CHAPTER 4

The expected utility of playing a game

Alvin E. Roth

1 Introduction

This chapter is concerned with how the Shapley value can be interpreted as an expected utility function, the consequences of interpreting it in this way, and with what other value functions arise as utility functions representing different preferences.

These questions brought themselves rather forcefully to my attention when I first taught a graduate course in game theory. After introducing utility theory as a way of numerically representing sufficiently regular individual preferences, and explaining which comparisons involving utility functions are meaningful and which are not, I found myself at a loss to explain precisely what comparisons could meaningfully be made using the Shapley value, if it was to be interpreted as a utility as suggested in the first paragraph of Shapley's 1953 paper. In order to state the problem clearly, it will be useful to remark briefly on some of the familiar properties of utility functions.

First, utility functions represent preferences, so individuals with different preferences will have different utility functions. When preferences are measured over risky as well as riskless prospects, individuals who have the same preferences over riskless prospects may nevertheless have different preferences over lotteries, and so may have different expected utility functions.

Second, there are some arbitrary choices involved in specifying a utility function, so the information contained in an individual's utility function is really represented by an equivalence class of functions. When preferences are defined over riskless prospects without any information about relative intensities of preference, then the class of utility functions equivalent to a given utility function u consists of all monotone transformations of u. When preferences are defined over risky prospects as well, then the

class of expected utility functions equivalent to a given expected utility function u consists of all positive linear transformations of u. That is, the (only) arbitrary elements in an expected utility function are the choice of the zero point and unit.

A *meaningful* statement about preferences, in terms of a utility function, must be true for every equivalent utility function. (In just this sense it is not a meaningful statement about temperature to say that water boils at between six and seven times the temperature at which it freezes: This is a statement about the Fahrenheit temperature scale that does not hold in the equivalent Celsius scale.) Similarly, because different individuals' expected utility functions have arbitrary origins and units, they are not comparable. For example, they cannot meaningfully be added. That is, no information about preferences is conveyed by saying that a particular outcome maximizes the sum of the utilities of the players in a game, because this is not independent of the scale of each utility function: If an individual's utility is multiplied by 100 (which yields an equivalent representation of the individual's preferences), the outcome that maximizes the sum of the utilities would not stay the same in general.[1]

The original derivation of the Shapley value does not resemble the derivation of utility functions, in that all conditions are stated directly on the value function, so there is no clear connection to underlying preferences. Hence, the following questions present themselves.

1. *If the Shapley value is to be interpreted as a utility, why is it unique?* Won't different individuals with different preferences and risk postures have different utility functions? If so, what can be said about those preferences for which the Shapley value is a utility function? What will other utility functions for games look like?

2. *What are the meaningful statements about preferences that can be conveyed by the Shapley value?* What are the arbitrary elements in the Shapley value as a utility – what normalization has been chosen? Under what circumstances can the Shapley value of a game be compared to the utility of other kinds of alternatives?

3. *What does the additivity axiom mean?* What statement about preferences is made by a utility function that relates the sum of the utilities of games v and w to the utility of another game, $v + w$?

4. *How can the efficiency axiom be interpreted in the context of a utility function?* It specifies that the values for each position in a game v must sum to $v(N)$: Is there some assumption hidden here that interpersonal comparisons can be made, and that sums of utilities are meaningful? If not, what is the significance of specifying the sum?[2]

To answer these questions, we need to consider preferences over games. The viewpoint I take is that the preferences in question are those of a single individual, faced with choices over positions in a game, and in different games. The resulting utility function can be thought of, like the Shapley value, as a function defined on games that assigns a real number to each position in a game. It turns out that whether such a utility function conforms to the efficiency axiom depends on the attitude of the individual in question to a certain kind of risk, and whether it conforms to the additivity axiom depends on the individual's attitude toward another kind of risk. When the individual is what I call "risk neutral" to both kinds of risk, then his or her expected utility for playing a game is equal to the Shapley value. Other attitudes toward risk yield other utility functions.

This chapter attempts to integrate the material originally presented in Roth (1977a,b,c). Section 2 briefly reviews how an expected utility for an abstract set of alternatives is derived. Sections 3 and 4 then consider how to apply and extend this treatment to include positions in games as alternatives. Section 5 considers the special case of simple games, and may be skipped by those interested only in the main ideas. Section 6 concludes by considering the answers to the questions posed earlier.

2 Utility theory

We summarize here an elegant axiomatization of expected utility developed by Herstein and Milnor (1953). A set M of alternatives is a *mixture set* if for any elements $a, b \in M$ and for any probability $p \in [0,1]$ we can associate another element of M, denoted by $[pa;(1-p)b]$ and called a *lottery* between a and b. (Henceforth the letters p and q will be reserved for elements of $[0,1]$.) We assume that lotteries have the following properties for all $a, b \in M$:

$$[1a;0b] = a,$$
$$[pa;(1-p)b] = [(1-p)b;pa],$$
$$[q[pa;(1-p)b];(1-q)b] = [pqa;(1-pq)b].$$

A *preference relation* on M is defined to be a binary relation $\succeq^*$ such that for any $a, b \in M$ either $a \succeq^* b$ or $b \succeq^* a$ must hold, and if $a \succeq^* b$ and $b \succeq^* c$ then $a \succeq^* c$. We write $a >^* b$ if $a \succeq^* b$ and $b \not\succeq^* a$, and $a \sim b$ if $a \succeq^* b$ and $b \succeq^* a$. (So $a >^* b$ means that the individual whose preferences we are considering prefers a to b; $a \succeq^* b$ means he likes a at least as well as b; and $a \sim b$ means he is indifferent between the two alternatives.) A real-valued function u defined on a mixture set M is an *expected utility*

function for the preference $\geq^*$ if it is order preserving (i.e., if for all a and b in M, $u(a) > u(b)$ if and only if $a >^* b$), and if it evaluates the utility of lotteries by their expected utility; that is, if for any lottery $[pa;(1-p)b]$,

$$u([pa;(1-p)b]) = pu(a) + (1-p)u(b).$$

If $\geq^*$ is a preference ordering on a mixture set M, then the following conditions ensure that an expected utility function exists:

Continuity: For any a, b, $c \in M$, the sets $\{p|[pa;(1-p)b] \geq^* c\}$ and $\{p|c \geq^* [pa;(1-p)b]\}$ are closed.

