GAMES AND ECONOMIC BEHAVIOR 7, 1–9 (1994)

Group Rationality, Utilitarianism, and Escher's Waterfall

KAUSHIK BASU*

Delhi School of Economics, Delhi - 100 007, India

Received July 18, 1991

Consider a game in which each player chooses between two strategies, H and S, and all players have the same payoff function. This could be, for instance, because all players are moral creatures committed to enhancing a common cause. Is it possible that in this game if each player chooses S instead of H (with other players' strategy choices held constant), he (hence, everybody) is better off but he is worse off if everybody chooses S? It is shown that the answer to this is yes, if the number of players is infinite (even if only countably so). This is demonstrated by constructing a paradoxical game referred to here as the "waterfall" paradox. Some implications of the paradox for models of economics are discussed. *Journal of Economic Literature* Classification Numbers: C70, D71. © 1994 Academic Press, Inc.

1. MOTIVATION

There is a long tradition in linguistic and moral philosophy that claims that language has at least two uses. First, there is the "normal" or "serious" use which allows us to communicate and exchange information. For its normal use what matters is "the meaning of what is said." Habermas (1989, p. 159) has described this as "the original mode of language." Secondly, there is what has been described as the "etiolated" or "parasitical" use (Austin, 1962) whereby language is used by the speaker to *bring*

* I am grateful to Abhijit Banerjee, John Conway, Uli Hege, David Lewis, Robert Pollak, Andrew Postlewaite, Ariel Rubinstein, Ekkehart Schlicht, Siddharth Sahi, and, especially, Debraj Ray for comments and discussion. I have also benefited from seminars at Princeton, Cornell, the Institute for International Economic Studies at Stockholm, and the University of Pennsylvania. Finally, I thank an anonymous referee of the journal for several helpful suggestions.

0899-8256/94 $6.00

about something in the world. I shall here make a related distinction between normal and "honest" speech, on the one hand, and parasitical and "strategic" speech on the other. If a murderer asks me about the whereabouts of a potential victim and I say, "He has gone left," knowing full well that he has gone right, I am using language strategically in order to bring about a certain kind of world. Note that if language was never used normally or honestly, then my saying "He has gone left" would have no effect.

It seems possible therefore to argue that if we made excessive use of language in a parasitical way for strategic purposes it would lose much of its value even as a normal instrument of communication. Now consider a world in which everyone is committed to the same objective function. While the objective function could be of any kind, let us for the sake of an example suppose that all individuals are committed utilitarians; and they practice this to the point where before uttering each sentence they calculate the total utility that will be generated as a consequence of it and then make those statements which generate the most utility. It seems possible to argue that such fastidious adherence to utilitarianism could be harmful for society in terms of utilitarian calculations itself, because, in such a world, language would cease to have power, communication would break down and so would many things which are predicated upon our ability to communicate. Such an argument, where utilitarianism is pitted against utilitarianism, has been made in the literature. Hodgson (1967) is quite explicit on this.

But is it possible to build a formal argument where utilitarianism gets pitted against utilitarianism? The purpose of this paper is to answer this question. It is shown that if the number of decisions to be taken in a society is finite then the answer must be negative. In an infinite-decision society it is possible that each person acting in utilitarian interest everywhere will end up creating a society inferior in utilitarianism terms itself.

Though I began with the case of language and utilitarianism, the point that this paper tries to illustrate is a more general one. Consider a game in which each player can choose from two possible strategies, S and H, and all players have the same payoff function. This could be because all players belong to a team with one common objective or that they are moral creatures all committed to one social welfare function.[1] The question being asked here is whether this game can have a dominant-strategy equilibrium which is Pareto inferior. More specifically, is it possible that when-

[1] It is therefore immediately obvious that my reference to utilitarianism, above, is an illustration of any consequentialist ethical system. It is, however, worth noting that some have considered the distinction between consequentialism, in general, and utilitarianism, in particular, to be blurred. Hammond (1990, p. 2), for instance, argues that "utilitarianism itself can be derived from the even more primitive principle called 'consequentialism'."

ever an individual switches from playing H to S (other players' choices remaining unchanged) he, and therefore everybody, becomes better off but if everybody played S everybody would get a lower payoff than what they would get if everybody played H? The answer to this is yes and this is demonstrated in Section 2 by constructing a game with a "paradoxical" outcome, which will be referred to here as the "Waterfall" paradox. A more difficult problem is to demonstrate this possibility for games which treat players "symmetrically," in a sense made precise below. This is demonstrated in Section 2 under the assumption that the axiom of choice is true. Section 3 discusses some implications of these "Waterfall" games for models of economics.

