28 Mixed and Behavior Strategies in Infinite Extensive Games

1 Introduction

We are concerned with infinite extensive games—not necessarily of perfect information—in which there may be a continuum of alternatives at some or all the moves; the games may also have unbounded or infinite play length. Our object is to define the notion of mixed strategy for such games, and to use this definition to prove the appropriate generalization of Kuhn's theorem on optimal behavior strategies in games of perfect recall [K_1]. Also, our methods give a solution to the conceptual problem raised by McKinsey under the heading "games played over function space" [Mc, pp. 355–357].

By-products are that our proof of Kuhn's theorem makes no use of the rather cumbersome "tree" model for extensive games, that it explicitly uses conditional probabilities (which are implicitly used by Kuhn), and that it explicitly proves that in a game which is of perfect recall for one player, that player can restrict himself to behavior strategies (this also is implicit in Kuhn's proof). Our proof is longer and more complicated than Kuhn's proof, but only because of the problems introduced by the non-denumerably infinite character of the game; the treatment of finite games by our methods would be considerably shorter.

2 Examples

We give four examples to motivate this study and to illustrate some of the difficulties.

In our first example, there are two players, the "attacker" and the "defender"; for concreteness, one may think of the attacker as a bomber. The attacker starts the play by choosing a course of action (such as a flight course). The defender has some mechanism (such as radar) for determining the course chosen by the attacker, and he decides on *his* course of action on the basis of the information he gets from this mechanism. But the mechanism is not perfect; it only gives an *apparent* attacker course x, which is distributed around the true attacker course z according to a known probability distribution (which may vary with z). Thus the defender gets some information about the attacker's course, but not perfect information.

This chapter originally appeared in *Advances in Game Theory*, Annals of Mathematics Studies 52, edited by M. Dresher, L. S. Shapley, and A. W. Tucker, pp. 627–650, Princeton University Press, Princeton, 1964. Reprinted with permission.

Denote by X the set of all possible apparent attacker courses, i.e., the set of possible information states of the defender. Denote by Y the set of courses of action available to the defender. Clearly a pure defender strategy is a function from X into Y. What about mixed strategies? If X and Y are finite, then there are only finitely many pure strategies, so there is no difficulty about defining mixed strategies. But in many cases the most appropriate model would be one in which X and Y are, say, copies of the unit interval. It is then still possible to define *some* kinds of mixed strategies; for example, we can mix finitely or denumerably many pure strategies, or we can adopt a fixed continuous distribution over Y regardless of what information we have—i.e., we can mix a continuum of pure strategies, each of which is a constant function from X into Y. But is this the best we can do? Can't we mix a continuum of pure strategies that are *not* constants?

A mixed strategy can be thought of as a probability distribution, i.e., a measure, on the set of all pure strategies. But before one defines a measure on a nondenumerable space, one must define a *measurable structure* on the space, i.e., one must define which subsets are measurable. It is by no means clear how this should be done in our case, or even what kind of measurable structure on the pure strategy space should be considered "appropriate" for this purpose.

For our second example we can do no better than quote McKinsey [Mc, p. 356]:

> "A game has four moves: in the first move P_1 (player 1) chooses a real number x_1; in the second move, P_2, knowing x_1, chooses a real number y_1; in the third move, P_1, knowing y_1, but having forgotten x_1, chooses a real number x_2; and in the last move, P_2, knowing y_1 and x_2, but not knowing x_1, chooses a real number y_2. (The payoff is then some function of the four variables x_1, x_2, y_1, and y_2.) A pure strategy for P_1 is now an ordered couple $\{a, f\}$, where a is a real number and f is a function of one real variable (it depends on y_1); and a pure strategy for P_2 is an ordered couple $\{g, h\}$, where g is a function of one real variable (it depends on x_1) and h is a function of two real variables (it depends on y_1 and x_2)....
>
> It is clear that the payoff function for a game of the type just described need not necessarily have a saddle point, and hence it is natural to suppose that the players will make use of mixed strategies...."

The difficulties that McKinsey goes on to describe correspond precisely to those we discussed in connection with the first example.

Our third example involves the notion of the *supergame* of a given game G. This is a game each play of which consists of a number of repeated plays of G; the payoff to the "superplay" usually is defined as some kind of average of the payoffs to the individual plays. The supergame and related notions[1] have received considerable attention in the literature; this is partly because supergames occur naturally in the applications, and partly because an analysis of a supergame sometimes yields clues as to rational behavior for a single play.[2]

Supergames are usually analyzed on a step-by-step basis; that is, it is assumed that each player decides on a strategy for each of the component plays separately. These "local" strategies may or may not depend on the outcomes of the previous component plays, and may be pure or mixed; but the possibility of mixing a number of pure "grand strategies" for the whole supergame is usually ignored. Of course this makes analysis of the supergame much easier.

The supergame may be considered a game in extensive form, a move being a choice of a pure strategy for a component play. Obviously it is a game of perfect recall—at each component play each player remembers what he knew at previous component plays. What we are doing when we limit analysis of the supergame to consideration of mixed strategies for the component plays is that we are considering only *behavior* strategies in the supergame. Now we lose no generality by this restriction if Kuhn's theorem on behavior strategies in games of perfect recall[3] applies, which is the case when the originally given game is finite and is only repeated finitely often. Wolfe [W, p. 15] has pointed out that Kuhn's theorem may be extended to games with infinite play length, and it is easily seen that we can also allow a denumerable infinity of alternatives at some (or all) of the moves. The difficulties enter when there may be a *continuum* of alternatives at some of the moves; in our case this corresponds to a G with a continuum of strategies.

What is the importance of supergames of games with a continuum of strategies? Suppose we wish to consider the supergame of a cooperative game. To analyze this supergame properly, we must formalize the pre-play bargaining for each component play. Such a formalization *must* involve a continuum of pure strategies for the bargaining session—for example we already have a continuum in the set of correlated strategies that can be offered by a player for the consideration of a coalition that he wishes to form.[4] Thus a satisfactory analysis of a cooperative supergame cannot proceed without first proving an analogue of Kuhn's theorem for the continuous case. Indeed it was this problem that originally motivated this study.