Substitutability: If a, $a' \in M$ and $a \sim a'$, then for any $b \in M$, $[\frac{1}{2}a;\frac{1}{2}b] \sim [\frac{1}{2}a';\frac{1}{2}b]$.

The utility function is unique up to an affine transformation; that is, if u is an expected utility function representing the preferences $\geq^*$, then so is v if and only if $v = c_1u + c_2$, where c_1 and c_2 are real numbers and $c_1 > 0$. Another way to say this is that in specifying a utility function u representing the preferences $\geq^*$, we are free to choose arbitrarily any alternatives a_1 and a_0 in M, such that $a_1 >^* a_0$, and set $u(a_1) = 1$ and $u(a_0) = 0$. When these arbitrary elements are specified, the value of $u(a)$ for any other alternative a is then completely determined by the preferences.[3] For example, if the alternative a is such that $a_1 \geq^* a \geq^* a_0$, then $u(a) = p$, where p is the probability such that $a \sim [pa_1;(1-p)a_0]$. (This follows since the utility of the lottery is p, its expected utility.)

3 Comparing positions in games

In what follows, we will consider for simplicity the class G of superadditive characteristic function games[4] v defined on a universe of positions N, where N is taken to be finite. To make comparison between positions in a game and in different games, we shall consider a preference relation defined on the set $N \times G$ of positions in a game. So $(i,v) >^* (j,w)$ means "it is preferable to play position i in game v than to play position j in game w." As before, $\sim$ will denote indifference, and $\geq^*$ will denote weak preference.

We consider preference relations that are also defined on the mixture set M generated by $N \times G$ (i.e., the smallest mixture set containing $N \times G$). That is, preferences are also defined over lotteries whose outcomes are positions in a game. Denote by $[q(i,v);(1-q)(j,w)]$ the lottery that, with probability q, has a player take position i in game v and, with

probability $1 - q$, take position j in game w. We henceforth consider only preference relations that have the standard properties of continuity and substitutability on M and that ensure the existence of an expected utility function unique up to the choice of origin and unit. Denote this function by θ, and write $\theta_i(v) \equiv \theta((i,v))$ and $\theta(v) \equiv (\theta_1(v), \ldots, \theta_n(v))$. Because θ is an expected utility function, $\theta_i(v) > \theta_j(w)$ if and only if the individual whose preferences are being modeled prefers to play position i in game v rather than position j in game w, and the utility of a lottery is its expected utility: that is,

$$\theta([p(i,v);(1-p)(j,w)]) = p\theta_i(v) + (1-p)\theta_j(w).$$

Recall that the games v we are considering are themselves defined in terms of some transferable commodity that reflects the expected utility of the players for some underlying outcomes (e.g., as in note 2). Some additional regularity conditions on preferences for positions in games will be needed in order that the preferences, and the resulting utility function for positions in games, be consistent with the underlying utility function in terms of which the games are defined.

It will be convenient to define, for each position i, the game v_i by

$$\begin{aligned} v_i(S) &= 1 \quad \text{if } i \in S, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

All positions other than i are null players in games of the form cv_i, so the player in position i may be sure of getting a utility of c. (This observation will provide the appropriate normalization for the utility θ.) Denote by v_0 the game in which all players are null players (i.e., the game $v_0(S) = 0$ for all S), and let G_{-i} be the class of games in which position i is null.

The first regularity condition we impose on the preferences is

R1. If $v \in G_{-i}$, then $(i,v) \sim (i,v_0)$. Also, $(i,v_i) >^* (i,v_0)$.

This condition says that being a null player in a game is not preferable to being a null player in any other game (in particular in the game v_0), and that the position (i,v_i) is preferable to playing a null position.

The second regularity condition is

R2. For all $i \in N$, $v \in G$, and for any permutation π, $(i,v) \sim (\pi i, \pi v)$.

This condition says simply that the names of the positions do not affect their desirability. An immediate consequence is that the utility function for games will obey the symmetry axiom.

Lemma 1. $\theta_{\pi i}(\pi v) = \theta_i(v)$.

By R1 we can choose (i,v_i) and (i,v_0) to be the unit and origin of the utility scale, so $\theta_i(v_i) = 1$ and $\theta_i(v_0) = 0$. These are the natural normalizations, reflecting the fact that a player in position i of game v_0 is assured of receiving a payoff of 0 (in terms of her underlying utility function for the outcomes of the games), and a player in position i of v_i is assured of receiving 1.

The last regularity condition reflects that the games v are defined in terms of an *expected* utility function.

R3. For any number $c > 1$ and for every (i,v) in $N \times G$,

$$(i,v) \sim [(1/c)(i,cv);(1 - 1/c)(i,v_0)].$$

Condition R3 reflects the fact that games v and cv are identical except for the scale of the rewards. These rewards are expressed in terms of a player's expected utility for the underlying consequences, so a player is indifferent between receiving a utility of 1 or of having the lottery that gives him or her a utility of c with probability $1/c$, and 0 with probability $1 - 1/c$. Condition R3 says that, whatever a player's expectation from playing position i in game v, it is related by the same sort of lottery to his or her expectation for playing position i in game cv.

