they risk death or serious injury, which yields the worst payoffs to each (1, 1). But if one driver cautiously swerves while the other dangerously drives straight ahead, then the cautious driver loses face and receives the third-best payoff of 2 while the dangerous driver wins a prestige victory and receives the best possible payoff of 4.

Among all strategically distinct two-person, two-strategy games, of which there are exactly 78 (Rapoport & Guyer, 1966), Chicken is the one that seems best suited to modelling the sociobiology of APD, partly because it occurs frequently in everyday strategic interactions involving risk taking and the threat of aggression, and partly because it has a number of strategic properties that seem applicable to APD. First, it has a uniquely compulsive quality that makes an invitation to play it impossible to refuse, inasmuch as declining a challenge to play Chicken amounts to playing a version of it and losing. Second, a player who appears resolutely committed to choosing the dangerous or antisocial *D* strategy is bound to win, provided only that the other player is rational, thus a person who cultivates a reputation for 'reckless disregard for the safety of self or others' (DSM-IV, criterion 5), 'lack of remorse for the mistreatment of others' (DSM-IV, criterion 7), or 'irritability and aggressiveness' (DSM-IV, criterion 4) is therefore likely to gain an advantage from the justifiable fear that this induces in any rational opponent; in fact there is authoritative anecdotal evidence that some professional criminals deliberately use precisely this form of brinkmanship (McVicar, 1981, pp. 225–226). The third strategic property, which is peculiar to Chicken, is the paradoxical sense in which it is rational to behave irrationally. A player who is seen to be irrational, impulsive, or unpredictable gains an advantage in Chicken, because there is a natural tendency to give a wide berth to anyone who shows a 'failure to conform to social norms' (DSM-IV, criterion 1), 'impulsivity or failure to plan ahead' (DSM-IV, criterion 3), or 'consistent irresponsibility' (DSM-IV, criterion 6). In Chicken-type encounters, a player thus gains advantage from deliberately appearing irrational, and there is anecdotal evidence that this ploy is also used in the pursuit of antisocial and criminal behaviour (e.g. Haldeman & DiMona, 1978, p. 83). The last strategic property of Chicken, and the one that is most relevant to the sociobiology of APD, is an aggregation effect that emerges when Chicken is transformed from a two-person to a multi-person game. This needs to be explained in some detail in a separate subsection.

Multi-person evolutionary model

We need to establish what will happen in an entire community in which individuals interact with one another in pairwise games with the strategic structure of Chicken. To explain the evolution of APD, we begin by assuming that the payoffs represent units of *Darwinian fitness* (the lifetime reproductive success of the individual players) and that the propensity to choose *C* or *D* is at least partly heritable. Taken together, these two assumptions imply that a player who receives a large payoff following a particular choice of *C* or *D* will leave relatively many offspring who tend to adopt the same strategy, and a player who receives a smaller payoff will leave relatively few. Consequently, strategies that yield high payoffs will tend to increase in relative frequency in the population, and less successful strategies will tend to decrease.

Patterns of behaviour may also be subject to evolution by a form of natural selection that occurs through learning and imitation rather than genetics (Dawkins, 1989). Dawkins called behaviour patterns that are subject to social evolution in this way *memes*—a word designed to resemble *genes*—because they are sustained by memory and mimicry. Like genes, memes are self-replicators: in suitable environments they produce multiple copies of themselves. Some are fitter than others, as measured by the number of people who adopt them, and only the fittest—those that people copy most often—survive in the struggle for existence. Social evolution can, of course, proceed much more rapidly than biological evolution, because it is not limited by the reproductive rate of the species, and a propensity to choose the cooperative/cautious *C* strategy or the dangerous/antisocial *D* strategy in Chicken-type interactions can clearly function as a meme and spread by social evolution. Our evolutionary model of APD is neutral as to whether the evolution proceeds biologically or socially, and it therefore does not rest on any assumptions about the heritability of antisocial behaviour, although there is evidence that it has a genetic component (Bock & Goode, 1996; Gottesman & Goldsmith, 1994; Schulsinger, 1972). In fact, biological and social evolution may even occur in parallel.

Suppose, therefore, that members of a community encounter one another in pairwise interactions with the strategic structure of Chicken. For mathematical simplicity, assume that within a specified time period each person plays one game of Chicken either with every other member of the community or with a random sample of the others and that everyone plays the same number of games.

