האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

ASSESSING STRATEGIC RISK

by

R. J. AUMANN and J. H. DREZE

Discussion Paper # 361 May 2004

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Assessing Strategic Risk*

R. J. Aumann[†] and J. H. Dreze[‡]

May 18, 2004

Abstract

In recent decades, the concept of subjective probability has been increasingly applied to an adversary's choices in strategic games. A careful examination reveals that the standard construction of subjective probabilities does not apply in this context. We show how the difficulty may be overcome by means of a different construction.

1. Introduction

Half a century ago, when decision theory and game theory were young, it was common to perceive a dichotomy between (i) *games against nature,* in which the "adversary" is a neutral "nature"—and (ii) *strategic games,* in which the adversary is an interested party. Games against nature were analyzed using several criteria, most prominent being the maximization of subjective expected utility—i.e., expected utility when the probabilities assigned to nature's moves are "subjective" or "personal," as in Savage [1954];[1] whereas strategic games were analyzed by minimax, or, more generally, strategic equilibrium in the sense of Nash [1951]. No need was seen to reconcile or even relate the approaches.

*The authors are grateful to Sergiu Hart, Martin Meier, Jean-Francois Mertens and Shmuel Zamir for important input. We alone remain responsible for the contents, including any errors or sloppy reasoning.

†Department of Mathematics and Center for Rationality, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel; raumann@math.huji.ac.il

‡CORE, Université Catholique de Louvain, 34, Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; dreze@core.ucl.ac.be

1 Related ideas occur in Ramsey [1931], DeFinetti [1937], Dreze [1961], and Anscombe and Aumann [1963].

In the ensuing years, the dichotomy gradually disappeared. It was recognized that games against nature and strategic games are in principle quite similar, and can—perhaps should—be treated similarly. Specifically, a player in a strategic game should be able to form subjective probabilities over the strategies of the other players, and then choose his own strategy so as to maximize his expected utility with respect to these subjective probabilities.

There is, however, a difficulty with applying the notion of subjective probability to strategic games—one that has, so far as we know, not been fully appreciated before. It has to do with the construction of the subjective probabilities. An event A is considered more likely than an event B if and only if the decision maker would rather "stake a prize" on A than on B, whatever the prize be. In a strategic game, the events of interest are the strategies of the other players. "Staking a prize" on strategy choices by adversaries is a modification of the payoff function, hence a modification of the game. This invalidates the whole process.

In this paper, we propose a resolution of this difficulty.

The nature of the difficulty that we resolve will be described more precisely in the following section. In Section 3, we describe the resolution informally. Section 4 is devoted to mathematical preliminaries, Section 5 to the formal statement of our main result, Section 6 to discussion, and Sections 7 and 8 to proving the main theorem.

2. The Difficulty

Before proceeding, we review and contrast the concepts of objective and subjective probability. *Objective* probabilities are associated with repeatable processes like coin tosses, roulette spins, dice throws, etc., in which considerations of symmetry determine the probability. *Subjective* probabilities are associated with one-time events like elections, the weather, a horse race or how an adversary will play a game. Reasonable people may be expected to agree on the numerical values of objective probabilities, but may well differ on subjective probabilities. That is why subjective probabilities are often called "personal:" They are associated with a particular person i. Roughly, one may think of i's subjective probability for an event E as that number p such that i would as soon have a dollar with objective probability p as a dollar contingent on E.

It will be useful to use the same terminology for games against nature and for strategic games. In either case, the "adversary"—be it nature or an interested party or parties—has several alternatives, called *strategies of the adversary.*

The decision maker—henceforth *protagonist*—also has several alternatives, called *strategies of the protagonist.* Together, the strategies of the adversary and of the protagonist determine the outcome of the game. Thus, each of the protagonist's strategies may be thought of as a function from the adversary's strategies to the possible outcomes: an "act" in the terminology of Savage [1954], a "horse lottery" in that of Anscombe and Aumann [1963].

The protagonist wishes to assign a subjective probability to each of the adversary's strategies. To do so, she[2] follows the procedure indicated at the end of the next-to-last paragraph. Specifically, for each of his (the adversary's) strategies s, she considers a hypothetical strategy r of her own that yields an outcome b if he plays any strategy other than s, and a different outcome a—preferred by her to b—if he plays[3] s. Clearly, she (weakly) prefers getting a for sure to playing r, and playing r to getting b for sure. So presumably, there is some number p between 0 and 1 such that she would as soon play r in the given game, as getting a with probability p, and b otherwise, an option henceforth denote $(a, b; p)$. This p is defined as her subjective probability for s.

We come now to the difficulty. The above is all right for games against nature, since *the choice of nature is not affected by the strategies that are available to the protagonist.* But in strategic games, the play of the adversary *is* strongly affected by the strategies available to the protagonist. In most cases, the hypothetical strategy r of the protagonist is indeed hypothetical—it is not actually available to her. Making it available changes the game, and so may well change the likelihood that the adversary plays s. But if it is *not* available, how can the protagonist consider it?

The reader might object that the situation is not all that different in games against nature, where, too, the strategy r may not be really available. But there, at least, one can consider a coherent hypothetical situation in which r *is* available. In strategic games, no such coherent hypothetical situation exists. On the one hand, r must be considered available, in order for the protagonist to evaluate it; on the other hand, it must *simultaneously* be considered unavailable, so that the adversary will play as in the original game. This internal contradiction is too much to swallow.

There is another difficulty, different from the above, namely that even if r really is available in the strategic game, the protagonist might not want to choose

[2]The protagonist is female. The other players are of indeterminate gender, so are referred to as "he," in accordance with standard usage before the feminist revolution.

[3]In other words, she considers "staking a prize on s," the prize being "a rather than b."

it, so that "comparison" with $(a, b; p)$ needs to be defined. This work does not resolve that difficulty, but it will be discussed in Section 6.5.

3. The Resolution

In this section, we state our result with a minimum of formality, while still striving for a maximum of clarity. Some compromises on both counts will be necessary. For a formal treatment, see Sections 4 and 5.