1. Such as that of stochastic game.

2. Cf. [Mc, the discussion at top of p. 134]; also [A_1] and [A_2, §10].

3. Kuhn's theorem asserts that in a game of perfect recall each mixed strategy m has an equivalent behavior strategy, i.e., a behavior strategy which yields the same payoff as m (to all players) no matter what the other players do.

4. Cf. [A_1, §6] or [A_2, §10].

For our last example, we start by recalling Ville's theorem [V]: Every 2-person 0-sum game on the unit square with continuous payoff has optimal mixed strategies, and hence a value. Now in a game on the unit square, each player picks a point in the unit interval $[0,1]$, with no knowledge of the point chosen by the other; the payoff is a function of the two points chosen. Fox [F] asked whether Ville's result could be extended to multimove games of imperfect (but partial) information, where each move consists of the choice of a point in the unit interval. In these games, the players alternatively choose points $y_i \in [0,1]$ for $i = 1, \ldots, 2n$, say. After the i^{th} move the player who is to make the $(i+1)^{\text{st}}$ move is informed of the value of $\psi_i(y_1, \ldots, y_i)$, where ψ_i is a real function. The payoff is $f(y_1, \ldots, y_{2n})$, where f is a continuous, real function. Note that the space of pure strategies of, say, player 1 is $\mathscr{F} = [0,1] \times F_3 \times \cdots \times F_{2n-1}$, where F_i is the set of all functions from the range of ψ_{i-1} to $[0,1]$.

The object of [F] is to show that in general Ville's result does not extend to this situation, and there is no value. Of course, before this can be done the concept of mixed strategy must be defined. When each ψ_i has finite range, then each F_i and therefore also $\mathscr{F}$ is a product of finitely many copies of $[0,1]$; in that case, therefore, mixed strategies can be defined as multidimensional distribution functions. When the range of ψ_i is a continuum, however, it is by no means clear how to define mixed strategies. One of Fox's examples, in which the ψ_i as well as f are continuous, is precisely of this kind.

Fox defined mixed strategies as finite mixtures of pure strategies. But this is not at all satisfactory, especially when one wants to show that a game does *not* have a value. Indeed, there is an example of a game on the unit square which has value 0 when distributions are admitted as mixed strategies, but for which the supinf of the payoff is -1 when one is restricted to finite mixtures of pure strategies [K_2, p. 118]. A priori, it could well be that the counterexamples of [F] vanish when a wider, more natural class of mixed strategies is admitted. In fact, it appears that this does not happen, since the proof apparently makes no essential use of the restrictive nature of the mixed strategies. It is the statement, rather than the proof, that is unsatisfactory; and the reason is that up to now there has been no definition of mixed strategy appropriate to such games.

3 Mixed Strategies

Let us take a closer look at the first example in the previous section; take X and Y to be copies of the unit interval. We shall need to consider probability distributions on X and Y, and as we remarked in the previous section, this involves defining measurable structures on them. Any such measurable structure should be rich enough to enable us to define the probability of an interval; this means it would have to contain all Borel sets. Let us denote by I the unit interval on which has been imposed a measurable structure consisting of all the Borel sets, and let us once and for all[5] take X and Y to be copies of I.

Henceforth we will write "m-" for "measurable."

Suppose the defender has adopted a strategy f, i.e., a function from X into Y, and that the action of chance and the strategy of the attacker have induced a probability distribution on X. The strategy f, acting on this X-distribution, should induce a distribution on Y. Does it? Suppose $B \subset Y$ is a Borel set. The probability that a member of B is chosen by the defender

$$= \text{prob}\{x : f(x) \in B\}$$
$$= \text{prob}\{f^{-1}(B)\}.$$

This expression is meaningless unless $f^{-1}(B)$ is measurable in X: The same holds for all m-subsets B of Y. In order to have an induced distribution on Y, we want the inverse image under f of a measurable set in Y to be measurable in X. In other words, we want f to be a measurable transformation. So we redefine a pure defender strategy; it is not just any function from X into Y, but an m-transformation.[6] We denote by Y^X the set of all m-transformations from X into Y.

A *mixed* strategy, then, should be a probability measure on Y^X, the latter having been endowed with an "appropriate" measurable structure R. Let us define a function $\varphi : Y^X \times X \to Y$ by $\varphi(f, x) = f(x)$. Suppose we again start out with a distribution on X, and suppose that the

5. We have adopted the smallest structure that fills our needs. An overly rich structure is self-defeating. For example, if the structure on X consists of *all* subsets, then the only measures on X are purely atomic (under the continuum hypothesis [S, p. 107]); if it consists of all Lebesgue measurable sets, then the only measures are sums of absolutely continuous and purely atomic ones (thus excluding all those with a singular non-atomic component). We therefore see that increasing the set of measurable sets beyond a certain point actually reduces the set of available measures. If we want all intervals to be measurable, the largest set of measures is obtained if we let the structure consist of the Borel sets. (In this connection we remark that there is a confusing misprint in [Mc, p. 357, line 7]; here "Lebesgue measurable" should read "Borel.")

6. This redefinition of pure strategy is a consequence of the demand that distributions on X induce distributions on Y. Besides being intuitively desirable, this is absolutely necessary for the formal analysis, as the reader will see later. Perhaps the most compelling intuitive argument, though, is that this is needed so that a pair of pure attacker and defender strategies should induce a *payoff* distribution, for example so that we should be able to assign a probability to the attacker's payoff being positive.

defender has chosen a mixed strategy; we wish to calculate the induced distribution on Y. For m-sets $B \subset Y$, the probability that the defender chooses a member of B

$$= \text{prob}\{(f, x) : f(x) \in B\}$$
$$= \text{prob}\{(f, x) : \varphi(f, x) \in B\}$$
$$= \text{prob}\{\varphi^{-1}(B)\}.$$

As before, we conclude that *the structure R must be chosen so that φ is an m-transformation*. But as we have shown elsewhere [A_3], there is no structure R for which this is so; *no* structure on Y^X is "appropriate"!

