The Breakdown of Social Contracts

Ken Binmore
Economics Department
University College London
Gower Street
London WC1E 6BT, UK

The Breakdown of Social Contracts[1]

by Ken Binmore

> All animals are equal but some are more equal than others.
>
> Orwell's *Animal Farm*

1 Introduction

Much has been written on revolution, rebellion and civil unrest.[2] Historians offer blow-by-blow accounts of the succession of events that led to the fall of this or that government. Sociologists and political theorists collate the circumstances that have precipitated revolutions in an attempt to find correlations that may tell us something about their common causes. Economists construct models that seek to predict when a revolution will occur by measuring the dissonance between the private aspirations of the citizens and realities of their lives under an oppressive regime. But none of these approaches get at the fundamental properties of human nature that make popular insurrections possible for us. Coups in which one elite replaces another are easy enough to understand, since humans would seem to differ little from other primates in their urge to claw their way up a dominance hierarchy. But what of the historically important uprisings fueled by resentment of the unjust or arbitrary use of power?

In this paper, I try to show how the language of game theory can be used to discuss questions concerning the stability of social contracts. The basic ideas are discussed at greater length in my two-volume work, *Game Theory and the Social Contract* (Binmore [10, 11]).

2 What is a social contract?

Hume [31] makes fun of the idea of an original contract as the basis of the legitimacy of the state. As he observes, the claim that we are morally obliged to obey the written and unwritten laws of the society into which we were born because our ancestors surrendered our natural rights at some ancient conclave is a naive fiction. Harsanyi [25] rejects all social contract theories on similar grounds. As he puts it:

[1]I am grateful to the Economic and Social Research Council and to the Leverhulme Foundation for funding this work through the Centre for Economic Learning and Social Evolution at University College London.

[2]Some typical references are Buchanan [12], Coleman [15], Davies [16], Gurr [22], Kuran [34, 35], Rice [43], Tullock [50, 51], and Zagorin [61].

> People cannot rationally feel committed to keep any contract unless they have *already accepted* a moral code requiring them to keep contracts. Therefore, morality cannot depend on a social contract because contracts obtain all their binding force from a *prior* commitment to morality.

If Harsanyi were right to argue that one cannot be a contractarian without believing that everyone is somehow committed to honor the terms of the social contract, then I would be forced to join Hume in poking fun at social contract theories. But I am not alone in thinking that Harsanyi's characterization of contractarianism is too narrow. Both Gauthier [19] and Mackie [37], for example, offer contractarian readings of Hume. Binmore [10] does the same for Harsanyi [24]. Such readings require that we abandon the quasi-legal sense in which the notion of a social contract has been traditionally understood.

This paper does not argue that members of society have an *a priori* obligation or duty to honor the social contract. On the contrary, the only viable candidates for a social contract are those agreements, implicit or explicit, that police *themselves*. Nothing enforces such a self-policing social contract beyond the enlightened self-interest of those who regard themselves as a party to it. Such duties and obligations as are built into the contract are honored, not because members of society are committed in some way to honor them, but because it is *in the interests* of each individual citizen with the power to disrupt the contract not to do so, unless someone else chooses to act against his own best interests by deviating first. The social contract therefore operates *by consent* and so does not need to rely on any actual or hypothetical enforcement mechanism. In game-theoretic terms, it consists simply of an agreement to coordinate on an *equilibrium*.[3]

When it is suggested that a social contract is no more than a set of common understandings among players acting in their own enlightened self interest, it is natural to react by doubting that anything very sturdy can be erected on such a flimsy foundation. Surely a solidly built structure like the modern state must be firmly based on a rock of moral certitude, and only anarchy can result if everybody just does what takes his fancy? As Gauthier [20, p.1] expresses it in denying Hume [28, p.280]: "Were duty no more than interest, morals would be superfluous".

I believe such objections to be misconceived. Firstly, there are no rocklike moral certitudes that exist prior to society. To adopt a metaphor that sees such moral certitudes as foundation stones is therefore to construct a castle in the air.

[3]There is admittedly sometimes a risk of misunderstanding when the term social contract is identified with the conventional rules that a society uses to coordinate on an equilibrium. It is incongruous, for example, that the common understanding in France that conversations be conducted in French should be called a "contract" Words other than contract—such as compact, covenant, concordat, custom, or convention—might better convey the intention that nobody is to be imagined to have signed a binding document or be subject to external enforcement. Perhaps the best alternative term would be "social consensus". This does not even carry the connotation that those party to it are necessarily aware of the fact. However, it seems to me that, with all its dusty encumbrances, "social contract" is still the only term that signals the name of the game adequately and that, as Gough [21, p.7] confirms, I am not altogether guilty of stepping outside the historical tradition in retaining its use while simultaneously rejecting a quasi-legal interpretation.

Society is more usefully seen as a dynamic organism, and the moral codes that regulate its internal affairs are the conventional understandings which ensure that its constituent parts operate smoothly together when it is in good health. Moreover, the origin of these moral codes is to be looked for in historical theories of biological, social, and political evolution, and not in the works of abstract thinkers no matter how intoxicating the wisdom they distill. Nor is it correct to say that anarchy will necessarily result if everybody "just" does what he wants. A person would be stupid in seeking to achieve a certain end if he ignored the fact that what other people are doing is relevant to the means for achieving that end. Intelligent people will *coordinate* their efforts to achieve their individual goals without necessarily being compelled or coerced by real or imaginary bogeymen.