2. "Waterfall" Games

The result that will be established first is this. If we have an infinite number of players then there exists a game in which everybody has the same objective function (i.e., they may, for instance, be maximizing player 1's utility or all their utilities summed up) but each individial's effort to maximize this objective function ends up with the players in a suboptimal equilibrium, suboptimality being defined in terms of the same objective function.

Let me make this more precise. Suppose N is the set of players or individuals, and Ω is the power set of N. Each player $i \in N$ can play one of two available strategies: H and S. Player i's pay-off function is f_i. That is, f_i: $\Omega \to R$, where R is the set of real numbers. This is to be interpreted as follows. If $A \in \Omega$ is the set of players who choose S (i.e., $N \backslash A$ is the set choosing H), then i's pay-off is $f_i(A)$. Since everything else in the description of a game is constant, in this paper we identify a game with a specification of pay-off functions $\{f_i\}_{i \in N}$.

We shall describe a game, $\{f_i\}_{i \in N}$, as a Prisoner's Dilemma if each player prefers to play S no matter what others play (i.e., S is a dominant strategy for each player), but everybody is worse off if everybody plays S as compared to the case where everybody plays H.

That Prisoner's Dilemma games exist for N finite or infinite is well-known and can be proved by construction. The question with which this paper is concerned is, however, more difficult to answer.

Suppose all players have the same pay-off function. That is, there exists a real-valued function, f, on Ω such that, for all $i \in N$, $f_i = f$. I shall refer to such a game as a *uniform game*. This could happen if all of them are perfectly united in their objective. For instance, they may be card-carrying utilitarians with each person acting so as to increase the sum of their

happiness or they could be a group of firms in an industry, totally united in their pursuit of maximizing joint profits or they may be employees of a firm all striving to maximize the firm's profit, as assumed in most economics textbooks. The first question I want to address is this. Does there exist a uniform game which is a Prisoner's Dilemma?

The answer is yes. I shall construct an example in a moment to demonstrate this. The proof requires that the number of players be infinite. A much more difficult question is: whether there exists a game which is uniform, a Prisoner's Dilemma and symmetric in pay-offs (in a sense to be made precise later)? Even here the answer is "yes" but my proof of this will make use of the axiom of choice.

Given that in the social sciences we often resort to the assumption of an infinite number of agents,[2] these games are not only of intrinsic interest but of relevance to our models.

Note first that with a finite number of agents, no matter how large, there cannot be a game which is uniform and a Prisoner's Dilemma. To see this consider two elements, A and B in Ω such that outcome A implies everyone plays H and B implies everyone plays S. That is, $A = \phi$, $B = N$.

Since there is a finite number of players we can go from A to B by changing H to S for one player at a time. If H is a dominated strategy for each player, every such switch must raise the pay-off (remember in a uniform game all players have the same pay-off function). Hence, $f(B) > f(A)$. So the game cannot be a Prisoner's Dilemma.

Now it will be shown that if N were infinite the above paragraph's claim would not be true. Somewhat surprisingly, this is so even if N is countably infinite. In fact, from here onward, I assume $N = \{1, 2, \ldots\}$; that is, N is the set of all positive integers. In other words it is possible that, beginning with any outcome, as each H is changed to an S, the pay-off rises, but the pay-off from everyone playing S is less than the pay-off from the case where everyone plays H. A picturesque analogus of this is Escher's well-known painting, "Waterfall", in which at each step water keeps flowing down but ends up on top.

Since in a uniform game all players have the same pay-off function, such a game is identified entirely by the pay-off function, $f: \Omega \rightarrow R$. Hence the main proposition being proved is this:

There exists a uniform game, f, such that

[2] In industrial organization theory the model of perfect competition and in general equilibrium theory Walrasian analysis are often rationalized in terms of an infinite number of firms and agents (Aumann, 1964; Hildenbrand and Kirman, 1976). In fact, while these models require a continuum of agents, the paradoxical result described here occurs even with a countably infinite set of agents. I write "even" because, as will be obvious later, my result is easily generalized to the uncountable case, though I work with the countable one.