We have attempted to define mixed strategies as distributions (i.e., probability measures) on Y^X, and have run into difficulties. In fact these difficulties can be overcome, at least to a certain extent (cf. [A_4]). In the current context, though, we find that a completely different approach is both more convenient and more natural. Let us recall the intuitive meaning of a mixed strategy: It is a method for choosing a pure strategy by the use of a random device. Physically, one tosses a coin, and according as to which side comes up chooses a corresponding pure strategy; or, if one wants to randomize over a continuum of pure strategies, one uses a continuous roulette wheel. Mathematically, the random device—the set of sides of the coin or of points on the edge of the roulette wheel—constitutes a probability measure space, sometimes called the *sample space*; a mixed strategy is a function from this sample space to the set of all pure strategies. In other words *what we have here is precisely a random variable whose values are pure strategies*. We previously attempted to work with something corresponding to the distribution of this random variable; now we propose to use the random variable itself.

Let us denote by Ω the measure space that results when we impose Lebesgue measure on I. All of our sample spaces will be copies of Ω. The intuitive justification for this is that every "real-life" random device is either "discrete," "continuous," or a combination of the two; that is, the sample space involved must either be finite or denumerable, or it must be a copy of I (with a measure that is not necessarily Lebesgue measure).[7] All such random devices can be represented by random variables whose sample space is actually a copy of Ω.

In our example, therefore, we should define a mixed strategy to be a function from Ω to the space Y^X of all pure strategies. We can expect that not all such functions will be "eligible" as mixed strategies, because

7. Physically, of course, all sample spaces are discrete and even finite; but it is often convenient to use a continuous or a denumerable model.

of the by now familiar condition that a mixed strategy and a distribution on X must induce a distribution on Y. Fortunately, the appropriate condition is *not* that the mixed strategy as defined above be a measurable transformation, because this would again involve defining a measurable structure on Y^X. To state the correct condition, we recall that to every function from Ω to Y^X there is a corresponding function from $\Omega \times X$ to Y; to $f : \Omega \to Y^X$ there corresponds the function $g : \Omega \times X \to Y$ defined by $g(\omega, x) = f(\omega)(x)$. The correct condition on a mixed strategy is that this *corresponding* function be an m-transformation. Thus we *define a mixed strategy to be an m-transformation from* $\Omega \times X$ *into* Y.

As it now stands, this definition of mixed strategy applies only to the highly simplified situation treated in the first example of the introduction. However, it can be extended without difficulty to more complicated, many-move games, as we shall show in Section 5.

4 Extensive Games

In this section we present the formal structure for extensive games that we will use in the sequel. We first give the definitions, then discuss their intuitive meaning and their relation to other definitions in the literature.

An m-space is called *standard*[8] if it is either finite or denumerable with the discrete structure (i.e., all subsets are measurable), or if it is isomorphic[9] with I. Most m-spaces that one "encounters in practice" are standard; for example, any Borel subset of any Euclidean space or of Hilbert space is standard.

DEFINITION A *game from an individual player's viewpoint*, or simply a *game*, consists of

(i) A (finite or infinite) sequence $Y_1, Y_2, \ldots$ of standard m-spaces called *action spaces*;

(ii) A corresponding sequence $X_1, X_2, \ldots$ of standard m-spaces called *information spaces*;

(iii) A set Z called the set of *strategies of the opponents*;

(iv) A sequence of functions

$g_i : Z \times Y_1 \times \cdots \times Y_{i-1} \to X_i,$

called *information functions*, which for each fixed $z \in Z$, are m-transformations on $Y_1 \times \cdots \times Y_{i-1}$ into X_i;

8. This use of the word is due to Mackey [M].

9. An *isomorphism* is a one-one correspondence that is measurable in both directions.

(v) A standard m-space H called the *payoff space*;

(vi) A function

$h : Z \times Y_1 \times Y_2 \times \cdots \to H$

called the *payoff function*. The payoff function is assumed to be an m-transformation for each fixed $z \in Z$.

Intuitively, the game is played as follows: First the "opponents," including chance, each pick a strategy; the composite of these strategies is a member of z of Z. Next, our player is informed of the value of $g_1(z)$; this is a member of X_1, and represents our player's state of information for his first move. Our player then chooses a member y_1 of Y_1. Next, he is informed of the value of $g_2(z, y_1)$; this is a member of X_2, and on the basis of this he must choose a member y_2 of Y_2. Next, he is informed of the value of $g_3(z, y_1, y_2)$; the game continues in this way. The payoff is determined as a function of the strategy z chosen by the opponents, and the actions y_1, $y_2, \ldots$ taken by our player. Usually it will be most convenient to take the payoff space H to be a Euclidean space of dimension equal to the number of players. However this need not always be so,[10] and since we do not use any particular form for H in the sequel, we have left H as general as possible.

Note that up to the present we have not assumed that our player remembers anything on the occasion of a given choice except what he is told by the value of the function g_i. This can be made plausible if we think of the choices of y_1, $y_2, \ldots$ as being made by distinct agents of our player, who are not allowed to communicate with each other.

The mappings g and h have been assumed to be m-transformations in the variables y_i for the familiar reason, namely to ensure that distributions on the domain spaces induce a distribution on the range space. This has not been required for the variable z in order to avoid the necessity of defining a measurable structure on the strategy space Z, which leads to difficulties, as we have seen. The results should thus be conceived as holding for each z separately. In a particular case it might be possible to integrate over some components of z (e.g., that belonging to chance); this can be done without difficulty after the results have been established for fixed z.

The above definition is a compromise between the normal and extensive forms of a game. The game has been retained in extensive form for our player, but has been normalized for the other players. Even for finite

10. For instance, for some purposes it is convenient to consider the payoff to a supergame as being simply the sequence of payoffs to the individual plays, rather than the average (in some sense) of these payoffs.