The extent to which simple implicit agreements to coordinate on an equilibrium can generate high levels of cooperation among populations of egoists is not something that is easy to appreciate in the abstract. That *reciprocity* is the secret has been repeatedly discovered, most recently by the political scientist Axelrod [8] in the eighties and the biologist Trivers [49] in the seventies. However, Hume [29, p.521] had already put his finger on the relevant mechanism three hundred years before:[4]

> . . . I learn to do service to another, without bearing him any real kindness: because I foresee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or others. And accordingly, after I have serv'd him and he is in possession of the advantage arising from my action, he is induc'd to perform his part, as foreseeing the consequences of his refusal.

In spite of all the eighteenth-century sweetness and light, one should take special note of what Hume says about foreseeing the consequences of refusal. The point is that a failure to carry out your side of the arrangement will result in your being *punished*. The punishment may consist of no more than a refusal by the other party to deal with you in future. Or it may be that the punishment consists of having to endure the disapproval of those whose respect is necessary if you are to maintain your current status level in the community. However, nothing excludes more active forms of punishment. In particular, the punishment might be administered by the judiciary, if the services in question are the subject of a legal contract.

At first sight, this last observation seems to contradict the requirement that the conventional arrangements under study be *self-policing*. The appearance of a contradiction arises because one tends to think of the apparatus of the state as somehow existing independently of the game of life that people play. But the laws that societies make are not part of the rules of this game. One *cannot* break the rules of the game of life, but one certainly can break the laws that man invents. Legal rules are nothing more than particularly well-codified conventions. And policemen, judges

[4] Note that he goes beyond simple reciprocity between two individuals. If someone won't scratch my back, a third party may fail to scratch his.

and public executioners do not exist outside society. Those charged with the duty of enforcing the laws that a society formally enacts are themselves only players in the game of life. However high-minded a society's officials may believe themselves to be, the fact is that society would cease to work in the long run if the duties assigned to them were not compatible with their own individual incentives. I am talking now about corruption. And here I don't have so much in mind the conscious form of corruption in which officials take straight bribes for services rendered. I have in mind the long-term and seemingly inevitable process by means of which bureaucracies gradually cease to operate in the interests of those they were designed to serve, and instead end up serving the interests of the bureaucrats themselves.

Game theorists rediscovered Hume's insight that reciprocity is the mainspring of human sociality in the early fifties when characterizing the outcomes that can be supported as equilibria in a repeated game. The result is known as the *folk theorem,* since it was formulated independently by several game theorists in the early fifties (Aumann and Maschler [5]). The theorem tells us that external enforcement is unnecessary to make a collection of Mr Hydes cooperate like Dr Jekylls. It is only necessary that the players be sufficiently patient and that they know they are to interact together for the foreseeable future. The rest can be left to their enlightened self interest. The next section introduces the terminology necessary to operationalize the result for the purposes of this paper.

3 Sustaining Equilibria

Moral philosophers are traditionally classified as deontologists or consequentialists. The former argue that morality lies in doing your duty regardless of the consequences. I believe that the moral absolutes of deontologists are actually intuitions acquired by taking note of the rules one follows when honoring whatever social contract is current. In game-theoretic terms, deontologists emphasize the importance of understanding how to operate the strategies that *sustain* an equilibrium. Consequentialists emphasize the importance of the criteria used to *select* a new equilibrium when the circumstances change.

This section briefly reviews the language that game theory uses in discussing how equilibria are sustained. For a naturalist like myself, this is the kind of language into which one needs to translate deontological theories of the Right in order to evaluate the extent to which they succeed in capturing important aspects of our social contract. Axelrod [8], Binmore [10, 11], Schotter [45], Skyrms [47] and Sugden [48] are among those who have written accessible books that popularize the relevant ideas. Schelling [44], Lewis [36] and Ulmann-Margalit [52] say similar things less formally.

The idea of a *Nash equilibrium* is the most fundamental notion of game theory. It is a strategy profile that assigns a strategy to each player that is an optimal reply to the strategies it assigns to the other players. A book that claimed to be an authoritative guide on rational play in games would necessarily have to recommend

a Nash equilibrium in each game for which it made a specific recommendation. Otherwise, at least one player would choose not to follow the book's advice if he believed that that the other players were planning to play as advised. Since everyone would figure this out in advance, the book's claim to be authoritative could therefore not be sustained.

The game Chicken of Figure 1(b) has two players, Adam and Eve. Adam chooses one of the two rows and Eve chooses one of the two columns. Adam's payoffs lie in the southwest of each cell and Eve's lie in the northwest. A *pure strategy* for a player in a game is a complete description of what the player plans to do whenever he or she might be called upon to make a decision. In Chicken, a player has to choose only between *dove* and *hawk*These are therefore his pure strategies. A *mixed strategy* arises when a player randomizes over his pure strategies—perhaps by tossing a coin or rolling dice.

	dove	*hawk*
dove	2 2	3 0
hawk	0 3	1 1

(a) Prisoners' Dilemma

	dove	*hawk*
dove	1 1	2 0
hawk	0 2	−1 −1

(b) Chicken

	dove	*hawk*
dove	−1 −1	2 0
hawk	0 2	−1 −1

(c) Battle of the Sexes

Figure 1: Canonical one-shot games

Chicken has three Nash equilibria. Two of these are pure-strategy equilibria. The third is a mixed equilibrium. The Nash equilibria in pure strategies for Chicken are $(dove, hawk)$ and $(hawk, dove)$. To see this in the case of $(dove, hawk)$, notice that $hawk$ is the best reply for Eve to Adam's choice of $dove$, while $dove$ is simultaneously the best reply for Adam to Eve's choice of $hawk$. Neither player would therefore have a motive to deviate from a book that recommended $(dove, hawk)$ for Chicken, unless there was reason to suppose that the opponent might also deviate. The mixed strategy for Chicken calls for both players to use each of their two pure strategies with probability $\frac{1}{2}$. If Eve plays this way, $dove$ and $hawk$ are equally good for Adam. Adam therefore does not care which he uses, and so he might as well play each with probability $\frac{1}{2}$. If he does so, then the same reasons show that Eve might as well play each of $dove$ and $hawk$ with probability $\frac{1}{2}$. Both Adam and Eve will then be making a best reply to the choice made by the other.