Following standard practice in decision theory, we rely on the primitive concept of "preference." But, we use preferences on *strategies,* which is *not* standard practice. Discussion and interpretation are postponed to Section 6.5.

Before the difficulty can be resolved, we must clarify what it is that we require from a "resolution." What, precisely, do we want subjective probabilities to do?

The first requirement is that the protagonist be able to use them to evaluate her options—in our terminology, her strategies. Specifically, starting out with the protagonist's preferences,[4] we want to construct utilities and subjective probabilities that "represent" those preferences: inter alia, such that she prefers one strategy to another if and only if its expected utility is greater.[5] Call this the "representation" requirement (or property).

There is also another, equally important desideratum. In Section 2 we wrote that one may think of i's subjective probability for an event E as that number p such that i would as soon have a dollar with objective probability p as a dollar contingent on E. Differently put, the constructed subjective probabilities should be interchangeable, preference-wise, with objective probabilities that have the same numerical value. Call this the "interchangeability" requirement (or property).

Before proceeding, we introduce some terminology and notation. A *lottery* is an objective probability distribution.[6] Unless otherwise indicated, a *strategy* is a strategy of the protagonist. A *consequence* is an outcome of the game. A *mixed strategy* is a lottery on strategies; a *mixed consequence* is a lottery on consequences.

To fulfill the above requirements, we posit a preference order on the set Δ of all lotteries λ whose prizes may be *either* a pure strategy of the protagonist, *or*

[4] Over her strategies and over the possible outcomes of the game.

[5] When the constructed subjective probabilities are assigned to the adversary's choices, and the constructed utilities are assigned to the outcomes of the game.

[6] Anscombe and Aumann call this a *roulette lottery;* they also have *horse lotteries,* which are similar to what we here call a "strategy" of the protagonist. See Section 6.3.

a consequence. This enables us to compare strategies with consequences, and so, to scale the utilities for strategies and for consequences to the same scale.

Operationally, such a lottery λ results either in (i) the outright selection of a specified pure outcome of the game, or in (ii) the game being played, with the protagonist choosing a specified pure strategy. More specifically, alternative (ii) results in awarding to the protagonist the outcome associated by the game with a specified strategy of the protagonist, combined with the strategy actually chosen by the adversary when playing the game. For each strategy s of the adversary, λ yields a mixed consequence λ_s in a natural way: if λ chose a consequence, then λ_s chooses the same consequence; and if λ chose a pure strategy r, then λ_s chooses the outcome of the game when the protagonist chooses r and the adversary chooses s. Note that all mixed consequences are in Δ, as are all mixed strategies; thus the preferences on Δ apply also to mixed consequences and to mixed strategies.

The following two assumptions are made:

N-M: The preference order satisfies the usual assumptions of von Neumann-Morgenstern utility theory[7]; and

MONOTONICITY: If one lottery λ always yields a mixed consequence preferred to that yielded by another lottery λ', no matter what the adversary does, then λ is preferred to λ'; likewise for weak preference.[8]

We then have

MAIN THEOREM: There exists a function on the consequences, called a *utility* function, and a probability distribution on the adversary's strategies, called a *subjective probability distribution,* such that one lottery is preferred to another if and only if its expected utility is greater.

Here the expected utility is calculated using the objective probabilities that define the relevant lotteries, and the subjective probabilities for the adversary's strategies.

That the utilities and subjective probabilities in this theorem enjoy the representation property is immediate. They also enjoy the interchangeability property, since expectations do not change when subjective probabilities are replaced by objective ones.

[7]As set forth in, say, von Neumann and Morgenstern [1944], or Luce and Raiffa [1957].

[8]Preference or indifference.

4. Formal Treatment: Preliminaries

In the following, terms being defined are italicized.

The set of all probability distributions[9] on a finite set A is denoted $\Delta(A)$; the probability assigned by an α in $\Delta(A)$ to an a in A is denoted α^a. Note that if $\alpha, \alpha' \in \Delta(A)$ and $t \in (0,1)$, then also $t\alpha + (1-t)\,\alpha' \in \Delta(A)$. Abusing our notation, we write α and a interchangeably if $\alpha^a = 1$; that is, we do not distinguish between a and a lottery that chooses a with certainty. No confusion should result.

A *preference order* $\succsim$ on $\Delta(A)$ is a transitive[10], reflexive[11], and complete[12] binary relation on $\Delta(A)$. If $\alpha \succsim \beta$ and $\beta \succsim \alpha$, write $\alpha \sim \beta$ and say that α is *indifferent* to β. If $\alpha \succsim \beta$ and $\alpha \nsim \beta$, write $\alpha \succ \beta$ and say that α is *preferred* to β. An *N-M utility* for $\succsim$ is a real valued-function u on $\Delta(A)$ such that.

(4.1) $\quad \alpha \succsim \alpha'$ iff $u(\alpha) \geq u(\alpha')$, and

(4.2) $\quad u(t\alpha + (1-t)\,\alpha') = tu(\alpha) + (1-t)\,u(\alpha')$.

Various mutually equivalent axiom systems for N-M utility theory are available (von Neumann and Morgenstern [1944], Luce and Raiffa [1957], and others). We say that a preference order *satisfies the axioms of von Neumann-Morgenstern utility theory*—and call it an *N-M preference order*—if it satisfies any one of those systems.

PROPOSITION 4.3: An N-M preference order on $\Delta(A)$ has an N-M utility.

5. The Main Theorem

The viewpoint taken here is that of a single player, the *protagonist,* also called *Rowena;* it is her subjective probabilities for the strategy choices of the other players that we will define. Also, the preferences appearing below are hers, as are the utilities. It is convenient to combine all the other players into a single one, called *Colin;* we will see that no loss of generality is involved.

A *game* G consists of

a finite set R with members r (Rowena's *pure strategies),*

a finite set S with members s (Colin's *pure strategies),*

[9] Non-negative real-valued functions whose values sum to 1.