(1) For all $A \in \Omega$, and for all $i \in N \backslash A$, $f(A \cup \{i\}) > f(A)$ and
(2) $f(N) < f(\phi)$.

This is proved by actually constructing an example. Let $\hat{f}$ be the following pay-off function. If $A \in \Omega$, $\hat{f}(A)$ is defined as follows. If $\#N\backslash A < \infty$ (that is, A is such that the number of players who choose H is finite) then $\hat{f}(A) = -\#N\backslash A$. Next suppose $\#N\backslash A = \infty$. Then $\hat{f}(A)$ is the number (in decimals) formed by writing 1 in the ith place after the decimal if $i \notin A$ (i.e. player i has chosen H), and 2 in the ith place after the decimal if $i \in A$. To avoid misunderstanding, let me state this a little more formally. Given $A \in \Omega$, define

$$x_i(A) = \begin{cases} 2 & \text{if } i \in A \\ 1 & \text{if } i \notin A. \end{cases}$$

If $\#N\backslash A = \infty$, then $\hat{f}(A) = \cdot x_1(A)x_2(A) \ldots$, where $\cdot x_1(A)x_2(A) \cdots$ is a decimal number where the number in the ith place after the decimal is $x_i(A)$.

It follows that, $\hat{f}(N) = 0$ and $\hat{f}(\phi) = \cdot 111 \ldots$. Hence $\hat{f}$ satisfies (2). It is easy to check that it satisfies (1) as well.

Observe that in the game, $\hat{f}$, the impact on the pay-off of a player's decision between H and S depends on who the player is. Player 1, for instance, has a much bigger effect on the pay-off, than player 58. Now, let us suppose that we want the game to be symmetric or anonymous. Some of the applications I discuss in the next section makes anonymity a reasonable property. But a pay-off function can be anonymous in several senses. I shall here consider a weak anonymity property, which will be called finite-anonymity.

Earlier a particular "play" of the game was denoted by a set $A \in \Omega$. There is an alternative characterization according to which a play is a sequence $\{x_i\}_{i \in N} \equiv \{x_i\}$, in which $x_i \in \{H, S\}$, for all i. The jth element in $\{x_i\}$ denotes what player j has chosen. Clearly we could think of a function, h, which for every play, $\{x_j\}$, specifies the set of players who have chosen S. That is $h(\{x_j\}) = \{i \in N | x_i = S\}$. The pay-off from $\{x_i\}$ will be, distorting earlier terminology a little, written as $f(\{x_i\})$. If we were more cautious we would write $f(h(\{x_i\}))$ instead of $f(\{x_j\})$.

Next, we need to define a finite permutation. The mapping $\sigma: N \to N$ is a *finite-permutation* if there exists a finite set $A \subset N$ such that, for all $i \notin A$, $\sigma(i) = i$ and the restriction of σ on A is a permutation (i.e. it is one-to-one and onto on A).

Finally, we say that the uniform game f satisfies *finite-anonymity* if the following is true: If $\{x_i\}$ and $\{y_i\}$ are plays that there exists a finite-permutation, σ, such that $\{x_i\} = \{y_{\sigma(i)}\}$, then $f(\{x_i\}) = f(\{y_i\})$.

The second, and main, paradoxical result is this: There exists a uniform game which satisfies (1), (2), and finite-anonymity.

Interestingly, it does not seem possible to give a constructive proof of this. In other words, what is being claimed is that though I cannot construct such a game, I can prove that such a game must exist.

In order to prove this define the binary relation, $\sim$, on Ω as follows: For all $X, Y \in \Omega$, $X \sim Y$ iff $\#X\backslash Y$ and $\#Y\backslash X$ are both finite. It will now be shown that $\sim$ is an equivalence relation. Its reflexivity and symmetry are obvious. To check that $\sim$ is transitive, suppose $X \sim Y$ and $Y \sim Z$. Hence $\#X\backslash Y$, $\#Y\backslash X$, $\#Y\backslash Z$, and $\#Z\backslash Y$ are finite. It is easy to see that $X\backslash Z \subset (X\backslash Y) \cup (Y\backslash Z)$. Let $x \in X\backslash Z$. If $x \notin Y$, then $x \in X\backslash Y$. Suppose $x \in Y$. Since $x \notin Z$, it follows $x \in Y\backslash Z$.

Hence $\#X\backslash Z$ is finite and by an anlogous proof $\#Z\backslash X$ is finite.