In spite of all the huffing and puffing to the contrary, the only possible outcome of rational play in the Prisoners' Dilemma of Figure 1(a) is the unique Nash equilibrium $(hawk, hawk)$ (Binmore [10]). If the human Game of Life were the one-

shot Prisoners' Dilemma, we therefore would not have evolved as social animals. Fortunately, our Game of Life is better modeled as a repeated game.

In the *indefinitely repeated* Prisoners' Dilemma, Adam and Eve play the Prisoners' Dilemma over and over again until some random event intervenes to bring their relationship to an end. The random event is modeled by postulating that each time they finish playing a round of the Prisoners' Dilemma, there is a fixed probability p that they will never play again. Interest then centers on the case when p is very small, so that the players will have good reason to believe that they have a long-term relationship to nourish and preserve.

The strategic considerations in an indefinitely repeated game are totally different from those in a one-shot game, because the introduction of time permits the players to reward and punish their opponents for their behavior in the past.

Figure 2(a) is based on the the Prisoners' Dilemma of Figure 1(a). It shows the four payoff pairs $(2, 2)$, $(0, 3)$, $(3, 0)$ and $(1, 1)$ that can result if the players restrict themselves to pure strategies in the one-shot version of the game. The broken line encloses the convex hull[5] of these four points. The unique Nash equilibrium outcome for the one-shot Prisoners' Dilemma is located at $(1, 1)$. To facilitate comparison, the equilibrium outcomes for the repeated Prisoners' Dilemma are given on a per-game basis. Thus the points in the shaded region R of Figure 2(a) show the long-run *average* payoffs corresponding to all Nash equilibria in the indefinitely repeated Prisoners' Dilemma in the case when the probability p that any repetition is the last becomes vanishingly small.

Notice that R is a *large* set because the indefinitely repeated Prisoners' Dilemma has *many* Nash equilibria. One of these calls for Adam and Eve to use *hawk* at every repetition of the game. Repeating the game does not therefore guarantee that rational players will cooperate. But neither is the possibility of rational cooperation excluded, since $(2, 2)$ is also a member of the set R of equilibrium outcomes.

It is easy to verify that $(2, 2)$ is a Nash equilibrium outcome for the indefinitely repeated Prisoners' Dilemma. Consider the strategy TIT-FOR-TAT. This calls for a player to begin by using *dove* in the repeated game and then to copy whatever move the opponent made at the previous stage. If Adam and Eve both stick with TIT-FOR-TAT, *dove* will get played all the time. Moreover, it is a Nash equilibrium for both players to stick with TIT-FOR-TAT.

To see this, it is necessary to check that neither can profit from deviating from TIT-FOR-TAT if the other does not. Suppose it is Adam who deviates by playing *hawk* at some stage. Eve does not deviate, and hence she continues by copying Adam. Eve therefore plays *hawk* in later stages until Adam signals his repentance by switching back to *dove*. If p is sufficiently close to 0, Adam's income stream during

[5] The convex hull of a set S is the smallest convex set containing S. In the current context, it represents the set of all pairs of expected payoffs that can result if the players *jointly* randomize over their pure strategies. For example, Adam and Eve can organize the pair of expected payoffs $(1, 2)$ by agreeing to play $(hawk, dove)$ with probability $\frac{1}{3}$ and $(dove, hawk)$ with probability $\frac{2}{3}$. Adam will then expect $1 = \frac{1}{3} \times 3 + \frac{2}{3} \times 0$ and Eve will expect $2 = \frac{1}{3} \times 0 + \frac{2}{3} \times 3$.

Figure 2: Equilibrium outcomes in repeated games

his period of deviance will be approximately $3, 1, 1, \ldots, 0$ instead of $2, 2, 2, \ldots, 2$. It follows that his deviation will have been unprofitable. The TIT-FOR-TAT strategy therefore has a built-in provision for punishing deviations. If both believe that the other is planning to use TIT-FOR-TAT, neither will have a motive for using an alternative strategy. Thus it is a Nash equilbrium for both players to stick with TIT-FOR-TAT.[6]

Figures 2(b) and 2(c) show how the preceding discussion for the Prisoners' Dilemma needs to be modified for the games Chicken and Battle of the Sexes given in Figures 1(b) and 1(c). These diagrams show that the fine structure of a game that is to be repeated indefinitely is often largely irrelevant. Once attention has been directed away from the infertile one-shot case, the question ceases to be *whether* rational cooperation is possible. Instead, one is faced with a bewildering variety of different ways in which the players can cooperate rationally, and the problem becomes that of deciding *which* of all the feasible ways of cooperating should be selected.

[6]The strategy TIT-FOR-TAT has been used to illustrate this point because Axelrod [8] emphasized this particular strategy in his influential *Evolution of Cooperation*. However, there are many other symmetric Nash equilibria that lead to the cooperative outcome $(2, 2)$, some of which are at least as worthy of attention as TIT-FOR-TAT.

This observation puts the question of what is the "right" game to serve as a paradigm for the problem of human cooperation on the sidelines. Once it is appreciated that reciprocity is the mechanism that makes things work, it becomes clear that it is the *fact* of repetition that really matters. The structure of the game that is repeated is only of secondary importance.