[10] $\alpha \succsim \beta$ and $\beta \succsim \gamma$ imply $\alpha \succsim \gamma$.

[11] $\alpha \succsim \alpha$.

[12] $\alpha \succsim \beta$ or $\beta \succsim \alpha$.

a finite set C with members c *(pure consequences),* and

a function $h : R \times S \to C$ (the *outcome function*[13]*).*

Members $\rho, \rho', ...$ of $\Delta(R)$ are called *mixed strategies of Rowena;* members $\gamma, \gamma', ...$ of $\Delta(C)$ are called *mixed consequences;* members $\lambda, \lambda', ...$ of $\Delta(R \cup C)$ (henceforth simply Δ) are called *hybrid lotteries* (or simply *lotteries).* If $\lambda \in \Delta$, set $\lambda = t\rho + (1-t)\gamma$, where $\rho \in \Delta(R)$, $\gamma \in \Delta(C)$, and $t \in [0,1]$. For each pure strategy s of Colin, let $\rho_s \in \Delta(C)$ be the mixed consequence that results when Rowena plays ρ and Colin plays s, and let $\lambda_s := t\rho_s + (1-t)\gamma$ be the mixed consequence that results when Rowena uses the lottery λ and Colin plays s. Call a preference order $\succsim$ on Δ *monotonic* if $\lambda \succsim \lambda'$ whenever $\lambda_s \succsim \lambda'_s$ for all s, and $\lambda \succ \lambda'$ whenever $\lambda_s \succ \lambda'_s$ for all s.

Main Theorem: Let $G = (R, S, C, h)$ be a game, $\succsim$ a monotonic preference order on Δ that satisfies the axioms of von Neumann-Morgenstern utility theory. Then there exist an N-M utility u on $\Delta(C)$, and a probability distribution p on S, such that for any lotteries λ, λ',

(5.1) $\lambda \succsim \lambda'$ if and only if $\sum_{s \in S} p_s u(\lambda_s) \geq \sum_{s \in S} p_s u(\lambda'_s)$.

Define $u(\lambda) := \sum_{s \in S} p_s u(\lambda_s)$; then (5.1) becomes

(5.2) $\lambda \succsim \lambda'$ if and only if $u(\lambda) \geq u(\lambda')$.

In words, Rowena evaluates lotteries by their expected utility, when she ascribes probability p_s to Colin's choosing his pure strategy s.

6. Discussion

6.1. The Decision Criterion

The main theorem provides an unequivocal answer to the question raised in the introduction: consistent decision making obeys the same basic logic in games against nature and in games of strategy, namely the logic of subjectively expected utility. This calls for (i) parallel assessment of the utilities of consequences and the probabilities of contingencies, both assessments being summarized in cardinal measures; (ii) the integration of utility and probability considerations through the calculation of expected utility.

[13]W.l.o.g. we could take $C = R \times S$ and let h be the identity; but nothing would be gained thereby, the notation would become more cumbersome, and the ideas less transparent.

It is remarkable that this simple and elegant approach has general validity across contexts as different as games against nature and games of strategy. The main contribution of our analysis is the identification of the domain to which the principles of consistent decision making should be applied, namely hybrid lotteries over consequences and *strategies.* The unusual element here is the appearance of "preferences" over strategies. This is both *natural,* because strategies are *the* decision variable of the players; and *dérangeant*—a pain in the neck—because preferences over strategies are not directly observable, and typically, cannot be observed without changing the game (see Section 6.5).

The nature of preferences over strategies is brought out by the main theorem: one strategy r is "better" than another, r', if r carries higher expected utility, at the subjective probabilities that summarize the uncertainties of the situation.

In a way, the theory seems upside down. In practice, people *first* make probability (and utility) assessments, then use them to construct their preferences. Here, probabilities and utilities are *derived* from preferences; the preferences come first. But this is not different from the standard analysis of games against nature; it is not our innovation. The point is that decision theory is not so much a derivation of utilities and probabilities, but a way of putting some consistency and rationality into the process of using them.

6.2. Non-uniqueness

No claim of uniqueness is made for the subjective probabilities in the main theorem. The indeterminacy is due to our considering preferences only over strategies actually available to the protagonist. A sufficient condition for uniqueness is that there exist $|S|$ pure strategies $r^1, ..., r^{|S|}$ for which the vectors $(u(r^j_s))_{s \in S}$, $j = 1, ..., |S|$, are linearly independent; but this is by no means necessary. Thus if Rowena has just one pure strategy, yielding her outcomes with utilities $(0, 1, ..., 1)$, and this strategy is indifferent to an outcome with utility 0, then she must assign probability 1 to Colin's first strategy.

For a simple example where uniqueness fails, let G be a two-person game, $\succsim$ a monotonic N-M preference order on the hybrid lotteries. Then there exist a probability distribution p on the set S of Colin's strategies in G satisfying our conclusions. At least one of Colin's strategies, say s_0, must be assigned a positive probability, p_{s_0}, by p. Now consider a different game, G', which is the same as G except that the strategy s_0 is doubled—occurs in two copies, s'_0 and s''_0, with $h(r, s'_0) = h(r, s''_0)$ for all r; in particular, Rowena has the same strategies in G as in G'. Then precisely the same order, $\succsim$, when considered as a preference order on the hybrid lotteries in G', will also be monotonic and N-M; and the probability distributions that satisfy our conclusions will include any probability distribution p' that coincides with p for strategies s other than s_0' and s_0'', and such that $p_{s_0'} + p_{s_0''} = p_{s_0}$. Since $p_{s_0} > 0$, there is a continuum of such p'.

Needless to say, indeterminacy does not really matter (as is clear in the example just considered). What matters is the preference ordering, which is represented by the utilities together with *any* appropriate subjective probability distribution. The whole theory will apply with any such probabilities.