From the viewpoint of a single player, suppose that the proportion of other players choosing C is k. The proportion k can vary from zero, if none of the others chooses C, to unity if all of the others choose C, and the proportion of the others choosing D must be $(1-k)$. At the end of the specified time period, after all of the two-person encounters, the average payoffs, according to the figures in Table 1, will be as follows:

$$\text{Average payoff to a } C \text{ chooser} = 3k + 2(1-k) = k + 2, \tag{1}$$

$$\text{Average payoff to a } D \text{ chooser} = 4k + 1(1-k) = 3k + 1. \tag{2}$$

A player who uses a so-called 'mixed strategy' receives a weighted average of the payoffs specified by (1) and (2): for example, a player who randomizes between choosing C half the time and D half the time receives a total expected payoff of $(k+2)/2 + (3k+1)/2 = 2k + 1\frac{1}{2}$. These are average payoffs; to specify actual payoffs to an individual player an additional term would have to be added to each equation to represent random errors (what game theorists call 'trembles') arising from the varying circumstances of particular encounters; but this is unnecessary for our purposes.

In order to produce a graphical representation of this simple mathematical model we need to fix some numerical values of the payoff functions (1) and (2). First, the average payoff to a C chooser if none of the other players chooses C (that is, if $k=0$), which is found by setting k equal to zero in (1), is 2. Second, the average payoff to a C chooser if all of the other players choose C, found by setting $k=1$ in (1), is 3. Third, the average payoff to a D chooser if none of the other players chooses C (if $k=0$), found by setting k equal to zero in (2), is 1. Last, the average payoff to a D chooser if all of the other players choose C, found by setting $k=1$ in (2), is 4.

Results

The final form of the model, with the parameters set to the values given in Table 1, is shown in Fig. 1. The end-points of each payoff function are the average payoffs to a player choosing C or D when none of the others chooses C (when $k=0$ at the left) or when all of the others choose C (when $k=1$ at the right).

Dynamics

The first point to notice in Fig. 1 is that the point of intersection of the two payoff functions represents a stable equilibrium, any deviation from which tends to be self-correcting because of the replicator dynamics built into the model. When a relatively small proportion of the other players choose C—that is, when k is small to the left of the intersection—the C function lies above the D function, which means that the average payoff from a C choice is higher than from a D choice. This in turn means that C choosers will replicate faster than D choosers, so that if choosing C is at all heritable or imitable, the number of C choosers will increase relative to D choosers

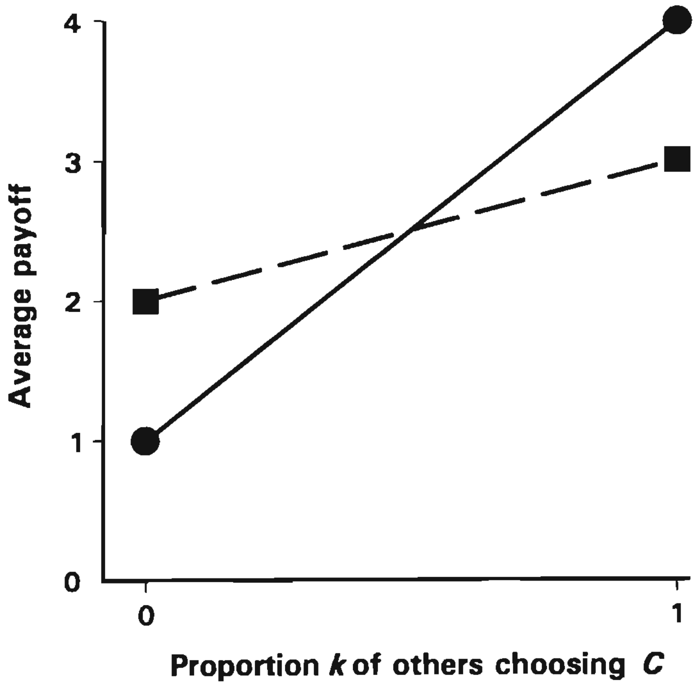

Figure 1. Multi-person Chicken or Hawk–Dove game, based on the payoffs given in Table 1, where the proportion *k* of other players choosing *C* varies from zero to unity. ■—■, payoff to a *C* chooser; ●—●, payoff to a *D* chooser.