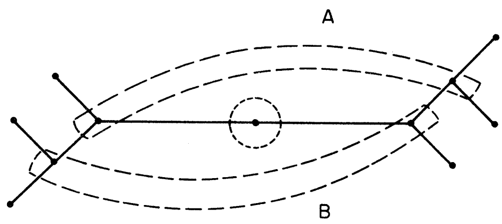

Figure 1

games, the theorem on games of perfect recall is best stated for one player at a time; the process of normalizing the game for the other players enables us to focus attention on the single player and thus simplifies the proofs.

Not all finite extensive games in the sense of Kuhn [K_1] (or in the more general sense of Isbell [I]) are included in the above definition; however all games of perfect recall are included, as are all finite extensive games in the sense of von Neumann and Morgenstern [N–M]. The condition for a Kuhn game to be included is that the game can be "serialized," timewise, for the player in question. For example the game in Figure 1 does not come under our definition if the information sets A and B belong to the same player, but does come under our definition if they belong to distinct players.[11] Of course the possibility of serialization is not at all equivalent with perfect recall (but the latter implies the former).

Most extensive-game models used by authors other than those mentioned above are similar to Kuhn's model, and the same remarks apply.

Next, we define games of perfect recall in our model.

DEFINITION A game is said to be of *perfect recall* if there are sequences of m-transformations

$$u_j^i : X_i \to Y_j, \quad j < i$$

and

$$t_j^i : X_i \to X_j, \quad j < i$$

such that

$$u_j^i g_i(z, y_1, \ldots, y_{i-1}) = y_j$$

and

$$t_j^i g_i(z, y_1, \ldots, y_{i-1}) = g_j(z, y_1, \ldots, y_{j-1}).$$

11. The example is taken from [K_1].

Intuitively, u is the function by which a player remembers what he previously did, and t is the function by which he remembers what he previously knew.

Note that we have given an analytic definition of games of perfect recall which, while retaining complete generality, avoids the cumbersome geometric tree model. This has been made possible by the device of normalizing the game for all but one player.

5 Formal Definitions of Mixed and Behavior Strategies; Kuhn's Theorem

We may assume without loss of generality that the X_i and Y_i are all copies of I;[12] for if one of them is only finite or denumerable, we can always add a continuum of identical copies. The cartesian products $\times_i X_i$ and $\times_i Y_i$ will be denoted $\mathbf{X}$ and $\mathbf{Y}$ respectively,[13] and their members will be denoted $\mathbf{x} = (x_1, x_2, \ldots)$ and $\mathbf{y} = (y_1, y_2, \ldots)$. We remind the reader that the phrase "sample space" means a copy of Ω. Sample spaces will be denoted by Ω, $\boldsymbol{\Omega}$, Ω_1, etc.; the measures on them by λ, $\boldsymbol{\lambda}$, λ_i, etc., respectively.

In this and the following three sections, the word "subset," when applied to an m-space, will always mean "measurable subset," and $B \subset Y$, when Y is an m-space, will mean "B is an m-subset of Y."

DEFINITION A *mixed strategy* is a sequence $\mathbf{m} = (m_1, m_2, \ldots)$ of m-transformations $m_i : \Omega \times X_i \to Y_i$, where Ω is a fixed sample space. A *behavior strategy* is a mixed strategy $\mathbf{b}$, such that for $i \neq j$, $b_i(\cdot, x_i)$ and $b_j(\cdot, x_j)$ are mutually independent random variables ($x_i \in X_i$ and $x_j \in X_j$ being arbitrary).[14]

Every triple $(\omega, \mathbf{m}, z)$ consisting of a member of the sample space, a mixed strategy, and a strategy of the opponents uniquely determines a member $\mathbf{v}(\omega; \mathbf{m}, z)$ of $\mathbf{Y}$; $\mathbf{v} = (v_1, v_2, \ldots)$ is defined recursively by

$$v_i = m_i(\omega, g_i(z, v_1, \ldots, v_{i-1})).$$

Intuitively, v is the sequence of choices that actually occur when the game is played. Furthermore every pair $(\mathbf{m}, z)$ uniquely determines a distribution (i.e., measure) μ on $\mathbf{Y}$; this is defined for $\mathbf{B} \subset \mathbf{Y}$ by

$$\mu(\mathbf{B}) = \mu(\mathbf{B}; \mathbf{m}, z) = \lambda\{\omega : \mathbf{v}(\omega; \mathbf{m}, z) \in \mathbf{B}\}.$$

12. For the definitions of I and Ω see Section 3.

13. The reader should be careful to distinguish between boldface letters $\boldsymbol{X}$, $\boldsymbol{\Omega}$, $\boldsymbol{\lambda}$, etc., and ordinary letters X, Ω, λ, etc.

14. See Section 9 for a discussion of this definition.

Intuitively, μ is the distribution of the random variable $\mathbf{v}(\cdot; \mathbf{m}, z)$. Two mixed strategies are said to be *equivalent* if for each $z \in Z$, they determine the same distribution on $\mathbf{Y}$.

We are now ready to state

KUHN'S THEOREM In a game of perfect recall, every mixed strategy has an equivalent behavior strategy.

6 Lemmas for the Proof of Kuhn's Theorem

Since we will make extensive use of conditional probabilities where the probability of the condition vanishes, we briefly review the properties of such conditional probabilities. Let Ω be a sample space with measure λ, let Y be a copy of I, and let $v : \Omega \to Y$ be an m-transformation. For arbitrary $\Gamma \subset \Omega$ and $y \in Y$, we are interested in the conditional probability of Γ given that $v(\omega) = y$. Note that the condition $v(\omega) = y$ may well have probability zero, that is, we may have $\lambda\{\omega : v(\omega) = y\} = 0$, possibly even for all y. However, it can be proved [H, pp. 206–209] that for each y there is an essentially unique[15] probability measure on Ω, denoted[16] by cond $\lambda(\cdot \mid v(\omega) = y)$, such that for each $B \subset Y$ and $\Gamma \subset \Omega$, we have

$$\int_B \text{cond } \lambda(\Gamma \mid v(\omega) = y)\, d\lambda v^{-1}(y) = \lambda(v^{-1}(B) \cap \Gamma). \tag{A0}$$

Formula (A0) is the analogue of the familiar "partition formula" in elementary probability theory. That cond $\lambda(\cdot \mid v(\omega) = y)$ is a probability measure follows from the fact that Ω is standard [H, p. 210, example 5]; this is the only place where we use standardness.