6.3. Hybrid Lotteries

Our treatment is similar to Anscombe and Aumann's [1963] treatment of games against nature (henceforth A-A), in that it takes the concept of objective probabilities as given and constructs subjective probabilities from them. This contrasts with Savage [1954], who does not formally differentiate between objective and subjective probabilities, constructing both at once. We have already remarked (Section 2) that our "strategies" are essentially the same as A-A's "horse lotteries:" Both are functions from the adversary's strategies to consequences. But in A-A, *all* such functions are available to the protagonist, at least in principle; whereas here—in strategic games—only the strategies specified in the definition of the game, and their mixtures, are available. This necessitates a rather different treatment.

Specifically, here Rowena's preferences are on a single set—$\Delta(C \cup R)$—of "hybrid" lotteries, each yielding consequences, strategies, or both. In contrast, A-A use two separate sets of lotteries: one—$\Delta(C)$—consists of mixed consequences only; the other—$\Delta(R)$—of mixed strategies only. For A-A that is enough, because in games against nature, the available strategies (members of R) include those described in the third paragraph of Section 2 above, which enable direct definitions of the subjective probabilities. But in a strategic game G, these strategies are not in general available, and the resulting space of lotteries is too "sparse" to allow defining subjective probabilities directly.

To see this, suppose given a preference order on the space $\Delta(C) \cup \Delta(R)$ consisting only of mixed consequences and mixed strategies. So we have preference orders on each of $\Delta(C)$ and $\Delta(R)$, and a "matching" of $\Delta(C)$ with $\Delta(R)$. Assume that each of the two preference orders is N-M, that the one on $\Delta(R)$ is monotonic, and that there is also monotonicity as "between" $\Delta(C)$ and $\Delta(R)$; i.e., if Rowena (weakly) prefers a mixed consequence γ to all outcomes of a mixed strategy ρ, no matter what Colin does, then she (weakly) prefers γ to ρ, and similarly in the opposite direction. In utility terms, we get N-M utility functions u on $\Delta(C)$ and $\Delta(R)$ that are *calibrated to the same scale,* so that it is meaningful to compare the utilities of mixed consequences and mixed strategies, and appropriate to use the same notation—u—for both.

A preference order of this kind is called *admissible.* Admissibility is about all one can ask for without going to hybrid lotteries. One might hope that it is enough to yield subjective probabilities with the requisite properties (representation and interchangeability), but it isn't.

For example, consider Game 1 (see display), consequences c being denoted by their utilities $u(c)$. Suppose $u(T) = \frac{3}{4}$ and $u(B) = \frac{1}{2}$. This induces an admissible preference order[14] on $\Delta(C) \cup \Delta(R)$. If q were the subjective probability of Colin's playing left, (5.2) would yield $\frac{3}{4} = u(T) = q \cdot 1 + (1-q) \cdot 0 = q$, and $\frac{1}{2} = u(B) = q \cdot \frac{2}{3} + (1-q) \cdot \frac{1}{3}$, so $q = \frac{1}{2}$, a contradiction.

	L	R
T	1	0
B	2/3	1/3

Game 1

Admitting hybrid lotteries—i.e., all members of $\Delta(C \cup R)$—adds requirements, as a result of which the example will no longer work. Indeed, let λ be the hybrid lottery yielding $1, 0$, and T with probability $\frac{1}{3}$ each. So $u(\lambda) = \frac{1}{3}1 + \frac{1}{3}0 + \frac{1}{3}u(T)$. But $B \sim \frac{1}{3}\lambda + \frac{2}{3}0$, so if $u(T) = \frac{3}{4}$, then $u(B) = \frac{1}{3}u(\lambda) + \frac{2}{3}0 = \frac{1}{9} + \frac{1}{9}u(T) = \frac{1}{9} + \frac{1}{12} = \frac{7}{36}$, which differs from the value of $\frac{1}{2}$ that the example stipulates for $u(B)$. Monotone preferences over $\Delta(C \cup R)$ imply subjective probabilities, as per the main theorem.

Another potential option is to forgo the "matching" between $\Delta(C)$ and $\Delta(R)$ (which "drives" the example) and to proceed from separate N-M preference orders on $\Delta(C)$ and $\Delta(R)$, obeying monotonicity on $\Delta(R)$. With this, one can indeed get subjective probabilities[15] that enjoy the representation property, but not interchangeability; indeed, if we can't compare mixed strategies with mixed outcomes,

[14] Monotonicity on $\Delta(R)$ is vacuously fulfilled, since for mixed strategies ρ and ρ', preferences between ρ_L and ρ'_L are opposite to those between ρ_R and ρ'_R. As between $\Delta(C)$ and $\Delta(R)$, let $\rho := \alpha T + (1-\alpha)B \in \Delta(R)$. Then $\rho \sim \frac{1}{2} + \frac{1}{4}\alpha$, which is always strictly between the consequences $\frac{2}{3} + \frac{1}{3}\alpha$ and $\frac{1}{3} - \frac{1}{3}\alpha$ that may result when ρ is played. Thus if γ is $\geq$ each ρ_s, then $\gamma \geq \frac{2}{3} + \frac{1}{3}\alpha > \frac{1}{2} + \frac{1}{4}\alpha \sim \rho$, and if each ρ_s is $\geq \gamma$, then $\rho \sim \frac{1}{2} + \frac{1}{4}\alpha > \frac{1}{3} - \frac{1}{3}\alpha \geq \gamma$.

[15] See Footnote 22.

interchangeability becomes meaningless. Hybrid lotteries thus remain as the only option if we are to get subjective probabilities with the appropriate properties.

And indeed, though they may appear artificial at first, closer examination reveals them as strategies in perfectly natural extensive games, of a kind often seen in real life. Specifically, each hybrid lottery λ may be seen as yielding a mixed strategy ρ with probability t, and a mixed outcome γ with the complementary probability $1 - t$. Denote by G^λ the extensive game in which nature chooses, with respective probabilities t and $1 - t$, whether G is to be played or whether the outcome is to be γ; then the lottery λ is equivalent (for the protagonist) to playing ρ in G^λ. The essential point is that one would not expect the adversary to play G^λ differently[16] from G, so the protagonist's preferences between the different lotteries λ accurately reflect her estimate of how the adversary will play G.