and thus *k* will increase towards the intersection. But when relatively many of the others choose *C*, that is, when *k* is large, a player gains a higher average payoff from choosing *D* than *C*, so *C* choosers will decrease relative to *D* choosers and *k* will decrease towards the intersection. At the point of intersection of the payoff functions a player receives the same payoff for a *C* choice as for a *D* choice, and the model is in equilibrium. At the equilibrium point, and only there, the two strategies are equally effective in terms of Darwinian (or social) fitness, and any deviation from the mixture at that point will tend to be self-correcting. This implies that, whatever the starting value of *k*, the community will tend to evolve by natural or meme selection towards the stable equilibrium at the intersection point, with a certain proportion of the population choosing *C* and the rest choosing *D*.

Results with large k

The specific value of *k* at which the payoff functions intersect depends on the payoffs in the basic two-person game from which the model is constructed. The payoffs shown in Table 1, which were used to produce Fig. 1, were chosen for maximum simplicity by using the numbers 4, 3, 2 and 1 for a player's best, second-best, third-best, and worst outcomes. When the parameters are modified within the ordinal structure that defines the game of Chicken, the basic properties of the model remain intact, and in

Table 2. The game of Chicken with revised parameters

		II	
		C	*D*
I	*C*	99, 99	50, 100
	D	100, 50	1, 1

particular the payoff functions always intersect as in Fig. 1, but the point of intersection varies.

This leads naturally to the question, what parameters might provide a realistic representation of APD? As mentioned earlier, surveys have shown that the prevalence of APD is relatively stable at approximately 2 per cent, which suggest that $k \approx .98$ at the point of intersection. This would be achieved if, for example, the average payoffs were as follows: for joint *C* choices, 99; for joint *D* choices, 1; for a *C* choice against a *D* choice, 50; and for a *D* choice against a *C* choice, 100 (see Table 2).

The intersection of the payoff functions, with *C* and *D* choices yielding the same average payoff, occurs where the values of the payoff functions are equal. With the parameters shown in Table 2, equality occurs where

$$99k + 50(1-k) = 100k + 1(1-k),$$

which simplifies to $k = .98$ as required. These are not the only parameters that yield a value of $k = .98$, but they do illustrate a property that has emerged as a finding of the dynamic model, namely that an equilibrium point with such a small proportion of *D* choosers occurs only when the average gain from choosing *D* rather than *C* against a *C* chooser is vastly less than the average loss to the player on the receiving end of that encounter—in the example just given the payoff of the dangerous exploiter goes up from 99 to 100, but the cooperative victim's payoff falls all the way from 99 to 50.

The model turns out to provide a possible explanation for the far higher prevalence of antisocial behaviour among young men than older men or women, for which there is ample research evidence (Ellis, 1990; Robins *et al.*, 1991). Wilson & Daly (1985) have argued that Darwinian fitness among males is limited chiefly by access to fertile females, whereas female fitness is limited by physiological and energy constraints, therefore high-status males can enhance their fitness by monopolizing the reproductive potential of several females, whereas females cannot profit to the same extent from multiple mates. This implies that young men have experienced intense within-sex reproductive competition during the evolutionary history of the species, and that this has resulted in greater variance in fitness among males than females. On the assumption that the greater the payoff discrepancy between winners and losers in any competition, the greater the expected expenditure of effort and tolerable risk, Wilson and Daly argued that this may explain the 'young male syndrome' of risky or violent competitive behaviour. But in common with many other informal sociobiological theories, the argument is inconclusive, because as Wilson & Daly pointed out, in conflicts between men of different ages, 'one might instead predict . . . that it is the *older*, not the younger, who has less to lose and should therefore be readier to employ dangerous competitive tactics' (p. 70, italics in original).