4 Selecting Equilibria

Deontological theories of the Right focus on the rules that must be followed to sustain an equilibrium. Consequentialist theories of the Good focus on the mechanisms that have evolved along with the human species to move a society from one social contract to another when the circumstances change. I believe that we are unique among social species in having available two separate and distinct equilibrium selection devices at our disposal. The first is to turn the choice of an equilibrium over to a leader. This solution to the equilibrium selection problem creates a society organized along the same lines as the dominance hierarchies of chimpanzees or baboons. The second mechanism is to use *fairness* as a coordinating device in the manner still practiced by the hunter-gatherer societies that continue to occupy marginal habitats around the world.

This paper suggests that the fundamental cause of popular uprisings lies in the tension that exists between these two rival devices for selecting among equilibria in our Game of Life. Maryanski and Turner [38] go further when they suggest that mankind is doomed to live a life of frustration inside the "social cage" created when we found a way round our genetic disposition to coordinate using fairness norms and began instead to recognize the authority of leaders. My own view is more optimistic, since the institutions of democracy provide a means of reconciling the two equilibrium selection devices. On the other hand, the social contracts of societies whose leaders are not held in check by constitutional checks and balances are always at risk of collapse if the equilibrium selected by the leadership diverges too far from the equilibrium perceived as fair by the rank and file.

I believe that retelling this old idea in the language of game theory may perhaps make it possible to construct models that quantify the relevant phenomena. To this end, the next two sections seek to tie down what is involved in using leadership or fairness as equilibrium selection devices,

5 Leadership and Authority

As explained in the next section, modern foraging societies have no bosses. Moreover, their social contracts are equipped with mechanisms that are designed to inhibit the emergence of bosses. My guess is that these social mechanisms exist because such subsistence societies cannot afford to take the risk of allowing a reformer to persuade them to experiment with their traditional survival techhniques.

But the immediate point is that the existence of such leaderless societies implies that humans do not need bosses to live in societies. So why do we have them? What is the source of their authority?

One popular argument holds that leaders are necessary because, like Uncle Joe Stalin, they know what is good for us better than we know ourselves. But whether leaders know what they are doing better than their followers or not, they can be very useful to a society as a coordinating device for solving the equilibrium selection problem in games for which the traditional methods are too slow or uncertain. In a sailing ship in a storm or in a nation at war, one cannot afford to wait for due process to generate a compromise acceptable to all. Henry Ford told us that history is bunk, but at least it teaches us that the way to get a society moving together in a crisis is to delegate authority to a single leader.

The mention of authority may make it seem that one cannot discuss leadership without stepping outside the class of phenomena that can be explained by the folk theorem. But the authority of a leader does not need to be founded in some theory of the divine right of kings, or in a Hobbesian social contract theory in which citizens surrender their rights to self-determination in return for security, or in some metaphysical argument purporting to prove it rational to subordinate one's own desires to the general will as perceived by the head of state. As Hume [30] puts it:

> Nothing appears more surprising to those who consider human affairs with a philosophical eye, than the ease with which the many are governed by the few, and the implicit submission with which men resign their own sentiments and passions to those of their rulers. When we inquire by what means this wonder is effected, we shall find that, as Force is always on the side of the governed, the governors have nothing to support them but opinion. It is therefore on opinion only that government is founded, and this maxim extends to the most despotic and most military governments as well as to the most free and most popular.

In short, the authority of popes, presidents, kings, judges, policemen and the like is just a matter of convention and habit. Adam obeys the king because such is the custom—and the custom survives because the king will order Eve to punish Adam if he fails to obey. But why does Eve obey the order to punish Adam? In brief, who guards the guardians?

Kant [32, p.417] absurdly thought that to answer this question is necessarily to initiate an infinite regress, but the proof of the folk theorem is explicit in *closing* the chains of responsibility. Eve obeys because she fears that the king will otherwise order Ichabod to punish her. Ichabod obeys because he fears that the king will otherwise order Adam to punish him. The game theory answer to *quis custodiet ipsos custodes?* is therefore that we must all guard each other by acting as official or unofficial policemen in keeping tabs on our neighbors. In the particularly simple case of a society with only two persons, Adam and Eve tread the strait and narrow path of rectitude because both fear incurring the wrath of the other.

To take a crude example, if Adam is the leader in the one-shot Battle of the Sexes of Figure 1(c), he could play fair by tossing a coin to decide which of the two pure Nash equilibria to nominate. If he nominates $(hawk, dove)$ and Eve believes that he will therefore play $hawk$, it is optimal for her to play $dove$. Similarly, if he believes that his nomination of $(hawk, dove)$ will induce her to play $dove$, then it is optimal for him to play $hawk$. A similar convergence of expectations applies if he nominates $(dove, hawk)$. But experience strongly suggests that the opportunities for Adam to abuse his position of authority by cheating are too tempting to resist. Over time he will learn to bias the coin in his favor, so that the equilibrium $(hawk, dove)$ is chosen more often than the equilibrium $(dove, hawk)$. Eventually, he or his successors will convince themselves that they have a right to choose the equilibrium $(hawk, dove)$ all the time.[7] Justice is therefore always a rare commodity in authoritarian states.

6 Fairness

Knauft [33] argues that the evolution of authority in human societies can be seen in terms of a U-shaped curve, in which dominance-structured prehuman societies give way to anarchic bands of human hunter-gatherers that were then replaced by the authoritarian herding and agricultural societies with which recorded history begins. As Erdal and Whiten [17] document, the evidence is strong that leadership in modern hunter-gatherer societies lies only in influencing the consensus: "But when a consensus has been reached, no-one has to follow it against their will—there is no enforcement mechanism."