6.4. Application to Games against Nature

If in a game against nature, we allow the decision maker to consider only a *specified set* of "acts" (or "horse lotteries" in A-A)—and lotteries among them and their consequences—we get a theory formally identical to the one developed above. To be sure, in games against nature, one can always consider arbitrary "hypothetical" acts, without running into the conceptual difficulty that motivates this paper (Section 2). Nevertheless, the decision maker might have difficulty in expressing meaningful preferences as between highly hypothetical acts, and so prefer to restrict consideration to acts actually available. Our theorem applies to such situations. Similar considerations appear in the decision literature under the heading of "incomplete preferences."

6.5. Preferences among Strategies

In the behavioralist tradition of Savage [1954, p.18] and of revealed preference theory [Samuelson, 1950], preferences are derived from choices. Formally, the primitives consist of a *choice set* X—consisting of objects to be chosen—and a *choice function* f, taking each subset B of X to some member of B; in the interpretation, $f(B)$ is the object that the protagonist chooses when she must choose from among the objects in B. From this, one *derives* preferences over X, by defining $x \succ y$ if and only if $f(\{x, y\}) = x$. This has the advantage of "operationalism" or "observability:" choices are in principle observable, preferences not.

[16] For each player i, the matrix of G^λ is obtained from that of G by multiplying the whole matrix by the constant t and adding the constant $(1 - t)u^i(\gamma)$, where u^i is i's utility.

In our context, this poses a serious problem. Given two strategies ρ and ρ' of Rowena, one would have to define $\rho \succ \rho'$ if and only if Rowena would choose ρ *when she is confined to choosing between* ρ *and* ρ'. But confining her to $\{\rho, \rho'\}$ changes the game!

We have no unequivocal solution to this problem. In this work, preferences among strategies are treated as a primitive; we have found no totally satisfactory way of giving them a directly observable interpretation (see also Section 6.6.1).

While this is undoubtedly serious, it is not crippling. Certainly, operationalism and observability are vital principles. The question is, where do they fit in? Must the theory be *defined* in terms of observable entities, or must it have observable *implications*? Does one *start* with something observable, or *end* with it? One could perhaps go either way, but the second option—observable implications—certainly does not seem unreasonable.

It is, indeed, the way of most of science. None of the fundamental entities of physics—mass, energy, distance, gravitational force, etc., etc.—are directly observable. These theoretical entities interact in complex ways, which eventually, sometimes, allows one to measure them. Here, analogously, one starts with the non-operational idea of preference among strategies; from this, one develops the theory of utility and subjective probability, which eventually enables an "operational" characterization of choices.

There is a second point to be made. Suppose you must choose between x and y. You prefer x, and you choose x. Do you choose it because you prefer it, or do you prefer it because you choose it? Obviously the first makes more sense. Conceptually, preference comes before choice, it accounts for choice. So it makes sense to build the theory around preference and *derive* choice from it, just as actually happens in a person's mind.

Third, as mentioned in Section 6.1, one may think of decision theory not so much as a derivation of utilities and probabilities, but as a way of putting some consistency and rationality into the process of using them. The starting point—choices, preferences, or utilities and probabilities—is largely a matter of convenience: convenience and elegance in constructing the theory, spareness and transparency in assumptions, and so on. Certainly, observability must have a place, but not necessarily right at the start.

Finally, Rowena really does think in terms of choosing between any two of her strategies. People often do that when making decisions; they consider the options pairwise. It is part of the process of choosing between many options. Colin *knows* that that is what Rowena is doing; indeed, he himself is doing the same.

In contrast, the original difficulty, described in Section 2, involves Rowena considering a strategy that is not, in fact, available. As we have said, making it available will change Colin's view of the game, and so will affect his choice of strategies. That is quite different from considering the *available* strategies pairwise.

6.6. Alternative Approaches

6.6.1. Side Bets

A potentially operational approach to defining and eliciting preferences over strategies relies on "side bets." The idea is to construct, for each strategy ρ of Rowena, a different game G_t^ρ, defined as a lottery that results either in (i) G being played, or (ii) Colin getting 0 for sure, and Rowena getting ρ_s if Colin played s, for all s in S, where the respective probabilities of (i) and (ii) are t and $1-t$, and where the players must decide *beforehand* on the strategies they must use if the lottery chooses G.

On the face of it, the incentives of neither player seem changed; it would appear that G_t^ρ is "equivalent" to G, in the sense that both players should play in G_t^ρ precisely as they do in G. If that were the case, one could say that ρ is "preferrred" to another strategy ρ' of Rowena if and only if Rowena would rather play in G_t^ρ than in $G_t^{\rho'}$.

But the appearance that G_t^ρ is "equivalent" to G is deceiving. For an example,[17] let G be Game 2 in the display. In this game, one might well expect the players to play BR, which is Pareto dominant. Now consider Game 3, obtained from G by adding 8 times Rowena's row T to her payoffs. This is $9G_{1/9}^T$; if G_t^ρ is equivalent to G, then so is $9G_t^\rho$, so one would expect the players to play BR in Game 3 as well. But by the same token, adding 8 times Colin's column L to his payoffs should not matter either; this brings us to Game 4, commonly known as the "Stag Hunt[18]." But here, it is far from compelling that the players would play BR, which is Pareto dominated by TL. Indeed, in their celebrated work on equilibrium selection, Harsanyi and Selten [1987] pick BR in Game 2 and TL in Game 4.

[17]Communicated by S. Hart.

[18]See O'Neill [1994, pp.1004-5] for a discussion of this game and some of the literature on it.

	L	R
T	$1,1$	$0,0$
B	$0,0$	$7,7$

Game 2

	L	R
T	$9,1$	$0,0$
B	$8,0$	$7,7$

Game 3

	L	R
T	$9,9$	$0,8$
B	$8,0$	$7,7$

Game 4

Consequently, the games G_t^ρ may not be substituted for G in an attempt to base preferences among strategies on observed choices, so the attempt fails.