Let Ω, λ, Y, and v be as above, Y' a copy of I, $g : \Omega \times Y \to Y'$ an m-transformation, $B' \subset Y'$ and $B \subset Y$.

LEMMA A Under the above conditions

$$\int_B \text{cond } \lambda(\{\omega : g(\omega, y) \in B'\} | v(\omega) = y)\, d\lambda v^{-1}(y)$$
$$= \lambda\{\omega : g(\omega, v(\omega)) \in B' \text{ and } v(\omega) \in B\}.$$

15. Actually cond λ is defined uniquely only up to a set of y which is of (λv^{-1})-measure 0. But all our statements will hold for any particular version of cond λ, so the particular choice can be made arbitrarily.

16. We trust that our notation for conditional probabilities, though not standard, is sufficiently transparent as to cause no confusion. There are good reasons for using it rather than one of the standard notations.

Remark The unusual feature of the integral on the left-hand side is that the subset of Ω of which the conditional probability is being taken—the set $\{\omega : g(\omega, y) \in B'\}$—varies with the condition y. If it were not that cond λ is defined essentially uniquely as a probability measure (for example if Ω were not standard), the integral would have no meaning, because cond λ could be assigned an arbitrary value for each y. What the lemma says is that since the condition asserts $v(\omega) = y$, we may substitute $v(\omega)$ for y on the left side of the | sign, and then obtain the correct answer by using (A0).

Proof Let $C \subset \Omega \times Y$ be defined by $C = g^{-1}(B')$. Denoting by C^y the section $\{\omega : (\omega, y) \in C\}$, we obtain that Lemma A is equivalent to

$$\int_B \text{cond}\, \lambda(C^y \,|\, v(\omega) = y)\, d\lambda v^{-1}(y)$$
$$= \lambda\{\omega : (\omega, v(\omega)) \in C \text{ and } v(\omega) \in B\}. \tag{A1}$$

Both sides of (A1), as functions of C, are measures on $\Omega \times Y$ (since cond λ is a measure on Ω for each y). Hence it is sufficient to prove (A1) when C is a rectangle $\Gamma \times A$ in $\Omega \times Y$. In this case the left side of (A1) becomes

$$\int_{A \cap B} \text{cond}\, \lambda(\Gamma \,|\, v(\omega) = y)\, d\lambda v^{-1}(y)$$
$$= \lambda\{\Gamma \cap v^{-1}(A \cap B)\}$$
$$= \lambda(\omega : \omega \in \Gamma \text{ and } v(\omega) \in A \text{ and } v(\omega) \in B\}$$
$$= \lambda\{\omega : (\omega, v(\omega)) \in \Gamma \times A \text{ and } v(\omega) \in B\}$$
$$= \lambda\{\omega : (\omega, v(\omega)) \in C \text{ and } v(\omega) \in B\}.$$

This demonstrates (A1) when C is a rectangle, and (A1) and therefore also Lemma A follows in the general case.

Now let us return to our game. First we introduce some further notation. We write $\mathbf{Y}_i = Y_1 \times \cdots \times Y_i$. Similarly, for $\mathbf{y} \in \mathbf{Y}$, we write $\mathbf{y}_i = (y_1, \ldots, y_i)$. If $B_1 \subset Y_1$, $B_2 \subset Y_2, \ldots,$ then we write $\mathbf{B}_i = B_1 \times \cdots \times B_i$, and $\mathbf{B} = B_1 \times B_2 \times \cdots$. The symbol $\mathbf{B}$ will always be reserved for a rectangle of this kind.

Let us consider a mixed strategy $\mathbf{m}$ with sample space Ω, a strategy z of the opponents, and a sequence $\mathbf{y} \in \mathbf{Y}$. Then for each $i = 1, 2, \ldots$ we may define a sequence $\mathbf{v}^i = (v_1^i, v_2^i, \ldots) = \mathbf{v}^i(\omega, \mathbf{y}; \mathbf{m}, z)$ inductively as follows:

$$v_j^i = \begin{cases} y_j, & \text{for } j < i \\ m_j(\omega, g_j(z, v_{j-1}^i)), & \text{for } j \geqslant i. \end{cases}$$

We have $\mathbf{v}^1 = v$, and $v_i^i = m_i(\omega, g_i(z, \mathbf{y}_{i-1}))$, which is the decision on the i^{th} play if $\mathbf{y}_{i-1}$ has been chosen on the previous plays. Denote $(v_i^i, \ldots, v_k^i)$ by $\hat{v}_k^i$ (for $k \geqslant i$). Note that v_j^i depends only on $\mathbf{y}_{i-1}$ rather than on all of $\mathbf{y}$, so we may write $v_j^i(\omega, \mathbf{y}_{i-1}; \mathbf{m}, z)$ rather than $v_j^i(\omega, \mathbf{y}; \mathbf{m}, z)$. As $\mathbf{m}$ and z will be fixed throughout most of this discussion, we will usually write $v_j^i(\omega, \mathbf{y}_{i-1})$, and omit explicit mention of $\mathbf{m}$ and z. The expression $v_j^i(\cdot, \mathbf{y}_{i-1})^{-1}(B_j)$ means $\{\omega : v_j^i(\omega, \mathbf{y}_{i-1}) \in B_j\}$. For future reference note that

$$v_j^{i+1}(\omega, (\mathbf{y}_{i-1}, v_i^i(\omega, \mathbf{y}_{i-1}))) = v_j^i(\omega, \mathbf{y}_{i-1}). \tag{B1}$$

Next, remembering that $\mathbf{y}$ is fixed, define a sequence λ, $\lambda_{\mathbf{y}_1}$, $\lambda_{\mathbf{y}_2}, \ldots$ of measures on Ω as follows:

$$\lambda_{\mathbf{y}_i}(\Gamma) = \operatorname{cond} \lambda_{\mathbf{y}_{i-1}}(\Gamma \mid v_i^i(\omega, \mathbf{y}_{i-1}) = y_i)$$

(where of course $\lambda_{\mathbf{y}_0}$ stands for λ).