The fact that modern hunter-gatherers operate social mechanisms that prevent potentially authoritarian leaders from getting established does not imply that their societies do not enforce norms. On the contrary, the evidence is that the social contract operated by a hunter-gatherer community is enforced with a rod of iron. No individual exercizes a leadership role, but the relatively small size of a hunter-gatherer band makes it possible for $public\ opinion$ to fulfill the same function. When Adam asks himself whether he should offer some of his meat to Eve, he knows very well that he will be relentlessly mocked and ridiculed by the band as a whole should he fail to share in the customary fashion. Full-scale ostracism would follow if he nevertheless persisted in behaving unfairly.

Reports that modern hunter-gatherer communities share on a quasi-utilitarian basis are consistent with the view that public opinion serves as a substitute for a leader in such societies, but it is hard to share the enthusiasm expressed by some anthropologists for the oppressive social mechanisms by which discipline is maintained. Envy is endemic. For example, among the !Kung of the Kalahari

[7] Economists should note that it is only by accident that $(hawk, dove)$ happens to be the so-called Stackelberg equilibrium of the Battle of the Sexes. But Adam is not a leader in the Stackelberg sense—he doesn't publically make his move $before$ Eve, leaving her with a take-it-or-leave-it problem. His initial choice of an equilibrium is merely a signal that commits nobody to anything.

desert, nobody cares to keep a particularly fine tool for too long. It is passed along to someone else as a gift lest the owner be thought to be getting above himself. But such gifts do not come without strings. In due course, a fair return will be expected. In some foraging societies, the close attention to the accountancy of envy in such a social contract makes progress almost impossible. According to Hayek's [26, p.153] definition, the citizens of such a society are free because they are subject to no man's will, but it would be a bad mistake for libertarians to idolize such societies. They would do better as a role model for the socialist utopia that Marx envisaged would emerge after the apparatus of the state had withered away.

My *Game Theory and the Social Conract* speculates at length on the reasons that fairness evolved as a coordinating device among out hunter-gatherer ancestors (Binmore [11]). I argue that the foraging bands of prehistory must have operated a much less tightly organized social contract than their modern counterparts. In particular, the freedom enjoyed by subgroups to strike out on their own in a world largely empty of humans makes it unlikely that public opinion could have been an effective weapon for punishing deviates. Food-sharing arrangements must therefore have been self-policing. If one takes this seriously, the same arguments that lead one to predict that a modern foraging society will share on a quasi-utilitarian basis suggest that prehistoric bands must have shared on a quasi-egalitarian basis that can be modeled in terms of the proportional bargaining solution of cooperative game theory. This conclusion matches encouragingly with empirical work from psychology on attitudes to justice.

It now seems to be almost uncontroversial that we are born with the deep structure of language wired into our brains. I think that the same is true of the deep structure of our sense of justice. This would explain the (limited) success that psychologists have enjoyed in predicting that problems of social exchange will be resolved by equalizing the ratio of each person's gain to his worth.[8] As in Wilson [60], the theory is usually called "modern" equity theory, although it originates with Aristotle[9] [4], and has been neglected in recent years after being introduced to social psychologists by Homans [27] and Adams [1, 2] more than thirty years ago.

Although modeling the manner in which fairness norms work is of the first importance, this paper can do no more than register the fact that theories exist that are compatible with what is known of our evolutionary history and which enjoy some empirical support. It is necessary to move on to the question of how the fair social contracts of our foraging ancestors were displaced by the authoritarian social contracts of traditional farming societies.

Cohen [13] attributes the origins of agriculture to a food crisis in prehistory that

[8] See, for example, Adams and Freedman [3] Austin and Hatfield [6], Austin and Walster [7], Baron [9], Cohen and Greenberg [14], Furby [18], Mellers [39], Mellers and Baron [40], Messick and Cook [41], Pritchard [42], Wagstaff *et al* [53, 56, 55], Walster *et al* [57, 58, 59]. Wagstaff [54] has a user-friendly book in draft that sets the philosophical scene, and reviews the history and current status of modern equity theory. Selten [46] provides an account of the theory which is easily accessible to economists.

[9] What is just ... is what is proportional—*Nichemachean Ethics*.

arose when human hunter-gatherer bands had expanded until the locally available habitat was no longer able to support their economies. The response to this over-population problem was twofold. One adaptation allowed foraging to continue in marginal habitats through the use of tightly organized social contracts in which population-size is kept under firm control, as in modern hunter-gatherer societies. The other proved to be the mainstream cultural adaptation: the emergence of agriculture and herding as new modes of production.

The organization necessary both to exploit the increasing returns to scale available in these new modes of production and to prevent the surplus from being appropriated by outsiders, made it necessary to abandon the anarchic structure of prehistoric foraging bands. Instead authority began to be vested in leaders. This readoption of the hierarchical organization typical of ape societies did not require a new set of biological adaptations. We did not lose our capacity to submit to leadership when we acquired the new program that permitted our protohuman ancestors the flexibility necessary to sustain the anarchic life-style of hunter-gatherers with a whole world into which to expand. Even in the uncompetitive ambience of a modern foraging society, our natural urge to dominate one another is not extinguished by our natural urge to be fair. Otherwise social mechanisms that inhibit dominance behavior would not be necessary.

Anthropologists attribute the social retooling necessary for the transition back to the type of hierarchical social contract needed to maintain a communal farming society to *cultural* evolution. The time available seems too short for a further *biological* adaptation to have been responsible.