Lemma 2. For any $c \geq 0$ and any $(i,v) \in N \times G$, $\theta_i(cv) = c\theta_i(v)$.

Proof: Without loss of generality we can take $c \geq 1$ (because if $c = 0$, the result follows from condition R1 and the normalization that $\theta_i(v_0) = 0$, and if $0 < c < 1$ we can simply consider $c' = 1/c$). By R3

$$(i,v) \sim [(1/c)(i,cv); (1 - 1/c)(i,v_0)],$$

so

$$\begin{aligned}\theta_i(v) &= \theta([(1/c)(i,cv); (1 - 1/c)(i,v_0)]) \\ &= (1/c)\theta_i(cv) + (1 - 1/c)\theta_i(v_0) \\ &= (1/c)\theta_i(cv).\end{aligned}$$

These regularity conditions, together with the normalization that $\theta_i(v_i) = 1$ and $\theta_i(v_0) = 0$, place some constraints on the utility function θ that allow us to interpret it as an extension of the underlying utility function defining the games. (We can regard the alternative (i,cv_i) as "embedding" in the mixture space M of positions in games the underlying payoffs of the games themselves, because the opportunity to play position

i in the game cv_i is essentially the same as being given a prize with utility c, and $\theta_i(cv_i) = c$.) We will call a utility function on M normalized in this way and satisfying R1–R3 an *extended* utility function, because it extends to the space of positions in games the utility function used to define the games. However, infinitely many extended utility functions still could arise, because the preferences that an agent could have over games still have many degrees of freedom. In particular, we turn now to consider an individual's attitude toward different kinds of risk.

4 Risk posture

We distinguish between two kinds of risk. *Ordinary risk* involves the uncertainty that arises from lotteries, whereas *strategic risk* involves the uncertainty that arises from the strategic interaction of the players in a game.

4.1 *Ordinary risk*

Recall that when we consider preferences defined over money, we say that an individual is "risk neutral" if his utility for any lottery is equal to its expected monetary value. Analogously, we say that an individual is "risk neutral to ordinary risk over games" if her preferences obey the following condition.

Neutrality to ordinary risk over games:

$$(i, (qw + (1 - q)v)) \sim [q(i,w);(1 - q)(i,v)].$$

The condition says that the individual is indifferent between the alternative on the right, which is a lottery that will result in playing position i in either game w or game v, and the alternative on the left, which is to play position i in the game whose characteristic function is equal to the expected value of the characteristic function of the lottery. That is, consider some coalition $S \subset N$. Its expected worth in the lottery on the right is $qw(S) + (1 - q)v(S)$, which is precisely its worth in the game on the left. So a player is risk neutral with respect to ordinary risk over games if he or she is indifferent between playing position i in the "expected game" $qw + (1 - q)v$ or to having the appropriate lottery between the games w and v.

Note that $v = (1/c)cv + (1 - (1/c)v_0)$, so neutrality to ordinary risk over games implies regularity condition R3. In fact, it is a much stronger

condition, and in Section 6 we briefly consider why an individual might not be neutral to ordinary risk over games, even if he or she was risk neutral in terms of the transferable commodity used to define them. However, the next result shows that this kind of risk neutrality is just what is involved in assuming that the utility function θ is additive.

Theorem 1 (Additivity). $\theta(v + w) = \theta(v) + \theta(w)$ for all $v, w \in G$ if and only if preferences are neutral to ordinary risk over games.

Proof: For each $i \in N$,

$$\theta_i(v + w) = \theta_i(2(\tfrac{1}{2}v + \tfrac{1}{2}w)) = 2\theta_i(\tfrac{1}{2}v + \tfrac{1}{2}w)$$

by Lemma 2. But by ordinary risk neutrality over games,

$$\theta_i(\tfrac{1}{2}v + \tfrac{1}{2}w) = \theta_i([\tfrac{1}{2}(i,v);\tfrac{1}{2}(i,w)]) = \tfrac{1}{2}\theta_i(v) + \tfrac{1}{2}\theta_i(w),$$

because θ is an expected utility function. So $\theta_i(v + w) = \theta_i(v) + \theta_i(w)$. The other direction is equally straightforward, after the initial task of proving that additivity of an extended utility function (together with continuity) implies the conclusions of Lemma 2.

There is uncertainty in playing a game even if no lotteries are involved. In Roth (1977a,b) this was called *strategic* risk. Given that an individual is neutral to ordinary risk over games, we will now show that the individual's posture toward strategic risk uniquely determines his or her utility for a position in a game.

4.2 *Strategic risk*

Any game with more than one strategic (i.e., nondummy) position involves some potential uncertainty as to the outcome, arising from the interaction of the strategic players. To describe a given player's preferences for situations involving strategic risk, it will be convenient for us to consider it on the games v_R defined for each subset R of N by

$$\begin{aligned} v_R(S) &= 1 \quad \text{if } R \subset S, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

A "pure bargaining game" of the form v_R is essentially the simplest game that can be played among r strategic players. (The cardinality of sets R, S, T, . . . is denoted by r, s, t, ...)

Define the *certain equivalent* of a strategic position in a game v_R to be the number $f(r)$ such that the prospect of receiving $f(r)$ for certain is exactly as desirable as the prospect of playing the strategic position.[5] That is, $f(r)$ is the number such that, for $i \in R$, $(i,v_R) \sim (i,f(r)v_i)$. Note that $f(1) = 1$, and that $f(r)$ is a measure of a player's opinion of his or her own bargaining ability in pure bargaining games of size r.