Conclusions

Is there a sociobiological explanation for the small but relatively stable minority of people with APD found in most large communities? The Prisoner's Dilemma game cannot provide a realistic basis for an explanation of this phenomenon, but a multi-person compound game based on Chicken or the Hawk–Dove game provides a possible explanatory mechanism. The fact that Chicken is a quintessentially dangerous game (the DSM-IV diagnostic criteria for APD include 'impulsivity or failure to plan ahead', 'irritability and aggressiveness', and 'reckless disregard for the safety of self or others'), together with several of its unusual strategic properties discussed above, make it a natural and perhaps obvious choice for modelling the sociobiology of APD. Most importantly, the multi-person compound game that has been developed from it in this article was shown to possess a stable evolutionary equilibrium at which a proportion of the population choose the cooperative or cautious strategy and the rest choose the dangerous or antisocial strategy, and with plausible parameters the latter proportion is approximately 2 per cent, corresponding to the approximate prevalence of APD. What this implies is that when more than 2 per cent of the population exhibit APD-type antisocial behaviour, cooperative behaviour pays better, presumably because antisocial behaviour is too dangerous in those social environments; but when less than 2 per cent of the population exhibit antisocial behaviour, the danger of retaliation is less and antisocial behaviour pays better.

What is important is not the specific numerical proportion of people with APD, which may be slightly higher than 2 per cent in the United States (Robins *et al.*, 1991), but the fact that there exists in every community an equilibrium point at *some* fixed proportion. The factors that determine the equilibrium proportion in a given community include everything that determines the payoffs in the basic game. For example, heavily populated inner-city communities provide greater anonymity than rural communities and may therefore offer greater opportunities for antisocial behaviour to succeed and to go undetected, which implies that the average payoff advantage from choosing the antisocial D strategy rather than the cooperative C strategy against a C chooser may be greater in inner-city communities. The effect of increasing this payoff would be to raise the right-hand extremity of the payoff to a D chooser in Fig. 1 and thus to move the intersection point to the left, which means that the proportion of D choosers at the evolutionary equilibrium would increase. This may explain why the prevalence of APD is higher in inner-city than in rural communities (Robins *et al.*, 1984, 1991, pp. 280–283). According to the model, a stable evolutionary equilibrium point exists in every community, but the specific parameters of the model, and therefore the precise proportions of cooperative and antisocial individuals at the equilibrium, depend on local circumstances.

There are two possible interpretations of the mixed-strategy equilibrium in an evolutionary game of this type. Either the population evolves to a dimorphic mixture of cooperative and antisocial phenotypes, in the proportions specified by the parameters of the model, in this case 98 per cent and 2 per cent respectively, or else it evolves to a form in which every individual shows a propensity to randomize between behaving cooperatively and antisocially in the required proportions. The first interpretation is clearly more likely to be correct in this case, because most people seldom or never manifest the types of behaviour characteristic of APD, at least once they have reached adulthood, before which APD cannot be diagnosed according to DSM-IV (American Psychiatric Association, 1994, p. 650).

If the model described in this article is applicable to the real world, then it reveals an interesting property of APD behaviour that has not been formalized before, namely the fact that (in Western industrial societies at least) the relative gain to the person performing an antisocial, predatory, or exploitative act must on average be very slight compared to the relative loss suffered by the person on the receiving end of it. For example, a mugger may gain only a small amount from assaulting or even killing a passer-by, but the consequences for the victim may be catastrophic. This turns out to be a necessary consequence of the empirical observation that the prevalence of APD is in the region of 2 per cent: even assuming this figure to be only approximately correct, the model requires the relative gain to the antisocial exploiter to be very much less than the relative loss to the victim. It is perhaps not entirely obvious that the prevalence of APD is necessarily low when the D chooser's average gain is much less than the C chooser's average loss, and it may be regarded as a finding that transcends our common-sense understanding of APD. Furthermore, if the evolution of antisocial behaviour is at least partly genetic, then the fact that men have experienced more intense reproductive competition than women during the evolutionary history of the species, and that the variance in fitness is greater among men than women, may explain why risky and antisocial behaviour appears to largely a 'young male syndrome' in all societies (Wilson & Daly, 1985).

Can this model provide an explanation for the criminal, delinquent, and generally antisocial or predatory forms of behaviour that are prevalent at low frequencies in most large communities? Such forms of behaviour are, of course, due to multiple complex causes, and no single explanation is likely to be complete, but a vastly disproportionate amount of antisocial behaviour is apparently attributable to a small minority of people with APD, and this highly significant minority may possibly be maintained through frequency-dependent Darwinian selection or a form of social evolution that mimics it. If that is the case, then some mechanism similar to the one presented in this article must be at work. There must in any event be some explanation for the relative stability of APD prevalence rates, and no other plausible suggestion has appeared in the literature on APD.