It has been argued that the human species paid a heavy price for the opportunity to become farmers. When social evolution erected an authoritarian superstructure on a biological foundation that had evolved to permit our ancestors to live a free-wheeling leaderless existence, a war began between part of our biological nature and our social conditioning. Commentators like Maryanski and Turner [38] believe that we are still fighting this war. In the language of game theory, their characterization of a modern industrial society as a social cage is expressed by saying that our habituated use of leadership as an equilibrium selection device conflicts with our natural instinct to employ fairness for this purpose.[10]

7 Deselecting Equilibria

This section turns to the relative stability of authoritarian and egalitarian social contracts. At first sight, game theory would seem to have nothing to say on this subject, since both types of social contract can be realized as Nash equilibria of our

[10] My guess is that we succeed in tolerating leaders by inventing the social fiction that they are responsible as individuals for the capabilities of the groups they coordinate. The worthiness that would be attributed to the group if it were a person is then conferred on its leader. His claim to more than his fair share is thereby rationalized away. But maintaining such a charade is endemically stressful.

Game of Life. However, the notion of a Nash equilibrium fails to capture the possibility that a whole group of individuals may succeed in orchestrating a simultaneous deviation. One might seek to deal with such an eventuality by introducing one of the various definitions of a "coalition-proof equilibrium" that have been proposed. But such equilibria typically fail to exist. That is to say, any social contract can be overthrown if the right kind of coalition can get its act together. However, it is necessary to bear in mind that a revolutionary coalition is itself a society that needs a social contract of its own if its members are to make their collective power felt. This social contract typically uses our built-in sense of justice as its coordinating focus.

Recall that I follow the anthropological line that sees the emergence of hierarchical human societies as a social adaptation required by the need to turn to farming in order to cope with increasing population pressure. But such authoritarian forms of social organization have to operate within the framework of a biologically determined fairness norm that evolved as a coordinating tool among the leaderless foraging bands of prehistory. A tension therefore exists between the conventional authority of leaders and the instinctive urge to coordinate using fairness criteria.

The instabilities created by our failure to be properly adapted to authoritarian social cages are not detected by the folk theorem, because the folk theorem depends on a notion of equilibrium that only considers deviations by one individual at a time. However, leaders who are too partial in choosing an equilibrium that favors the group that put them in power, risk creating a coalition for mutual protection amongst those they treat unfairly. Such an alienated group will treat the *nomenclatura* as outsiders and coordinate on a rebellious equilibrium of its own, in which the first tenet is never to assist a boss in punishing one's own kind. Alexander Hamilton [23, p.3] explains how demagogues take over the leadership of such alienated groups by temporarily facilitating coordination on the fair equilibrium that serves as its focus. But if such demagogues are propelled into power, history shows that they soon become as corrupt as the tyrants they replace.

The recent revolution in Zaire exhibits all the typical hallmarks of a classical popular uprising. It remains to be seen how long it takes for the new elite to become indistinguishable from their predecessors. But the western democracies have no grounds for complacency. It is true that we have evolved institutions that can take the sting out of the incipient conflict between the rival coordination mechanisms of fairness and leadership. First, we constrain our leaders by a system of checks and balances that are intended to limit the extent to which leaders can act without some measure of popular support. Second, the constitution provides regular opportunities for switching leaders. But although these measures can have the effect of preventing leaders taking a society to a markedly unfair equilibrium and keeping it there, they achieve this aim only haphazardly. Nor is it clear that matters are improving. American Presidents are nowadays far more powerful than they were ever intended to be. In the United Kingdom, Mrs Thatcher made it clear that British Prime Ministers have even greater opportunities to ride roughshod over constitutional arrangements that were originally intended to check the power of a

monarch who is now a mere figurehead.

8 Classifying Political Attitudes

The reflections offered in this paper on the stability of the social contracts that our all-too-human nature permits also suggest a radical revision of the classical taxonomy of political systems. Rather than seeing political philosophy in terms of a battle between a utilitarian left and a libertarian right, it seems to me that we need to think in terms of a battle between social contracts based on the authority of individuals or elites and those based on fairness norms. I refer to the former as $neofeudal$ and the latter as $whiggish$.[11] In adopting this terminology, I do not mean to suggest that even a whiggish society can dispense with leaders altogether under modern circumstances. Without entrepreneurs, we would never find the Pareto-frontier of the set of feasible social contracts. Nor is due process appropriate when quick decisions need to be made. But a fair society needs to hold its leaders in check—as the founding fathers of the American Republic knew only too well when they wrote its constitution.

It seems to me that one needs at least two dimensions to come anywhere near capturing the richness of current political attitudes. Figure 3(a) uses two axes to separate the plane into four regions that I could untendentiously have labeled unplanned centralization, unplanned decentralization, planned decentralization and planned centralization. But the language of economics is so dismally dull that I have translated these terms into neofeudalism, libertarianism, whiggery and utilitarianism. A journalist would go further down this road and interpret a utilitarian as a bleeding-heart, big-spending liberal and so on, but I prefer to keep my prejudices under slightly firmer control.

In terms of the traditional left-right political spectrum, utilitarianism sits out on the socialist left and libertarianism sits out on the capitalist right. However, the orthogonal opposition that I think should supercede the sterile and outdated dispute between left and right contrasts whiggish societies in which fairness is used to coordinate collective decisions with neofeudal societies that delegate such decisions to individuals or elites. All large organized states of historical times have been neofeudal in character—including those that think of their prime characteristic as being capitalist or socialist. As is obvious from the totalitarian regimes operating before the Second World War in Germany and Japan, and the species of social consensus that has operated so far in both countries since the war, capitalism does not need libertarian political institutions to flourish. Equally, as is shown by the experience of Britain after the Second World War, a country can ruin its economy by turning to socialism without any need to abandon freedom and democracy along the way. One therefore goes astray in seeking to draw conclusions about the relative

[11] In honor of the authors of the Glorious Revolution of 1688 and its culmination in the American War of Independence.