Using the terminology of Roth (1977a), we say that the preference is *neutral to strategic risk* if $f(r) = 1/r$ for $r = 1, \ldots, n$. The preference is *strategic risk averse* if $f(r) \leq 1/r$, and *strategic risk preferring* if $f(r) \geq 1/r$. (Note that preferences may be none of these; e.g., if $f(2) > 1/2$ but $f(3) < 1/3$.) The utility of playing a position in a game v_R is given by the following lemma.

Lemma 3.

$$\begin{aligned} \theta_i(v_R) &= f(r) && \text{if } i \in R, \\ &= 0 && \text{otherwise.} \end{aligned}$$

Proof: If $i \notin R$, then $v_R \in G_{-i}$ and $\theta_i(v_R) = \theta_i(v_0) = 0$, by R1. If $i \in R$, then $\theta_i(v_R) = \theta_i(f(r)v_i) = f(r)\theta_i(v_i) = f(r)$, by Lemma 2.

If preferences are neutral to ordinary risk over games, then Theorem 1 implies that the utility function is completely determined by the numbers $f(r)$, because the games v_R are an additive basis. We have the following result.

Shapley value theorem. The Shapley value is the utility function of an individual who is both neutral to ordinary risk over games and neutral to strategic risk. That is, when preferences are neutral to both kinds of risk,

$$\theta_i(v) = \phi_i(v) = \sum_{S \subset N} \frac{(s-1)!(n-s)!}{n!} [v(S) - v(S-i)].$$

Proof: Neutrality to ordinary risk over games implies that θ is additive, and strategic risk neutrality implies that θ agrees with the Shapley value ϕ on all games of the form v_R. Because the games v_R constitute a basis of the space of games, it follows that θ agrees with ϕ on all games.

As the Shapley value theorem and its proof make clear, neutrality to ordinary risk together with different strategic risk preferences (as ex-

pressed by the numbers $f(r)$, $r = 2, \ldots, n$) will determine utility functions that differ from the Shapley value. These other utility functions for games are given by the next result.

Representation theorem. When preferences are neutral to ordinary risk over games, the utility function θ has the form

$$\theta_i(v) = \sum_{T \subset N} k(t)[v(T) - v(T - i)], \tag{1}$$

where

$$k(t) = \sum_{r=t}^{n} (-1)^{r-t} \binom{n-t}{r-t} f(r).$$

Proof: Every game v is a sum of games of the form v_R. In fact (see Shapley 1953 and Chapter 2 of this volume), $v = \Sigma_{R \subset N} c_R v_R$, where $c_R = \Sigma_{T \subset R}(-1)^{r-t} v(T)$. By Lemma 2 and Theorem 1,

$$\theta_i(v) = \sum_{R \subset N} c_R \theta_i(v_R) = \sum_{\substack{R \subset N \\ i \in R}} c_R f(r) = \sum_{\substack{R \subset N \\ i \in R}} \sum_{T \subset R} (-1)^{r-t} v(T) f(r).$$

Reversing the order of summation, we obtain

$$\theta_i(v) = \sum_{T \subset N} \left\{ \sum_{\substack{R \subset N \\ R \supset (T \cup i)}} (-1)^{r-t} f(r) \right\} v(T).$$

If we denote the term in braces by $g_i(T)$, then we note that $g_i(T) = -g_i(T - i)$ when $i \in T$. So

$$\theta_i(v) = \sum_{\substack{T \subset N \\ i \in T}} g_i(T)[v(T) - v(T - i)].$$

But there are $\binom{n-t}{r-t}$ coalitions of size r that contain T, so

$$g_i(T) = \sum_{r=t}^{n} (-1)^{r-t} \binom{n-t}{r-t} f(r) = k(t).$$

Because $[v(T) - v(T - i)] = 0$ unless $i \in T$, we are done.

An immediate consequence of this representation theorem is that when preferences are neutral to ordinary risk (i.e., when the utility function is additive), then the utility of playing a null position is 0, because the utility is the weighted sum of marginal contributions. The effect of strate-

gic risk neutrality, and the special feature of the Shapley value, is that the sum over positions equals $v(N)$. In a purely axiomatic framework (Roth 1977d), the theorem can be restated to say that any symmetric and additive value θ that gives null players 0 (or, equivalently, with the property that $\Sigma_{i\in T}\theta_i(v) = \Sigma_{i\in S}\theta_i(v)$ for any carriers T, S of a game v) is a weighted sum of marginal contributions as given in the theorem. Such values, which need not sum to $v(N)$, have subsequently been called *semivalues* (Dubey, Neyman, and Weber 1981; Einy 1987; Weber Chapter 7 this volume).

The semivalue that has received perhaps the most attention in the literature (cf. Banzhaf 1965; Coleman 1971; Owen 1975; Dubey 1975a; Roth 1977b,c; Dubey and Shapley 1979; Straffin Chapter 5 this volume) is the Banzhaf index $\beta' = (\beta'_1, \ldots, \beta'_n)$ given by

$$\beta'_i(v) = \sum_{S\subset N} \frac{1}{2^{n-1}} [v(S) - v(S - i)].$$

Banzhaf (1965) originally proposed a version of this index in the context of simple games (see Chapters 1 and 5), but the extension to general games is straightforward, the major difference being that the marginal contributions $v(S) - v(S - i)$ may take on values other than 0 and 1. The factor $1/2^{n-1}$ is a convenient normalization, but others could be chosen. The important point for the following result is that the normalization does not depend on the game v. (In some treatments of the Banzhaf index for simple games, the index is normalized so that $\beta_i = \beta'_i/\Sigma_{i\in N}\beta'_i$ for β' as defined here, so $\Sigma\beta_i = 1$. But this involves a different divisor for each game, so the resulting index is not additive–i.e., not neutral to ordinary risk.)