Like the G_t^ρ, the "hybrid" lotteries λ on which our analysis is based (Sections 3 and 6) involve playing G with some probability, and otherwise, Rowena getting an amount (in utility terms) that is fixed—i.e., independent of her choice. But in the G_t^ρ, the "fixed amount" is contingent on Colin's choice, whereas in λ it is not. Thus λ is in fact the same as G, with entries subjected to a positive linear transformation—which is indeed equivalent to G.

6.6.2. Ignorance

In games against nature, the probability of a state s of nature is the number p such that Rowena is indifferent between a lottery yielding a dollar (or some other desirable outcome) with objective probability p, and a bet r yielding a dollar if and only if s occurs. The fundamental difficulty set forth in Section 2 is that when s is a strategy of Colin in a strategic game G, then r is usually not available to Rowena; making it available may change Colin's view of the game, and so Rowena's probability that he will choose s.

One might think it enough to consider a situation in which r really *is* available to Rowena, but *Colin does not know that it is,* so his choices—and Rowena's probabilities for his making those choices—will not be affected. But that is not very satisfactory. What *does* Colin know? Does he consider it possible that r is available to Rowena? If he does, then already that changes the game G, and we have the same difficulty as before. If he considers it impossible—*knows* that r is unavailable—how can it be available? There is a basic incoherence in situations where something false is "known."

Similar remarks apply to other ignorance-based approaches. E.g., to interpreting a preference between strategies ρ and ρ' "operationally" by really confining Rowena to choosing between ρ and ρ', while stipulating that Colin "does not know" that. Or, to procedures applied "after the game is played," while stipulating that when playing the game, Colin "does not know" that the procedure will (or might) be applied.

A conceivable resolution is based on the idea of something that is possible but has probability zero. There could be a situation in which a priori it is commonly known that r may be available to Rowena, but that this happens with probability 0. After that, Rowena, but not Colin, is informed that r is indeed available. Thus Colin believes that with probability 1, the only strategies available to Rowena are those specified in G. He therefore plays as in G, and so Rowena can indeed use r (and other lotteries) to estimate her probabilities for Colin's strategies in G.

While this resolution *is* coherent, it is nevertheless unsatisfactory. To base the entire theory of strategic decision-making on the difference between probability 0 and impossibility will not pass muster—it is simply too precious.

6.7. Reversal of Order

A-A use a formal assumption called "Reversal of Order in Compound Lotteries." It may be interpreted as saying that if the outcome of a compound lottery depends in a *fixed, predetermined way* both on a roulette spin and a horse race, then it is immaterial whether the wheel is spun before or after the race. That is, a decision maker will be indifferent between these two possibilities, as long as he *cannot change* his horse bets after the roulette spin, or his roulette bets after the horse race (e.g., if he is not informed of the first lottery's outcome before the second one is performed). Though lacking substantive content in the A-A context, the assumption is required in their formalism[19].

In our treatment, no explicit assumption of this kind is necessary; whatever is needed in this direction is already implicit in the formalism. Indeed, it follows from monotonicity on $\Delta(R \cup C)$ that if $\lambda_s = \lambda'_s$ for all s, then $\lambda \sim \lambda'$, which may be considered an expression of the reversal of order assumption.

6.8. Value

It is natural to define the value v to Rowena of a game G as her expected utility when she plays the game optimally, given her estimate as to how Colin will play; i.e., in the notation of Section 6, $v := \max\{\sum_{s \in S} p_s u(r_s) : r \in R\}$. Unlike the minmax value of two-person zero-sum games, this does not call for Colin's play to be "optimal" in any sense; it is based only on Rowena's subjective estimate of how Colin will play. Indeed, in a two-person zero sum game, this definition may well yield results that are larger than the minmax value. To get the minmax value of

[19]See Dreze [1987] for a context in which a similar assumption does have substantive content.

two-person zero-sum games from this definition, one must assume common priors and common knowledge of rationality, as we show in forthcoming work.

6.9. Literature

Luce and Raiffa [1957, p.306] were among the earliest[20] to suggest assigning subjective probabilities to an adversary's choices in a strategic game; they wrote as follows: "The problem of individual decision making under uncertainty can be considered a one-person game against a neutral nature. Some of these ideas can be applied indirectly to individual decision making ... where the adversary is not neutral but a true adversary. ... One *modus operandi* for the decision maker is to generate an *a priori* probability distribution over the ... pure strategies ... of his adversary by taking into account both the strategic aspects of the game and ... 'psychological' information ... about his adversary, and to choose an act which is best against this ... distribution." They go on to explore the idea of "side bets" (see 6.6.1 above), noting some difficulties with it, and informally suggesting a possible way around them. No formal model was developed, and no definite conclusion reached.

It appears that Armbruster and Böge [1979] and Böge and Eisele [1979] were the first to construct formal models in which each player directly[21] assigns subjective probabilities to the strategy choices of the others. A relatively early application of this idea is Brandenburger and Dekel [1987]. The representation of the value of a game to a player as a subjectively expected utility is implicit in the work of Nau and McCardle [1990].

Kadane and Larkey [1982] suggested abandoning altogether all notions of equilibrium and any "interactive" analysis (i.e., analysis of each player's beliefs about the others' beliefs about his actions and beliefs). Instead, they proposed simply that each player form, in some unspecified and unrestricted way, a probability distribution over the other players' strategies, and then maximize against that. To form the probabilities, they suggested using disciplines like cognitive psychology

[20] They cite an earlier paper by Hodges and Lehmann [1952] who suggest that a player in a two-person zero-sum game might assign subjective probabilities to the eventuality that his adversary will make a "mistake." But this is not really in the spirit of this paper, nor of Luce and Raiffa's suggestion.

[21] Previously, Aumann [1974] had already used subjective probability in analyzing games; but in that analysis, players use "subjectively mixed strategies"—peg their pure strategy choices on events (like outcomes of horse races) whose probability is not agreed upon—rather than simply assigning a subjective probability to the other players' choices.

rather than decision or game theory.