Figure 3: Classifying political attitudes.

merits of the political aspects of the social contracts operating on the two sides of the Iron Curtain in the cold war from the fact that market economies outperform command economies. Insofar as the choice is between the political regimes of living memory, the issue for political philosophers is not how we organize our economies, but whether our children will find more fulfilment in a society in which the same oligarchs rule all the time, or in a society that rotates its oligarchs using a method traditionally regarded as fair.

A classification system that emphasizes the contrast between neofeudalism and whiggery puts corruption at the top of the agenda of problems facing the modern state. Everyone is willing to condemn the straightforward bribery and nepotism which afflict all societies, but we need to worry at least as much about the corruption of our institutions that arises when their officers cease to operate the institutions to further the purpose for which they were created, but instead imperceptibly and unknowingly come to use the power of their office to advance their own personal goals. To what extent are democratic elections fair now that rich men have learned how to get their puppets elected by employing spin doctors to reduce political debate to an exchange of meaningless advertising slogans? How is justice to be obtained in a law case when the other side has all the money? What poor man seeking his legal rights now expects due process to be respected by the various Jacks-in-office who gnaw at the heart of our public institutions? Our institutions were mostly founded with the most benevolent of intentions, but good intentions

are not enough to prevent time unraveling the firmest weave if loose ends are left at which it can tug. As the Monty Python catchphrase has it. "Nobody expects the Spanish Inquisition." But this is what corruption eventually made of the institution set up to spread the gospel of Jesus Christ.

Utilitarians and libertarians are as dissatisfied as whigs with our current neofeudal institutions, but respond by making the same mistake as Santayana's *Lucifer*, who rebelled against the *feasibility* constraint in God's decision to create the best of all possible worlds. But there is no point in designing ideal social systems whose workability depends on first changing human nature. Human nature is as it is, and no amount of wishing that it were different will make it so. We therefore have to resign ourselves to living in a second-best society because first-best societies are not stable. Utopians who seek to establish first-best societies are actually condemning us to live in whatever hell evolution eventually makes of their unstable utopia.

The founding fathers of the American Republic understood this point perfectly well when they built a system of checks and balances into the American Constitution in an attempt to confine neofeudalism to the Old World. But what remains of their construction is now hopelessly unfitted to meet the new forms of neofeudalism that have emerged in modern times. My own country does not even have a written Constitution or a Bill of Rights to obstruct the triumphal advance of neofeudalism. If we do not wish to destabilize our societies by creating a new class of outsiders who can focus their resentment around a call for natural justice, we need to rethink the thoughts of the classical liberals who wrote the American Constitution, as they would rethink them if they were alive today.

In summary, we need to put aside outdated thoughts about where we would like to locate society on a left-right spectrum. Choosing between utilitarianism and libertarianism makes as much sense as debating whether griffins make better pets than unicorns. We need to start thinking instead about how to move in the orthogonal direction that leads from neofeudalism to whiggery.

References

[1] J. Adams. Towards an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67:422–436, 1963.

[2] J. Adams. Inequity in social exchange. In L. Berkowitz, editor, *Advances in Experimental Social Science, Volume II*. Academic Press, New York, 1965.

[3] J. Adams and S. Freedman. Equity theory revisited: Comments and annotated bibiliography. In L. Berkowitz, editor, *Advances in Experminental Social Science, Volume IX*. Academic Press, New York, 1976.

[4] Aristotle. *Nichomachean Ethics*. Hackett, Indianopolis, 1985. (Translated by T. Irwin).

[5] R. Aumann and M. Maschler. *Repeated Games with Incomplete Information*. MIT Press, Cambridge, MA, 1995.

[6] W. Austin and E. Hatfield. Equity theory, power and social justice. In G. Mikula, editor, *Justice and Social Interaction*. Springer-Verlag, New York, 1980.

[7] W. Austin and E. Walster. Reactions to confirmations and disconfirmations of expectancies of equity and inequity. *Journal of Personality and Social Psychology*, 30:208–216, 1974.

[8] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[9] J. Baron. Heuristics and biases in equity judgments: A utilitarian approach. In B. Mellors and J. Baron, editors, *Psychological Perspectives on Justice: Theory and Applications*. Cambridge University Press, Cambridge, 1993.

[10] K. Binmore. *Playing Fair: Game Theory and the Social Contract I*. MIT Press, Cambridge, MA, 1994.

[11] K. Binmore. *Just Playing: Game Theory and the Social Contract II*. MIT Press, Cambridge, MA, 1998. (forthcoming).

[12] A. Buchanan. Revolutionary motivation and rationality. *Philosophy and Public Affairs*, 9:59–82, 1979.

[13] M. Cohen. *The Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture*. Yale University Press, New Haven, 1977.

[14] R. Cohen and J. Greenberg. The justice concept in social psychology. In R. Cohen and J. Greenberg, editors, *Equity and Justice in Social Behavior*. Academic Press, New York, 1982.

[15] J. Coleman. *Against the State: Studies in Sedition and Rebellion*. Penguin Books, London, 1990.

[16] J. Davies. Toward a theory of revolution. *American Sociological Review*, 27:5–19, 1962.