If the model is correct, and if the replicator dynamics of the model are driven mainly by social rather than biological evolution, then the battle to eliminate antisocial behaviour cannot be won by removing increasing numbers of offenders from society. For even if it were possible to remove all people with APD from society, others would emerge to fill the resulting strategic vacuum, although if the model is driven exclusively by biological evolution, then this inference cannot necessarily be drawn. In any event, a more realistic public policy would involve interventions at the societal level aimed at coping with antisocial behaviour rather than attempting to eliminate it. What are required are measures designed to raise the equilibrium point itself to a higher proportion of C choosers. This in turn would require altering the payoffs of the basic game in Table 2. It is clear from Fig. 1 that any social or legal changes that increase the payoff for cooperation with either cooperative or antisocial individuals should have the effect of reducing the proportion of antisocial individuals at the evolutionary equilibrium point, and that any changes that decrease the payoff for antisocial behaviour in interactions with either cooperative or antisocial individuals should have the same effect. Thus, for example, any social interventions that improve people's sense of community and thus make mutual cooperation more rewarding and victimization less traumatic should increase the payoffs for cooperation, and any increase in either the likelihood of detection or the severity of punishment for antisocial behaviour should reduce the average payoffs for D choices; in both cases the effect should be to reduce the proportion of antisocial individuals at the equilibrium point. Although the analysis presented in this article suggests that it may be impossible to rid society of antisocial behaviour, it also implies that it is worth diverting some of the resources currently devoted to the criminal justice system towards encouraging and rewarding cooperative behaviour and fostering a sense of community, especially in relatively anonymous inner-city areas of society.

Acknowledgements

Preparation of this article was supported by a Grant No. L122251002 to the first author from the Economic and Social Research Council as part of the research programme on Economic Beliefs and Behaviour. We are grateful to Ron Blackburn and an anonymous referee for helpful comments on an earlier draft of the article.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Bendor, J. (1993). Uncertainty and the evolution of cooperation. *Journal of Conflict Resolution*, **37**, 709–734.

Bock, G. R. & Goode, J. A. (Eds) (1996). *Genetics of Criminal and Antisocial Behaviour* (Ciba Foundation Symposium 194). Chichester: Wiley.

Cloninger, C. R. & Gottesman, I. I. (1987). Genetic and environmental factors in antisocial behavior disorders. In S. A. Mednick, T. E. Moffitt & S. A. Stack (Eds), *The Causes of Crime: New Biological Approaches*, pp. 92–109. New York: Cambridge University Press.

Cooper, J. E. (1994). *Pocket Guide to the ICD-10 Classification of Mental and Behavioural Disorders.* Edinburgh: Churchill Livingstone.

Colman, A. M. (1995*a*). *Game Theory and its Applications in the Social and Biological Sciences*, 2nd ed. Oxford: Butterworth-Heinemann.

Colman, A. M. (1995*b*). Prisoner's Dilemma, Chicken, and mixed-strategy evolutionary equilibria. *Behavioral and Brain Sciences*, **18**, 510–511.

Correctional Service of Canada. (1990). A mental health profile of federally sentenced offenders. *Forum on Corrections Research*, **2**, 7–8.

Davison, G. C. & Neale, J. M. (1994). *Abnormal Psychology*, 6th ed. New York: Wiley.

Dawkins, R. (1989). *The Selfish Gene*, 2nd ed. Oxford: Oxford University Press.

Ellis, L. (1990). Universal behavioral and demographic correlates of criminal behavior: Toward common ground in the assessment of criminological theories. In L. Ellis & H. Hoffman (Eds), *Crime in Biological, Social and Moral Contexts*, pp. 36–49. New York: Praeger.

Frank, R. H. (1987). If *Homo economicus* could choose his own utility function, would he want one with a conscience? *American Economic Review*, **77**, 593–604.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions.* New York: Norton.

Frank, R. H., Gilovich, T. & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology*, **14**, 247–256.