Since $\sim$ is an equivalence relation, there exists a partition P of Ω generated by $\sim$ such that if there exists $P_\alpha \in P$, such that $X, Y \in P_\alpha$ then $X \sim Y$.

Let Q be a set formed by choosing one element from each element of P. The existence of this set is guaranteed by the axiom of choice.

Define g: $Q \to R$ as follows:

Let $A \in Q$. Hence there exists a unique element, P_α, in P such that $A \in P_\alpha$.

(a) If $\phi \notin P_\alpha$, and $N \notin P_\alpha$,	then $g(A) = 1$.
(b) If $\phi \in P_\alpha$,	then $g(A) = \#A + 1$.
(c) If $N \in P_\alpha$,	then $g(A) = -\#(N\backslash A)$.

Note that ϕ and N cannot belong to the same element of the partition. They belong to the collection of, respectively, the finite sets and the cofinite sets. Also, we could have chosen ϕ and N from these sets in constructing Q. In what follows it is worth keeping in mind that A and $A \cup \{i\}$ are always in the same element of the partition.

Now let f^*: $\Omega \to R$ be an extension of g defined as follows. Let $A \in \Omega$. Find a $B \in$ Q such that $A, B \in P_\alpha \in P$, for some P_α. We set $f^*(A) = g(B) - \#(B\backslash A) + \#(A\backslash B)$.

It is easy to check that f* satisfies (1), (2), and finite-anonymity. I shall here demonstrate finite-anonymity, since (1) and (2) are obvious.

Let $x \equiv \{x_i\}$ and $y \equiv \{y_i\}$ be such that there exists a finite-permutation σ of N such that $\{x_i\} = \{y_{\sigma(i)}\}$. Hence it follows that $\#X\backslash Y = \#Y\backslash X < \infty$, where $X \equiv \phi(x)$ and $Y \equiv \phi(y)$. Therefore $X, Y \in P_\alpha \in P$, for some P_α. Let $B \in P_A \cap Q$. It is easy to see that $\#X\backslash B - \#B\backslash X = \#Y\backslash B - \#B\backslash Y$. This follows from the fact that

$$\#X\backslash B + \#B\backslash Y = \#X\backslash Y + \#(X \cap Y)\backslash B + \#B\backslash(X \cup Y), \text{ and}$$

$$\#Y\backslash B + \#B\backslash X = \#Y\backslash X + \#(X \cup Y)\backslash B + \#B\backslash(X \cup Y).$$

Hence, $f^*(X) = f^*(Y)$, thereby establishing that f^* satisfies finite-anonymity.

It may be asked as to why attention is restricted to *finite*-anonymity. This is because it is easy to show that a *fully* anonymous uniform game is incompatible with (1), let alone (1) *and* (2). Define $r: N \to N$ to be a *permutation* if r is one-to-one and onto on N. The uniform game f is *fully anonymous* if whenever $\{x_i\}$ and $\{y_i\}$ are plays such that for some permutation r, $\{x_i\} = \{y_{r(i)}\}$, then $f(\{x_i\}) = f(\{y_i\})$. Now let $\{x_i\}$ and $\{y_i\}$ be such that $x_i = H$ if and only if i is odd and $y_i = x_i$, if $i > 1$, and $y_1 = S$. By (1), $f(\{y_i\}) > f(\{x_i\})$. But note that there exists a permutation r such that $\{x_i\} = \{y_{r(i)}\}$. This is so if

$$\begin{aligned} r(i) = {} & i + 2, && \text{if } i \text{ is odd} \\ & i - 2, && \text{if } i \text{ is even but not 2} \\ & 1, && \text{if } i = 2 \end{aligned}$$

Hence, $f(\{x_i\}) = f(\{y_i\})$, which is a contradiction.

It would be interesting to develop notions of anonymity which lie between the polar extremes of finite and full anonymity and examine their compatibility with (1) and (2).

3. Remarks

What the "Waterfall" game demonstrates is, in some sense, the opposite of the widsom embodied in the well-known invisible hand theory. According to the latter, every individual working in his individual interest may lead to an outcome which is optimal for the group. According to the "Waterfall" game, every individual working in the group's interest may lead to an outcome which is suboptimal for the group. For models of economics the results of Section 2 are important because such models do often assume an infinity of agents. One response to this is to treat the paradoxical "Waterfall" effect as a *reductio ad absurdum* against the idea that, if a population is large and individuals are insignificant, we can safely model the population as infinite.[3] This would amount to a critique of models like that of perfect competition in economics.