LEMMA B Let $B_i \subset Y_i, \ldots, B_k \subset Y_k$. Then

$$\begin{aligned}\int_{B_i} \cdots \int_{B_k} & d\lambda_{\mathbf{y}_{k-1}} v_k^k(\cdot, \mathbf{y}_{k-1})^{-1}(\mathbf{y}_k) \cdots d\lambda_{\mathbf{y}_{i-1}} v_i^i(\cdot, \mathbf{y}_{i-1})^{-1}(y_i) \\ &= \lambda_{\mathbf{y}_{i-1}} \hat{v}_k^i(\cdot, \mathbf{y}_{i-1})^{-1}(\hat{B}_k^i),\end{aligned}$$

where $\hat{B}_k^i = B_i \times \cdots \times B_k$.

Proof We use reverse induction on i. The start, at $i = k$, is immediate. For the inductive step ($i + 1$ implies i) we have

$$\begin{aligned}\int_{B_i} \int_{B_{i+1}} \cdots \int_{B_k} &= \int_{B_i} \lambda_{\mathbf{y}_i} \hat{v}_k^{i+1}(\cdot, \mathbf{y}_i)^{-1}(\hat{B}_k^{i+1})\, d\lambda_{\mathbf{y}_{i-1}} v_i^i(\cdot, \mathbf{y}_{i-1})^{-1}(y_i) \\ &= \int_{B_i} \operatorname{cond} \lambda_{\mathbf{y}_{i-1}}(\{\omega : \hat{v}_k^{i+1}(\omega, (\mathbf{y}_{i-1}, y_i)) \in \hat{B}_k^{i+1}\} \mid \\ &\qquad v_i^i(\omega, \mathbf{y}_{i-1}) = y_i)\, d\lambda_{\mathbf{y}_{i-1}} v_i^i(\cdot, \mathbf{y}_{i-1})^{-1}(y_i).\end{aligned}$$

Applying Lemma A with $\lambda_{\mathbf{y}_{i-1}}$ instead of λ, B_i instead of B, Y_i instead of Y, $\hat{B}_k^{i+1}$ instead of B', $\hat{v}_k^{i+1}(\cdot, (\mathbf{y}_{i-1}, \cdot))$ instead of g, $\hat{Y}_k^{i+1}$ instead of Y', and $v_i^i(\cdot, \mathbf{y}_{i-1})$ instead of v, we obtain that the last expression above is equal to

$$\lambda_{\mathbf{y}_{i-1}}\{\omega : \hat{v}_k^{i+1}(\omega, (\mathbf{y}_{i-1}, v_i^i(\omega, \mathbf{y}_{i-1}))) \in \hat{B}_k^{i+1} \text{ and } v_i^i(\omega, \mathbf{y}_{i-1}) \in B_i\},$$

and from (B1) we deduce that this is equal to

$$\lambda_{\mathbf{y}_{i-1}}\{\omega : \hat{v}_k^i(\omega, \mathbf{y}_{i-1}) \in \hat{B}_k^i\}.$$

This completes the induction.

COROLLARY C Let $B_1 \subset Y_1, \ldots, B_k \subset Y_k$. Then

$$\int_{B_1} \cdots \int_{B_k} d\lambda_{\mathbf{y}_{k-1}} v_k^k(\cdot, \mathbf{y}_{k-1})^{-1}(y_k) \cdots d\lambda \mathbf{v}_1^{-1}(\mathbf{y}_1) = \lambda v_k^{-1}(\mathbf{B}_k).$$

COROLLARY D Let f be an m-transformation from Y_k to the real numbers. Then

$$\int_{B_1} \cdots \int_{B_k} f(\mathbf{y}_k)\, d\lambda_{\mathbf{y}_{k-1}} v_k^k(\cdot, \mathbf{y}_{k-1})^{-1}(y_k) \cdots d\lambda v_1^{-1}(y_1)$$
$$= \int_{B_k} f(\mathbf{y}_k)\, d\lambda \mathbf{v}_k^{-1}(\mathbf{y}_k).$$

Proof If f is the characteristic function of a rectangular parallelepiped in Y_k, this follows from Corollary C. The general case follows by the usual methods.

7 Further Lemmas

The object of this section is to prove that a family of distributions can be "inverted" to yield a family of random variables; the precise statement is Lemma F below. We shall need the following lemma, which says that a single distribution can be inverted to yield a single random variable.

LEMMA E Let f be a non-decreasing upper semi-continuous function[17] on I such that $f(0) \geqslant 0$ and $f(1) = 1$. For $0 \leqslant y \leqslant 1$ define

$$f^{-1}(y) = \begin{cases} \sup\{x : f(x) \leqslant y\}, & \text{if } \{x : f(x) \leqslant y\} \text{ is non-empty} \\ 0 & \text{if it is empty.} \end{cases}$$

Then

(1) f^{-1} is non-decreasing,

(2) f^{-1} is upper-semi-continuous,

(3) $f^{-1}(0) \geqslant 0, f^{-1}(1) = 1$, and

(4) $(f^{-1})^{-1} = f$.

The proof is straightforward, and will be omitted.

Now let X and Y be copies of I, and let $\mathscr{B}$ be the σ-ring of m-sets in Y. Let $\beta : \mathscr{B} \times X \to \Omega$ be a function which is measurable in X for each fixed $B \in \mathscr{B}$ and a probability in $\mathscr{B}$ for each fixed $x \in X$.