None of the authors cited above, nor indeed others who wrote later, seemed aware of the fundamental difficulty addressed by this paper: that the standard construction of subjective probabilities does not apply to strategic games.

7. Affine Monotonic Functions

For points x, y in $\mathbb{R}^n$, write $x > y$ if $x_i > y_i$ for all i, and write $x \geq y$ if $x_i \geq y_i$ for all i. A real-valued function f from a convex set D in $\mathbb{R}^n$ to $\mathbb{R}$ is called *affine* if $f(tx + (1-t)y) = tf(x) + (1-t)f(y)$ for all x, y in D and t in $(0, 1)$. It is called *monotonic* if $x > y$ implies $f(x) > f(y)$, and $x \geq y$ implies $f(x) \geq f(y)$, for all x, y in D.

PROPOSITION 7.1: Let H be a convex subset of $\mathbb{R}^n$, and f an affine monotonic real-valued function on H. Assume[22] that there are points z and z' in H with $z > z'$. Then there exist non-negative $q_1, ..., q_n$, not all of which vanish, and a real q_0, such that $f(x) = q_0 + \sum_{i=1}^n q_i x_i$ for all x in H.

In this section we prove 7.1; readers willing to accept the proposition on faith may proceed at once to the proof of the main result in the next section.

The origin $(0, ..., 0)$ of $\mathbb{R}^n$ is denoted $\mathbf{0}$. A *linear subspace* M of $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that, together with any two points x, y in it, and any real number t, contains $x+y$ and tx. A function f on M is *linear* if $f(x+y) = f(x)+f(y)$ and $f(tx) = tf(x)$ for all x, y in M and all real t. Note that a function on M is linear if and only if it is affine. A *hyperplane* in M is a set of the form $\{x \in M : f(x) = t\}$, where f is a linear function on M that does not vanish identically, and t is a constant.

LEMMA 7.2: If D is an open convex set in $\mathbb{R}^n$ and M a linear subspace of $\mathbb{R}^n$ that does not meet D, then there is a hyperplane in $\mathbb{R}^n$ that includes M and does not meet D; i.e., a linear function on $\mathbb{R}^n$ that vanishes on M and is negative on D.

PROOF: Eggleston [1958], p.19, Theorem 7.

Define $\mathbb{R}^n_- := \{x \in \mathbb{R}^n : x < \mathbf{0}\}$.

LEMMA 7.3: Let L be a linear subspace of $\mathbb{R}^n$ that intersects $\mathbb{R}^n_-$. Then any monotonic linear function f on L may be extended to a monotonic linear function on all of $\mathbb{R}^n$.

[22] The proposition is true as it stands even without this assumption, but the proof lies deeper.

Proof: Define $L_0 := \{x \in L : f(x) = 0\}$. Because f is monotonic, L_0 cannot meet $\mathbb{R}^n_-$. Use 7.2 to find a linear function f^* on $\mathbb{R}^n$ that vanishes on L_0 and is negative on $\mathbb{R}^n_-$. Let $-x_1$ be a point in $\mathbb{R}^n_- \cap L$. Then $-f(x_1) = f(-x_1) < 0$, so $f(x_1) > 0$. Similarly $f^*(x_1) > 0$. Possibly redefining f^* by multiplication by a positive constant, we may take $f(x_1) = f^*(x_1)$. So f and f^* coincide on L_0 and on x_1, which together span L, so they coincide on L. Thus f^* extends f, and since it is negative on $\mathbb{R}^n_-$, it is monotonic. This completes the proof of 7.3.

Proof of Proposition 7.1. By possibly applying a translation, we may suppose w.l.o.g. that $\mathbf{0}$ is in the interior of H, relative to the smallest linear manifold L that includes H. Since $\mathbf{0} \in L$, it follows that L is a linear space, and there is a unique extension of f to a linear function f' on L. Both z and z' are in H and so in L, so $z' - z$ is in L, so L intersects $\mathbb{R}^n_-$, since $z' - z < \mathbf{0}$. To see that f' is monotonic, first let $x > \mathbf{0}$. Then for sufficiently small positive ε, we have $\varepsilon x \in L$, and $\varepsilon x > 0$, so $\varepsilon f'(x) = f'(\varepsilon x) = f(\varepsilon x) > 0$, so $f'(x) > 0$. Similarly, $x \geq 0$ implies $f'(x) \geq 0$. Thus f' is monotonic as well as linear. Applying 7.3, we obtain an extension f'' of f' from L to a linear monotonic function on all of $\mathbb{R}^n$. Let $f''(x) = \sum_{i=1}^n q_i x_i$ for all x in $\mathbb{R}^n$; all linear functions on $\mathbb{R}^n$ have this form. It cannot be that all the q_i vanish, for then $f''(1, ..., 1) = 0 = f(\mathbf{0})$, contrary to the monotonicity of f''. To show that the q_i are non-negative, suppose, say, that $q_1 < 0$. Then $f''(1, 0, ..., 0) = q_1 < 0$, again contradicting the monotonicity of f''. So the q_i are non-negative and do not all vanish, proving[23] the proposition.

8. Proof of the Main Theorem

By Proposition 4.3, $\succsim$ has an N-M utility u; from 4.1, it follows that

(8.1) $\quad \lambda \succ \lambda'$ iff $u(\lambda) > u(\lambda')$.

W.l.o.g., we assume that there are pure consequences d and d' with[24]

(8.2) $\quad d \succ d'$.

Set $S := \{s_1, ..., s_n\}$. With each lottery λ in Δ, associate the point $u(\lambda_S) := (u(\lambda_{s_1}), ..., u(\lambda_{s_n}))$ in $\mathbb{R}^n$. Note that when λ chooses some mixed consequence γ for sure (rather than playing the game with positive probability)—i.e., when $\lambda \in \Delta(C)$—then $\lambda_{s_1} = ... = \lambda_{s_n} = \gamma$, so

(8.3) $\quad u(\lambda_S) := (u(\lambda_{s_1}), ..., u(\lambda_{s_n})) = (u(\gamma), ..., u(\gamma))$.