There are three remarks in this connection worth keeping in mind. First, for certain kinds of issues, an infinity of agents may not be an unrealistic

[3] This is the position which David Lewis seems to take (personal communication to the author, dated January 15, 1990).

assumption, as long as it is countable. This is because if we consider all people of present and future generations the number may well be infinite.

Secondly, in anonymous games the "Waterfall" effect seems to be a consequence of not just an infinite number of agents but also the axiom of choice. Hence, it may be possible to avert the paradoxical result by foregoing the axiom of choice, and indeed there are some areas of game theory which have tried to do without the axiom.

Finally, there is an interpretation under which the results of Section 2 could be thought of as occurring in models with a finite number of agents. Suppose there are n (finite) agents and each agent i takes t_i decisions. Each decision consists of choosing between S and H. For instance, an agent i might be having t_i points of time between now and next year at each of which he has to decide whether to do something (H) or procrastinate (S) to the next point of time, like in Akerlof's (1991) model. As long as $t_1 + \cdots + t_n$ is infinite, the same construction as in Section 2 is possible. It is important, however, to note that in this case there must exist one agent who controls an infinite number of decisions. That is, for some i, t_i must be infinite. Hence, in this case S being chosen everywhere is not a dominant strategy for player i. However, S being chosen everywhere could be thought of as *decision-wise dominant,* that is, for *each* of the t_i decisions, considered one at a time, S dominates H.

To close with a digression, I turn to a problem in welfare economics on which the "Waterfall" game throws some light. The distinction between act and rule consequentialism has always been considered ambiguous[4] and I have shared in the feeling. *Act* consequentialism requires that a particular act or decision should be undertaken if it brings about a social state which is desirable. *Rule* consequentialism recommends an act, if it is the case that everyone undertaking the act in similar situations leads to a desirable social state. The distinction between the two, however, seems questionable because by making the definition of what constitutes a "similar situation" sufficiently specific we can make rule consequentialism operationally indistinguishable from act consequentialism (see Smart, 1973, pp. 9–12, for discussion).

Our exercise in Section 2 clarifies the conditions under which the two moral systems can be distinguished. If the uniform game is given by the f^* defined above, then act utilitarianism would recommend that individuals use speech strategically (i.e., choose S), whereas rule utilitarianism would clearly not make such a recommendation. If a "rule" is taken to mean either of the following two recommendations: (i) Be strategic (S) and (ii) be honest (H), then rule utilitarianism would recommend honesty. But if

[4] See Smart (1973). His discussion is concerning act and rule *utilitarianism* but the same ideas are easily extended to the more general concept of consequentialism (see Sen, 1985).

we allow rules to be more complicated and take forms like "in the following circumstances, be honest" then rule utilitarianism may make more complicated recommendations. The argument of Section 2 may be seen as pushing us away from act consequentialism of any form, including utilitarianism, towards some form of deontological ethics because it highlights how act consequentialism can be self-defeating. The argument also shows that in some sense rule consequentialism is closer to deontological ethics than act consequentialism.

REFERENCES

AKERLOF, G. (1991), "Procrastination and Obedience," *Amer. Econ. Rev.* **81** (Papers and Proceedings).

AUMANN, R. J. (1964), "Markets with a Continuum of Traders," *Econometrica,* **32**.

AUSTIN, J. L. (1962), *How to do Things with Words*. London/New York: Oxford Univ. Press.

HABERMAS, J. (1989), *On Society and Politics: A Reader*. Boston: Beacon Press.

HAMMOND, P. J. (1990), "Interpersonal Comparisons of Utility: Why and How They are and Should be Made," mimeo, Florence.

HILDENBRAND, W., AND KIRMAN, A. P. (1976), *Introduction to Equilibrium Analysis*. Amsterdam: North-Holland.

HODGSON, D. H. (1967), *Consequences of Utilitarianism*. London/New York: Oxford Univ. Press.

SEN, A. (1985), "Well-Being, Agency and Freedom," *J. Philos.* **82**.

SMART, J. J. C. (1973), "An Outline of a System of Utilitarian Ethics," in *Utilitarianism: For and Against,* (J. C. C. Smart and B. Williams, Eds.). London/New York: Cambridge Univ. Press.