The Banzhaf index β' as normalized here is an extended utility function reflecting preferences averse to strategic risk and neutral to ordinary risk. We state without proof the following corollary of the representation theorem, from Roth (1977d).

Corollary. If $f(r) = 1/2^{r-1}$, then the extended utility function equals the Banzhaf index; that is, $\theta(v) = \beta'(v)$.

Thus the Banzhaf index is a utility function in which a player's utility for a strategic position in a game v_R is inversely proportional to the number of ways the $r - 1$ other strategic players can form coalitions.

5 Simple games

As discussed in Chapter 1, the Banzhaf index, like the Shapley–Shubik (1954) index, was proposed in connection with voting processes modeled by simple games. However, the characterization of the Shapley value and Banzhaf index for general games given here, like Shapley's axiomatic characterization of the Shapley value, makes crucial use of nonsimple games. If the universe of games we are interested in consists only of simple games, then symmetry, efficiency, and additivity do not uniquely characterize the Shapley value. In particular, the Banzhaf index β, normalized so as to sum to 1, also obeys these three axioms when they are applied only to simple games. The reason is that additivity (equivalently, neutrality to ordinary risk) loses all its force when applied only to simple games, because the class of simple games is not closed under addition. So if v and w are nontrivial simple games, $v(N) = w(N) = 1$ and the game $v + w$ is not simple, because $v(N) + w(N) = 2$. In this section we follow Roth (1977c) in considering how the Shapley–Shubik index can be (uniquely) characterized as a risk-neutral utility function defined on the class of simple games.

Dubey (1975a,b) axiomatically characterized the Shapley–Shubik index on the class of simple games by replacing additivity with the following axiom (which Weber, in Chapter 5, has called the *transfer axiom*).

Transfer axiom. For any simple games v, w,

$$\phi(v \vee w) + \phi(v \wedge w) = \phi(v) + \phi(w),$$

where the games $v \vee w$ and $v \wedge w$ are defined by

$$\begin{aligned} (v \vee w)(S) &= 1 \quad \text{if } v(S) = 1 \text{ or } w(S) = 1, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

and

$$\begin{aligned} (v \wedge w)(S) &= 1 \quad \text{if } v(S) = 1 \text{ and } w(S) = 1, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Perhaps the easiest way to understand the transfer axiom is to recast it in terms of preferences over games and lotteries over games, as a form of neutrality to ordinary risk over simple games. Viewed in that way, it takes the following form.

Ordinary risk neutrality for simple games: For all simple games v, w

$$[\tfrac{1}{2}(v,i);\tfrac{1}{2}(w,i)] \sim [\tfrac{1}{2}((v \vee w),i);\tfrac{1}{2}((v \wedge w),i)].$$

This condition specifies indifference between two lotteries. One lottery results in either the game v or the game w, and the other results in either the game $v \vee w$ or the game $v \wedge w$. What makes this a condition of risk neutrality is that any given coalition S has the same probability of being a winning coalition in either lottery. It follows immediately from the fact that θ is an expected utility function that if it is neutral to ordinary risk it obeys the transfer axiom.

In order to state all conditions on preferences in terms of simple games only, we also need to rewrite neutrality to strategic risk, because the game $f(r)v_i$ is not a simple game. The following condition involves only simple games.

Strategic risk neutrality for simple games: For all $R \subset N$ and $i \in R$,

$$(v_R,i) \sim [\tfrac{1}{r}(v_i,i);(1 - \tfrac{1}{r})(v_0,i)].$$

It is easy to see that, when the utility function is normalized as in the previous sections so that $\theta_i(v_i) = 1$ and $\theta_i(v_0) = 0$, strategic risk neutrality for simple games continues to imply that θ coincides with ϕ on the class of pure bargaining games.

Dubey proved the following result.

Proposition. The Shapley–Shubik index is the unique function ϕ defined on simple games that obeys Shapley's symmetry and carrier axioms as well as the transfer axiom.

In terms of utilities, we can now recast this result as follows.

Shapley–Shubik index theorem. The Shapley–Shubik index is the unique utility θ, normalized so that $\theta_i(v_i) = 1$ and $\theta_i(v_0) = 0$, corresponding to preferences that obey conditions R1 and R2 and that are neutral to both strategic and ordinary risk defined over simple games.

Proof: We have already observed that θ is symmetric (Lemma 1) and obeys the transfer axiom, and that for every $R \subset N$, $\theta(v_R) = \phi(v_R)$; that is, θ coincides with the Shapley–Shubik index on the pure bargaining games

v_R. To complete the proof of the theorem, we show that θ coincides with ϕ on every simple game v.

Let $R_1, R_2, \ldots, R_k \subset N$ be all the distinct minimal winning coalitions[6] of v. Then we say the game v is in class k, and note that $v = v_{R1} \vee v_{R2} \vee \cdots \vee v_{Rk}$. If v is in class $k = 0$, then $v = v_0$ and $\theta(v) = \phi(v) = 0$. If v is in class $k = 1$, then $v = v_{R1}$ is a pure bargaining game, and $\theta(v) = \phi(v)$.