17. I.e., $f(x) = \lim\sup_{y \to x} f(y) = \lim_{y \to x+} f(y)$.

LEMMA F Under the above conditions, there is a family of random variables whose distributions are given by β; more precisely, there is an m-transformation $b : \Omega \times X \to Y$ such that

$$\lambda\{\omega : b(\omega, x) \in B\} = \beta(B, x)$$

for each $x \in X$ and $B \in \mathscr{B}$.

Proof For $y \in Y$, define

$$\pi(x, y) = \beta([0, y], x).$$

Write $\pi_x = \pi(x, \cdot)$; π_x is a non-decreasing upper-semi-continuous function of y, so by Lemma E it has a well-defined inverse, which we denote b_x; set $b(\omega, x) = b_x(\omega)$.

LEMMA F1 $b(\omega, \cdot)$ is Borel measurable in x for each fixed ω.

Proof For $B \in \mathscr{B}$ we must show that $\{x : b(\omega, x) \in B\}$ is measurable in X. It is sufficient to show this when B is of the form $[0, y_0)$. Now

$$\begin{aligned}\{x : b(\omega, x) \in [0, y_0)\} &= \{x : \sup\{y : \pi(x, y) \leqslant \omega\} < y_0\} \\ &= \{x : \exists \text{ rational } r < y_0 \text{ such that } \pi(x, r) > \omega\} \\ &= \bigcup_{r<y_0} \{x : \pi(x, r) > \omega\} \\ &= \bigcup_{r<y_0} \{x : \beta([0, r], x) > \omega\} \\ &= \text{union of Borel sets} = \text{a Borel set.}\end{aligned}$$

This completes the proof of Lemma F1.

Next we show that b is measurable in the two variables simultaneously. It is sufficient to prove that sets of the form $[y_0, 1]$ have measurable inverse images. Indeed,

$$\begin{aligned}b^{-1}[y_0, 1] &= \{(\omega, x) : b(\omega, x) \geqslant y_0\} \\ &= \{(\omega, x) : (\forall \text{ rational } s),\ (s > \omega \Rightarrow b(s, x) \geqslant y_0)\} \\ &\qquad\qquad\text{(because of upper semi-continuity of } b\text{)} \\ &= \bigcap_s \{(\omega, x) : (b(s, x) \geqslant y_0) \text{ or } (s \leqslant \omega)\} \\ &= \bigcap_s (\{(\omega, x) : b(s, x) \geqslant y_0)\} \cup \{(\omega, x) : (s \leqslant \omega)\}) \\ &= \bigcap_s ((\Omega \times \{x : b(s, x) \geqslant y_0\}) \cup ([s, 1] \times X)),\end{aligned}$$

and this is Borel measurable in $\Omega \times X$ (by Lemma F1).

Finally, we show that

$$\lambda\{\omega : b(\omega, x) \in B\} = \beta(B, x).$$

It is sufficient to demonstrate this when B is of the form $[0, y]$. Then

$$\lambda\{\omega : b(\omega, x) \in [0, y]\} = \lambda\{\omega : b(\omega, x) \leqslant y\}$$
$$= \sup\{\omega : b_x(\omega) \leqslant y\} = b_x^{-1}(y).$$

But $b_x = \pi_x^{-1}$; so $b_x^{-1} = \pi_x$ (by Lemma E). Hence

$$\lambda\{\omega : b(\omega, x) \in [0, y]\} = b_x^{-1}(y) = \pi_x(y) = \pi(x, y) = \beta([0, y], x),$$

and the proof of Lemma F is complete.

8 Proof of Kuhn's Theorem

Fix m; we wish to find an equivalent behavior strategy, which we will call b. We first define the *distributions* β_i of the random variables $b_i(\cdot, x)$, and then only the random variables themselves. For $B \subset Y_i$ and $x \in X_i$, define

$$\beta_i(B, x) = \text{cond } \lambda(\{\omega : m_i(\omega, x) \in B\} \,|\, m_{i-1}(\omega, t_{i-1}^i(x)) = u_{i-1}^i(x) \,|\, \cdots$$
$$|\, m_1(\omega, t_1^i(x)) = u_1^i(x)).$$

The expression on the right is to be interpreted as an iterated conditional probability, similar to the definition of $\lambda_{\mathbf{y}_i}$. To underscore the similarity, note that

$$\beta_i(B, g_i(z, \mathbf{y}_{i-1})) = \lambda_{\mathbf{y}_{i-1}} v_i^i(\cdot, \mathbf{y}_{i-1}; \mathbf{m}, z)^{-1}(B). \tag{K1}$$

Let Ω_1, $\Omega_2, \ldots$ be a sequence of copies of Ω. According to Lemma F we can find $b_i' : \Omega_i \times X_i \to Y_i$ so that the $\beta_i(\cdot, x)$ are the distributions of the $b_i'(\cdot, x)$, i.e., so that

$$\lambda_i\{\omega_i : b_i'(\omega, x) \in B\} = \beta_i(B, x). \tag{K2}$$

Let $\boldsymbol{\Omega} = \Omega_1 \times \Omega_2 \times \cdots$, and note that $\boldsymbol{\Omega}$ is a copy of Ω; define a behavior strategy $b_i : \boldsymbol{\Omega} \times X_i \to Y_i$ by $b_i(\boldsymbol{\omega}, X_i) = b_i'(\omega_i, X_i)$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots)$. Let $B_1 \subset Y_1$, $B_2 \subset Y_2, \ldots$; for each n write $\mathbf{B}_n^* = \mathbf{B}_n \times Y_{n+1} \times \cdots$. To show that $\mathbf{m}$ and $\mathbf{b}$ are equivalent, it is only necessary to show that

$$\mu(\mathbf{B}_n^*; \mathbf{m}, z) = \mu(\mathbf{B}_n^*; \mathbf{b}, z) \tag{K3}$$

for every $z \in Z$ and every n and arbitrary choice of the B_i.