[17] D. Erdal and A. Whiten. Egalitarianism and Machiavellian intelligence in human evolution. In P. Mellars and K. Gibson, editors, *Modelling the Early Human Mind*. Oxbow Books, Oxford, 1996.

[18] L. Furby. Psychology and justice. In R. Cohen, editor, *Justice: Views from the Social Sciences*. Harvard University Press, Cambridge, MA, 1986.

[19] D. Gauthier. David Hume: Contractarian. *Philosophical Review*, 88:3–38, 1979.

[20] D. Gauthier. *Morals by Agreement*. Clarendon Press, Oxford, 1986.

[21] J. W. Gough. *The Social Contract*. Clarendon Press, Oxford, 1938.

[22] T. Gurr. *Why Men Rebel*. Princeton University Press, Princeton, 1970.

[23] A. Hamilton, J. Jay, and J. Madison. *The Federalist*. Everyman, London, 1992. (Edited by W. Brock. First published 1787–1788).

[24] J. Harsanyi. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, Cambridge, 1977.

[25] J. Harsanyi. Review of Gauthier's "Morals by Agreement". *Economics and Philosophy*, 3:339–343, 1987.

[26] F. Hayek. *The Constitution of Liberty*. University of Chicago Press, Chicago, 1960.

[27] G. Homans. *Social Behavior: Its Elementary Forms*. Hartcourt, Brace and World, New York, 1961.

[28] D. Hume. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. 3rd edition. Clarendon Press, Oxford, 1975. (Edited by L. A. Selby-Bigge. Revised by P. Nidditch. First published 1777).

[29] D. Hume. *A Treatise of Human Nature (Second Edition)*. Clarendon Press, Oxford, 1978. (Edited by L. A. Selby-Bigge. Revised by P. Nidditch. First published 1739).

[30] D. Hume. Of the first principles of government. In *Essays Moral, Political and Literary, Part I*. Liberty Classics, Indianapolis, 1985. (Edited by E. Miller. Essay first published 1758).

[31] D. Hume. Of the original contract. In *Essays Moral, Political and Literary*. Liberty Classics, Indianapolis, 1985. (Edited by E. Miller. Essay first published 1748).

[32] I. Kant. Theory and practice. In *The Philosophy of Kant*. Random House, New York, 1949. (Edited by C. Friedrich. First published 1793).

[33] B. Knauft. Violence and sociality in human evolution. *Current Anthropology*, 32:223–245, 1991.

[34] T. Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public Choice*, 61:41–71, 1989.

[35] T. Kuran. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press, Cambridge MA, 1995.

[36] D. Lewis. *Conventions: A Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.

[37] J. Mackie. *Hume's Moral Theory*. Routledge and Kegan Paul, London, 1980.

[38] A. Maryanski and J. Turner. *The Social Cage: Human Nature and the Evolution of Society*. Stanford University Press, Stanford, 1992.

[39] B. Mellers. Equity judgment: A revision of Aristotelian views. *Journal of Experimental Biology*, 111:242–270, 1982.

[40] B. Mellers and J. Baron. *Psychological Perspectives on Justice: Theory and Applications*. Cambridge University Press, Cambridge, 1993.

[41] D. Messick and K. Cook. *Equity Theory: Psychological and Sociological Perspectives*. Praeger, New York, 1983.

[42] R. Pritchard. Equity theory; A review and critique. *Organizational Behavior and Human Performance*, 4:176–211, 1969.

[43] E. Rice. *Revolution and Counter Revolution*. Blackwell, Oxford, 1991.

[44] T. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, 1960.

[45] A. Schotter. *The Economic Theory of Social Institutions*. Cambridge University Press, Cambridge, 1981.

[46] R. Selten. The chain-store paradox. *Theory and Decision*, 9:127–159, 1978.

[47] B. Skyrms. *Evolution of the Social Contract*. Cambridge University Press, Cambridge, 1996.

[48] R. Sugden. *The Economics of Rights, Cooperation and Welfare*. Basil Blackwell, Oxford, 1986.

[49] R. Trivers. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46:35–56, 1971.

[50] G. Tullock. *The Social Dilemma: The Economics of War and Revolution*. University Publications, Blacksberg VA, 1974.

[51] G. Tullock. *Autocracy*. Kluwer, Dordrecht, 1987.

[52] E. Ulmann-Margalit. *The Emergence of Norms*. Oxford University Press, New York, 1977.

[53] G. Wagstaff. Equity, equality and need: Three principles of justice or one? *Current Psychology: Research and Reviews*, 13:138–152, 1994.

[54] G. Wagstaff. Making sense of justice. Draft book, Psychology Department, University of Liverpool, 1997.

[55] G. Wagstaff, J. Huggins, and T. Perfect. Equal ratio equity, general linear equity and framing effects in judgments of allocation divisions. *European Journal of Social Psychology*, 26:29–41, 1996.

[56] G. Wagstaff and T. Perfect. On the definition of perfect equity and the prediction of inequity. *British Journal of Social Psychology*, 31:69–77, 1992.

[57] E. Walster, E. Berscheid, and G. Walster. New directions in equity research. *Journal of Personality and Social Psychology*, 25:151–176, 1973.

[58] E. Walster and G. Walster. Equity and social justice. *Journal of Social Issues*, 31:21–43, 1975.

[59] E. Walster, G. Walster, and E. Berscheid. *Equity: Theory and Research.* Allyn and Bacon, London, 1978.

[60] J. Wilson. *The Moral Sense.* Free Press, New York, 1993.

[61] P. Zagorin. Theories of revolution in contemporary historiography. *Political Science Quarterly*, 88:23–52, 1973.