Suppose that for games v in classes $k = 1, 2, \ldots, m$ it has been shown that θ is well defined and coincides with the Shapley–Shubik index. Consider a game v in class $m + 1$. Then

$$v = v_{R1} \vee v_{R2} \vee \cdots \vee v_{Rm} \vee v_R = w \vee v_R,$$

where w is a game in class m. Hence, by neutrality to ordinary risk over simple games (which implies that the utility θ obeys the transfer axiom),

$$\theta_i(v) = \theta_i(w \vee v_R) = \theta_i(w) + \theta_i(v_R) - \theta_i(w \wedge v_R).$$

But we show that the game $w \wedge v_R$ cannot be in a class higher than w, so by the inductive hypothesis the terms on the right side of the preceding expression are uniquely determined and equal to the Shapley–Shubik index. Consequently, we will have shown that $\theta(v) = \phi(v)$ for all simple games v.

To see that the game $w' = (w \wedge v_R)$ cannot be in a class higher than the game w, consider a minimal winning coalition S' of the game w'. By the definition of w' we know that $S' \supset R$ and $w(S') = 1$. If $S' = R$, then $w' = v_R$ and we are done (because except for the game v_0, every game has at least one minimal winning coalition). Otherwise, $S' = S \cup R$, where S is a minimal winning coalition in the game w. (Of course, S and R need not be disjoint.)

Consider now a coalition T' that is minimal winning in w'. Then $T' = T \cup R$, where T is minimal winning in w. If $T' \neq S'$, then $T \neq S$. Consequently, every minimal winning coalition in w' can be identified with a distinct minimal winning coalition in w, so w' cannot be in a class higher than w. This completes the proof.

6 Discussion

To see what has been accomplished by considering the Shapley value as a utility function, let us consider what kind of answers have been obtained to the questions raised in the introduction to this chapter.

1. The "uniqueness" of the Shapley value as a utility function for

games is associated with its risk neutrality. Perhaps a good way to think of this is by analogy with utility functions for money: The risk-neutral utility function is the one that evaluates lotteries at their expected value. Although few of us consider only the expected value when choosing among risky investments, for example, the expected value is nevertheless enormously important to know and can give us at least a rough indication of what our preferences are likely to be upon closer investigation. In the same way, the Shapley value gives an indication of what our preferences over positions in games are likely to be, even if we are not neutral to both strategic and ordinary risk over games. And for preferences that are neutral to ordinary risk over games, we have been able to characterize the utility functions that reflect different attitudes toward strategic risk. However, the systematic behavior of utility functions that reflect different attitudes toward ordinary risk over games remains an open question.

2. We have seen that the Shapley value "inherits" the normalization of the utility function used to define the games being considered. That is, underlying any game is a concrete set of outcomes that are represented by utility payoffs in terms of utility functions with arbitrary origin and unit. An individual's Shapley value is an extension of this utility function, with the same normalization. Thus the meaningful utility comparisons that can be made with the Shapley value are precisely those that can be made with expected utility functions. For example, in Chapter 1 the Shapley value was calculated for a simple model of the U.N. Security Council, yielding a Shapley value of .00186 for a rotating member and a Shapley value of .196 for a permanent member. Viewing the Shapley value as an expected utility function, we can now determine which statements about these numbers are meaningful comparisons reflecting the underlying preferences, and which are not. For example, an individual who is neutral to both ordinary and strategic risk would be indifferent between playing the game in the position of a permanent member or to having a lottery that gave a .196 probability of being a dictator in the game, and otherwise made him a null player (or for that matter having a lottery that gave a .196 probability of receiving any prospect with a utility of 1, and otherwise receiving a utility of 0). Similarly, such an individual would be indifferent between playing the position of a rotating member or having a lottery that gave her a probability of $p = .00186/.196 = .0095$ of playing the position of a permanent member and otherwise being a null player. To put it another way, this individual would prefer a 1 in 100 chance of being a permanent member (and a 99 in 100 chance of being a null player) to the prospect of being a rotating member. But it would not be a meaningful

comparison to say that the prospect of playing a position in a game is over 100 times as desirable as another prospect (which could be either a position in a game or a lottery over prizes), because this depends on the (arbitrary) normalization chosen for the underlying utility function.

3. We have seen that the additivity axiom on the value function is equivalent to assuming that the preferences that the value represents as a utility function are neutral with respect to ordinary risk over games. Perhaps the best way to understand what this entails at this point is to consider why an individual might *not* be neutral to this kind of risk. For example, let v be the three-person majority game given by $v(1) = v(2) = v(3) = 0$ and $v(12) = v(13) = v(23) = v(123) = 1$, and let $w = v_{\{12\}}$ be the two-person pure bargaining game (with player 3 a null player). Then the game $z = v + w$ is given by $z(1) = z(2) = z(3) = 0$, $z(13) = z(23) = 1$, $z(12) = z(123) = 2$. Although v and w are both symmetric among the nonnull players, z is not. In particular, the symmetry of v makes it not unreasonable to suppose that each of the two-person coalitions is as likely to form as any other, and that if the three-person coalition forms it will divide equally. So the fact that $\phi(v) = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$ seems reasonable, as does the fact that $\phi(w) = (\frac{1}{2},\frac{1}{2},0)$. So our evaluation of the two games separately is that each two-person coalition is equally likely to form in v, but in w the coalition $\{12\}$ will form.

Therefore the coalition $\{12\}$ should be especially easy to form in the game z because players 1 and 2 are essential for the game to be worth 2, and player 3 can make no further marginal contribution. (This is clearly reflected in the core of z, which is the single payoff vector (1,1,0).) But the Shapley value is $\phi(z) = (\frac{5}{6},\frac{5}{6},\frac{1}{3})$. That is, an individual whose utility is the Shapley value is indifferent between playing position 3 in the game z or in the game v. Although this preference may be consistent with plausible models of how the game might be played (because game z gives 3 a less advantageous position but has higher stakes than game v), I think that for most purposes I would personally prefer to play position 3 of game v rather than of game z. So, although neutrality to ordinary risk is an easy to understand and plausible condition on preferences that gives rise to tractable (i.e., additive) utility functions, it is by no means an inescapable requirement for plausible preferences, either for all individuals or for a given individual over all games.