[23]The term q_0 is due to the translation at the beginning of the proof.

[24]If all consequences are indifferent, take u to be identically 0, and p an arbitrary distribution.

Let H be the set of all the points $u(\lambda_S)$ when λ ranges over Δ. By 4.2, we have $tu(\lambda_s) + (1-t)u(\lambda'_s) = u(t\lambda_s + (1-t)\lambda'_s) = u((t\lambda + (1-t)\lambda')_s)$ for each s in S and t in $[0, 1]$ so

(8.4) $\quad tu(\lambda_S) + (1-t)u(\lambda'_S) = u((t\lambda + (1-t)\lambda')_S)$,

so H is convex. Moreover, if d and d' are as in 8.2, then by 8.3, H contains the points $z := (u(d), ..., u(d))$ and $z' := (u(d'), ..., u(d'))$, so by 4.1, $z > z'$. Thus H satisfies the hypotheses of 7.1.

Now define a function f on H by $f(x) := u(\lambda)$ for any λ for which $u(\lambda_S) = x$; that there is such a λ follows from $x \in H$, and that it doesn't matter which one we use follows from monotonicity: If $u(\lambda_S) = u(\lambda'_S)$, then $\lambda \sim \lambda'$, by 5.1 and monotonicity, so $u(\lambda) = u(\lambda')$, by 8.1. Next, we show that f is affine and monotonic. Indeed, let $x = u(\lambda_S)$ and $x' = u(\lambda'_S)$; then by 8.4 and 4.2,

$f(tx + (1-t)x') = f(tu(\lambda_S) + (1-t)u(\lambda'_S)) = f(u((t\lambda + (1-t)\lambda')_S)) =$
$= u(t\lambda + (1-t)\lambda') = tu(\lambda) + (1-t)u(\lambda') = tf(x) + (1-t)f(x')$, proving that

(8.5) $\quad f$ is affine.

To show that

(8.6) $\quad f$ is monotonic,

first let $x \geq x'$; then $u(\lambda_S) \geq u(\lambda'_S)$, so $u(\lambda_s) \geq u(\lambda'_s)$ for all s, so $\lambda_s \succsim \lambda'_s$ for all s, (by 4.1), so $\lambda \succsim \lambda'$ (by monotonicity), so $f(x) = u(\lambda) \geq u(\lambda') = f(x')$, by 8.1. Similarly, if $x > x'$, then $f(x) > f(x')$, proving 8.6. So by 7.1, there exist non-negative $q_1, ..., q_n$, not all of which vanish, and a real q_0, such that $f(x) = q_0 + \sum_{i=1}^n q_i x_i$ for all x in H. So $u(\lambda) = q_0 + \sum_{i=1}^n q_i u(\lambda_S)_i = q_0 + \sum_{i=1}^n q_i u(\lambda_{s_i})$. So by 8.1,

(8.7) $\quad \lambda \succsim \lambda'$ if and only if $\sum_{i=1}^n q_i u(\lambda_{s_i}) \geq \sum_{i=1}^n q_i u(\lambda'_{s_i})$.

Set $p_i = q_i / \sum_{i=1}^n q_i$; the denominator does not vanish because the q_i are non-negative and do not all vanish. So the p_i are non-negative and sum to 1; that is, they constitute a probability distribution on $S = \{s_1, ..., s_n\}$. The conclusion of the main theorem then follows from 8.7.

9. References

Anscombe, F. J., and R. J. Aumann (1963), "A Definition of Subjective Probability," *Ann. Math. Stat.* 34, 199-205 (**A-A** in the text).

Aumann, R. J. (1974), "Subjectivity and Correlation in Randomized Strategies," *J. Math. Econ.* 1, 67-96.

Brandenburger, A., and E. Dekel (1987), "Rationalizability and Correlated Equilibria," *Econometrica* 55, 1391-1402.

Dreze, J. H. (1961), "Les fondements logiques de l'utilité cardinale et de la probabilité subjective," in *La Décision,* Paris: CNRS, 73-87.

—— (1987), *Essays on Economic Decisions under Uncertainty,* Cambridge, at the University Press.

Eggleston, H. G. (1958), *Convexity,* Cambridge, at the University Press.

de Finetti, B. (1937), "La Prevision: Ses Lois Logiques, ses Sources Subjectives," *Ann. Inst. H. Poincaré* 7, 1-68.

Harsanyi, J. C., and R. Selten (1987), *A General Theory of Equilibrium Selection in Games,* Cambridge, Mass.: MIT Press.

Kadane, J. B., and P. D. Larkey (1982), "Subjective Probability and the Theory of Games," *Manag. Sci.* 28, 113-120.

Luce, R. D., and H. Raiffa (1957), *Games and Decisions,* New York: John Wiley.

Hodges, J. L. Jr., and E. L. Lehmann (1952), "The uses of previous experience in reaching statistical decisions," *Ann. Math. Stat.* 23, 396-407.

Nash, J. F. (1951), "Non-cooperative Games," *Ann. Math.* 54, 286-295.

Nau, R. F. and K. F. McCardle (1990), "Coherent behaviour in non-cooperative games", *J. Econ. Th.* 50, 424-44.

von Neumann, J., and O. Morgenstern (1944), *Theory of Games and Economic Behavior,* Princeton: Princeton University Press.

O'Neill, B., (1994), "Game Theory Models of Peace and War," in *Handbook of Game Theory with Economic Applications, Vol.2,* R.J. Aumann and S.Hart, eds., Amsterdam: Elsevier, 996-1053.

Ramsey, F. P., (1931), "Truth and Probability," Chapter VII of *The Foundations of Mathematics,* edited by R.B. Braithwaite, London: Routledge and Kegan Paul, 156-198.

Samuelson, P. A., (1950) "The Problem of Integrability in Utility Theory," *Economica* NS 17, 355-385.

Savage, L. J. (1954), *The Foundations of Statistics,* New York: John Wiley.