Gottesman, I. I. & Goldsmith, H. H. (1994). Developmental psychopathology of antisocial behavior: Inserting genes into its ontogenesis and epigenesis. In C. A. Nelson (Ed.), *Threats to Optimal Development: Integrating Biological, Psychological, and Social Risk Factors* (Minnesota Symposium on Child Psychology), vol. 27, pp. 69–104. Hillsdale, NJ: Erlbaum.

Haldeman, H. R. & DiMona, J. (1978). *The Ends of Power.* New York: Times Books.
Hamburger, H. (1979). *Games as Models of Social Phenomena.* San Francisco, CA: Freeman.
Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, **1**, 111–119.
Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology*, **53**, 7–16.
Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised.* Toronto: Multi-Health Systems.
Hare, R. D. (1993). *Without Conscience: The Disturbing World of the Psychopaths Among Us.* New York: Pocket Books.
Hare, R. D., Hart, S. D. & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for Antisocial Personality Disorder. *Journal of Abnormal Psychology*, **100**, 391–398.
Harpending, H. & Sobus, J. (1987). Sociopathy as an adaptation. *Ethology and Sociobiology*, **8**(3), 63S–72S.
Kramer, R. M., McClintock, C. G. & Messick, D. M. (1986). Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality*, **54**, 576–592.
Kuhlman, D. M. & Marshello, A. J. F. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in Prisoner's Dilemma. *Journal of Personality and Social Psychology*, **32**, 922–931.
Lazarus, J. (1995). Behavioural ecology and evolution. In D. Kimble & A. M. Colman (Eds), *Biological Aspects of Behaviour*, pp. 28–46. London & New York: Longman.
Liebrand, W. B. G. & Van Run, G. J. (1985). The effects of social motives on behavior in social dilemmas in two cultures. *Journal of Experimental Social Psychology*, **21**, 86–102.
McClintock, C. G. & Liebrand, W. B. G. (1988). The role of interdependence structure, individual value orientation and other's strategy in social decision making: A transformational analysis. *Journal of Personality and Social Psychology*, **55**, 396–409.
McVicar, J. (1981). Postscript. In S. Cohen & L. Taylor (Eds), *Psychological Survival: The Experience of Long-Term Imprisonment*, 2nd ed., pp. 221–229. Harmondsworth: Penguin.
Maynard Smith, J. (1976). Evolution and the theory of games. *American Scientist*, **64**(1), 41–45.
Maynard Smith, J. (1978). The evolution of behavior. *Scientific American*, **239**, 136–145.
Maynard Smith, J. & Price, G. R. (1973). The logic of animal conflict. *Nature*, **246**, 15–18.
Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, **18**, 460–485.
Nowak, M. & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, **364**, 56–58.
Page, J. D. (1947). *Abnormal Psychology: A Clinical Approach to Psychological Deviants.* New York: McGraw-Hill.
Rapoport, A. & Guyer, M. (1966). A taxonomy of 2×2 games. *General Systems*, **11**, 203–214.
Rasmusen, E. (1989). *Games and Information: An Introduction to Game Theory.* Oxford: Blackwell.
Robins, L. N., Helzer, J. E., Weissman, M. M., Orvaschel, H., Gruenberg, E., Burke, J. K. & Regier, D. A. (1984). Lifetime prevalence of specific psychiatric disorders in three cities. *Archives of General Psychiatry*, **41**, 949–958.
Robins, L. N., Tipp, J. & Przybeck, T. (1991). Antisocial personality. In L. N. Robins & D. A. Regier (Eds), *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*, pp. 258–290. New York: Free Press.
Schulsinger, F. (1972). Psychopathy: Heredity and environment. In M. Roff, L. N. Robins & M. Pollack (Eds), *Life History Research in Psychopathology: II*, pp. 102–119. Minneapolis, MN: University of Minnesota Press.
Selten, R. & Hammerstein, P. (1984). Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of 'tit for tat'. *Behavioral and Brain Sciences*, **7**, 115–116.
Shubik, M. (1991). *Game Theory in the Social Sciences: Concepts and Solutions.* Cambridge, MA: MIT Press.
Swingle, P. G. (1970). Dangerous games. In P. G. Swingle (Ed.), *The Structure of Conflict*, pp. 235–276. New York: Academic Press.
Wilson, M. & Daly, M. (1985). Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and Sociobiology*, **6**, 59–73.

Received 24 November 1995; revised version received 6 June 1996