4. Finally, we have seen that, when preferences are neutral to ordinary risk so that the utility function is additive, the vector θ of utility for each position in a game is "efficient" if and only if the preferences are neutral to strategic risk. The quotation marks reflect the fact that under the interpretation presented here the vector θ is not a distribution of utility among

different players but simply a vector of utilities for the different positions. Indeed, whether the vector $(\theta_1(v), \ldots, \theta_n(v))$ is even a *feasible* outcome of the game, let alone an efficient one, appears to arise in this context essentially by accident.[7] The risk-neutral utility – the Shapley value – always happens to coincide with an outcome of the game, but utility vectors that do not reflect neutrality to strategic risk do not share this property.[8] In any event no interpersonal comparisons are implied, because all comparisons are those of a single agent evaluating alternative positions.

As to whether we should expect individuals to be neutral to strategic risk, just as many individuals do not judge monetary lotteries only by their expected value, I imagine that many are not indifferent between bargaining among r individuals or receiving $1/r$ of the proceeds for sure. Certainly some aversion to strategic risk would appear to be justified by the experimental evidence, which reveals a nonnegligible frequency of disagreement (see Roth 1987), and by the growing theoretical understanding about how differences in information, ability to make commitments, or long-term concerns may lead to disagreements (see, e.g., the papers in Roth 1985 or Binmore and Dasgupta 1987). So, like additivity, efficiency arises from assumptions about preferences that are plausible but by no means inescapable.

In conclusion, the analogy between the Shapley value, which is the risk-neutral utility for playing a game, and the expected value, which is the risk-neutral utility for monetary gambles, seems to be a strong one. (Note that this is *not* because of the interpretation of the Shapley value as an expected marginal contribution. The Banzhaf index and other non-risk-neutral utilities can also be interpreted as expected marginal contributions; see, e.g., Weber Chapter 7 this volume.) When we consider a specific individual or a specific choice among games, we may be able to find a more precise indicator. But when we are considering a first approximation, both the expected value of monetary gambles and the Shapley value of transferable utility games seem to work in similar ways. And even if we conclude that most individuals are not risk neutral, the assumptions of risk neutrality implicit in the Shapley value, like the expected value, may be a more natural proxy for the utility of some unspecified individual than would any assumption of a particular risk posture.

NOTES

1 In the same way, the arithmetic mean of players' utilities is not meaningful. But the geometric mean of expected utilities is: This forms the basis for Nash's celebrated model of bargaining (see Nash 1950; Roth 1979).

2 Of course, a similar question arises concerning transferable utility games, in which the "transferable utility" payoffs to the players are assumed to sum to (at most) $v(N)$. No assumption that utilities are interpersonally comparable needs to be made to consider such a game. For example, if the payoffs are all in money and the players are all risk neutral, then the characteristic function form representation of the game simply involves a common (but still arbitrary) normalization of the players' utility functions. To see that no fundamental comparisons are involved, observe that we could construct a characteristic function form game among players, all of whom receive quite different commodities and among whom no actual physical transfers can take place. Consider three players, one of whom will ultimately be paid in French francs, one in baskets of fruit, and one in wine. For each player, a utility function is constructed for possible payoffs. The arbitrary elements in each utility representation are chosen without reference to the others. A given characteristic function game v defined on $N = \{1,2,3\}$ can now be created by allowing the members of each coalition $S \subset N$ (who can communicate by telephone and sign contracts as needed) to divide an amount $v(S)$ of a fictitious commodity–"utility money"–in any way they choose. At the conclusion of the game, each player may exchange whatever utility money he has earned for the amount of the commodity in which he is to be paid that gives him that amount of utility, according to the arbitrarily scaled utility function established for him before the game.

3 In general, for any element $x \in M$, the utility of x is

$$u(x) = (p_{ab}(x) - p_{ab}(a_0))/(p_{ab}(a_1) - p_{ab}(a_0))$$

where a, b, a_1, and a_0 are elements of M such that $a \geq^* x \geq^* b$ and $a \geq^* a_1 \geq^* a_0 \geq^* b$, and for any $y \in M$ such that $a \geq^* y \geq^* b$, $p_{ab}(y)$ is defined by $y \sim [p_{ab}(y)a;(1 - p_{ab}(y))b]$.

It can be shown that the numbers $p_{ab}(\cdot)$ are well defined, and the function $u(\cdot)$ is independent of the choice of a and b. Note that $u(a_1) = 1$ and $u(a_0) = 0$.

4 The class of superadditive games is sufficiently large, but we could consider a larger class of games without changing the results presented here.

5 We take the point of view that a player does not know who will occupy the other positions in a game. Consequently, her certain equivalent for a game v_R depends only on r.

6 A coalition $R \subset N$ is minimal winning in v if $v(R) = 1$ and if $S \subset R$, $S \neq R$, implies $v(S) = 0$.

7 Note that, by analogy, expected values of money gambles aren't necessarily feasible outcomes: for example, the 50-50 gamble for plus or minus one dollar has an expected value of 0, although that isn't a feasible outcome. For transferable utility games feasibility comes along with risk neutrality, but this does not appear to be the case for NTU games. It seems to me that this may be part of the trouble in interpreting the value for NTU games along the lines of the Shapley value for TU games (see the references in this connection in Chapter 1).

8 This is so for utilities that are strategic risk averse as well as strategic risk

preferring. It is clear that a vector of utilities that are strategically risk preferring will not always coincide with a feasible outcome: Because $f(r) > 1/r$, such a utility vector isn't a feasible outcome in a pure bargaining game v_R, because $rf(r) > v_R(N) = 1$. For a risk-averse utility, consider the Banzhaf index β', with $f(r) = 1/2^{r-1}$. For the three-person majority game, $\beta'(v) = (\frac{1}{2},\frac{1}{2},\frac{1}{2})$, so $\Sigma\beta'_i > v(N) = 1$. To see what is going on here, note that $v = v_{(12)} + v_{(13)} + v_{(23)} - 2v_{(123)}$, and so $\beta'_i(v) = \beta'_i(v_{(12)}) + \beta'_i(v_{(13)}) + \beta'_i(v_{(23)}) - 2\beta'_i(v_{(123)})$. But when $r = 2$, $1/2^{-1} = \frac{1}{2}$, so the Banzhaf index agrees with the Shapley value on the two-person pure bargaining games. But when $r = 3$, the strategic risk aversion of the Banzhaf agent comes into play, with $\beta'_i(v_{(123)}) = \frac{1}{4}$ for each $i = 1, 2, 3$ (in contrast to the Shapley value $\phi'_i(v_{(123)}) = \frac{1}{3}$). Because $v_{(123)}$ enters the expression for the three-person majority game v with a negative coefficient, this means that the relatively greater strategic risk aversion of the Banzhaf agent, which causes him to evaluate the three-person pure bargaining game less favorably than does the Shapley agent, nevertheless causes him to evaluate the three-person majority game more favorably. Thus, in the presence of neutrality to ordinary risk (i.e., additivity) differences in strategic risk aversion can have effects that are difficult to anticipate.

There are (at least) two ways to think about these effects of strategic risk aversion. On the one hand, they appear to parallel similar effects of ordinary risk aversion found in game-theoretic models of bargaining (Roth and Rothblum 1982; Harrington 1987). On the other hand, they are also intimately related to the assumption of neutrality to ordinary risk and the resulting additivity of the utility function, and so these effects may also provide some further cause to be cautious about the assumption of additivity.

REFERENCES

Banzhaf, John F. III [1965], "Weighted Voting Doesn't Work: A Mathematical Analysis," *Rutgers Law Review,* 19, 317–43.

Binmore, Ken and Partha Dasgupta (editors) [1987], *The Economics of Bargaining,* Oxford, Basil Blackwell.

Coleman, James S. [1971], "Control of Collectivities and the Power of a Collectivity to Act," in *Social Choice,* B. Lieberman, editor, Gordon and Breach, New York, pp. 269–300.

Dubey, Pradeep [1975a], "Some Results on Values of Finite and Infinite Games," Ph.D. thesis, Cornell University.

[1975b], "On the Uniqueness of the Shapley Value," *International Journal of Game Theory,* 4, 131–9.

Dubey, Pradeep, Abraham Neyman, and Robert Weber [1981], "Value Theory without Efficiency," *Mathematics of Operations Research,* 6, 122–8.

Dubey, Pradeep and Lloyd S. Shapley [1979], "Mathematical Properties of the Banzhaf Power Index," *Mathematics of Operations Research,* 4, 99–131.

Einy, Ezra [1987], "Semivalues of Simple Games," *Mathematics of Operations Research,* 12, 185–92.

Harrington, Joseph E., Jr. [1987], "The Role of Risk Preferences in Bargaining for

the Class of Symmetric Voting Rules," Department of Political Economy, Johns Hopkins, mimeo.

Herstein, I. N. and J. Milnor [1953], "An Axiomatic Approach to Measurable Utility," *Econometrica,* 21, 291–7.

Nash, John F. [1950], "The Bargaining Problem," *Econometrica,* 54, 155–62.

Owen, Guillermo [1975], "Multilinear Extensions and the Banzhaf Value," *Naval Research Logistics Quarterly,* 22, 741–750.

Roth, Alvin E. [1977a], "The Shapley Value as a von Neumann–Morgenstern Utility," *Econometrica,* 45, 657–64.

[1977b], "Bargaining Ability, the Utility of Playing a Game, and Models of Coalition Formation," *Journal of Mathematical Psychology,* 16, 153–60.

[1977c], "Utility Functions for Simple Games," *Journal of Economic Theory,* 16, 481–9.

[1977d], "A Note on Values and Multilinear Extensions," *Naval Research Logistics Quarterly,* 24, 517–20.

[1979], *Axiomatic Models of Bargaining,* Lecture Notes in Economics and Mathematical Systems, No. 170, Springer-Verlag, New York.

editor, [1985], *Game-Theoretic Models of Bargaining,* Cambridge University Press, Cambridge.

[1987], "Bargaining Phenomena and Bargaining Theory," *Laboratory Experimentation in Economics: Six Points of View,* A. E. Roth, editor, Cambridge University Press, Cambridge, pp. 14–41.

Roth, Alvin E. and Uriel Rothblum [1982], "Risk Aversion and Nash's Solution for Bargaining Games With Risky Outcomes, *Econometrica,* 50, 639–47.

Shapley, Lloyd S. [1953], "A Value for *n*-Person Games," *Contributions to the Theory of Games,* vol. II, H. W. Kuhn and A. W. Tucker, editors, Ann. of Math. Studies 28, Princeton University Press, Princeton, New Jersey, pp. 307–17.

Shapley, Lloyd S. and Martin Shubik [1954], "A Method for Evaluating the Distribution of Power in a Committee System," *American Political Science Review,* 48, 787–92.