Let us write $\boldsymbol{\omega}_n$ for $(\omega_1, \ldots, \omega_n)$. We first note that $\mathbf{v}_n(\boldsymbol{\omega}; \mathbf{b}, z)$ depends only on ω_n rather than on all of ω. In fact, if we define $\mathbf{w}_n(\boldsymbol{\omega}_n)$ recursively by

$$w_n(\omega_n) = b'_n(\omega_n, g_n(z, \mathbf{w}_{n-1}(\omega_{n-1}))),$$

and $\mathbf{w}_n = (w_1, \ldots, w_n)$, then

$$w_n(\boldsymbol{\omega}_n) = v_n(\boldsymbol{\omega}; \mathbf{b}, z). \tag{K4}$$

Henceforth we will use $\mathbf{v}$ for $\mathbf{m}$ exclusively (unless we explicitly indicate otherwise); thus $\mathbf{v}(\omega)$ will mean $\mathbf{v}(\omega; \mathbf{m}, z)$, and similarly for v^i_j, etc.

The proof of (K3) is by induction on n; the induction is easily started (at $n = 1$). For the inductive step (n implies $n + 1$) note that

$$\boldsymbol{\lambda}_n \mathbf{w}_n^{-1}(\mathbf{B}_n) = \mu(\mathbf{B}^*_n; \mathbf{b}, z)$$

because of (K4), where $\boldsymbol{\lambda}_n = \lambda_1 \times \cdots \times \lambda_n$; furthermore

$$\lambda \mathbf{v}_n^{-1}(\mathbf{B}_n) = \mu(\mathbf{B}^*_n; \mathbf{m}, z).$$

Since by induction hypothesis the two right sides are equal, it follows that the left sides also are; but since this holds for all $\mathbf{B}_n \subset \mathbf{Y}_n$, it follows that

$$\boldsymbol{\lambda}_n \mathbf{w}_n^{-1} = \boldsymbol{\lambda} v_n^{-1} \tag{K5}$$

as measures on $\mathbf{Y}_n$. Next, we have

$$\begin{aligned}
\mu(\mathbf{B}^*_{n+1}; \mathbf{b}, z) &= \boldsymbol{\lambda}\{\boldsymbol{\omega} : w_{n+1}(\boldsymbol{\omega}_{n+1}) \in \mathbf{B}_{n+1}\} \\
&= \boldsymbol{\lambda}_{n+1}\{\boldsymbol{\omega}_{n+1} : w_{n+1}(\boldsymbol{\omega}_{n+1}) \in B_{n+1} \text{ and } \mathbf{w}_n(\boldsymbol{\omega}_n) \in \mathbf{B}_n\} \\
&= \boldsymbol{\lambda}_{n+1}\{\boldsymbol{\omega}_{n+1} : b'_{n+1}(\omega_{n+1}, g_{n+1}(z, \mathbf{w}_n(\boldsymbol{\omega}_n))) \in B_{n+1} \text{ and } \mathbf{w}_n(\boldsymbol{\omega}_n) \in \mathbf{B}_n\} \\
&= \int_{\mathbf{w}_n^{-1}(\mathbf{B}_n)} \lambda_{n+1}\{\omega_{n+1} : b'_{n+1}(\omega_{n+1}, g_{n+1}(z, \mathbf{w}_n(\boldsymbol{\omega}_n))) \in B_{n+1}\}\, d\boldsymbol{\lambda}_n(\boldsymbol{\omega}_n) \\
&= \int_{\mathbf{B}_n} \beta_{n+1}(B_{n+1}, g_{n+1}(z, \mathbf{y}_n))\, d\boldsymbol{\lambda}_n \mathbf{w}_n^{-1}(\mathbf{y}_n),
\end{aligned}$$

(because of (K2) and the change of variables $\mathbf{y}_n = \mathbf{w}_n(\boldsymbol{\omega}_n)$)

$$= \int_{\mathbf{B}_n} \boldsymbol{\lambda}_{\mathbf{y}_n} v^{n+1}_{n+1}(\cdot, \mathbf{y}_n)^{-1}(B_{n+1})\, d\lambda \mathbf{v}_n^{-1}(\mathbf{y}_n)$$

(because of (K1) and (K5))

$$= \int_{B_1} \cdots \int_{B_n} \lambda_{\mathbf{y}_n} v_{n+1}^{n+1}(\cdot, \mathbf{y}_n)^{-1}(B_{n+1})\, d\lambda_{\mathbf{y}_{n-1}} v_n^n(\cdot, \mathbf{y}_{n-1})^{-1}(y_n)$$

$$\cdots d\lambda v_1^{-1}(y_1) \quad \text{(because of Corollary D)}$$

$$= \int_{B_1} \cdots \int_{B_n} \int_{B_{n+1}} d\lambda_{\mathbf{y}_n} v_{n+1}^{n+1}(\cdot, \mathbf{y}_n)^{-1}(y_{n+1}) \cdots d\lambda v_1^{-1}(y_1)$$

$$= \lambda \mathbf{v}_{n+1}^{-1}(\mathbf{B}_{n+1})$$

$$= \mu(B_{n+1}^*; \mathbf{m}, z). \quad \text{(because of Corollary C)}$$

This completes the proof of Kuhn's theorem.

9 Remarks on the Definition of Behavior Strategy

Intuitively, a player using a behavior strategy "postpones crossing his bridges until he gets to them;" that is, he randomizes independently on each occasion of a choice, rather than letting his choices be governed by a single randomization performed before the start of play. Thus a behavior strategy is a family of independent random procedures for choosing an action, one for each possible information state. The direct translation of this intuitive concept into technical language yields a sequence of functions $b_i : \Omega \times X_i \to Y_i$, where the functions $b_i(\cdot, x)$ are mutually independent random variables for distinct x, even when we are dealing with a single i. This is a little different from the definition of Section 5, in which it was demanded only that the $b_i(\cdot, x)$ be independent for distinct i; that is, the player randomizes independently at each stage, but before the information for that stage is received, rather than afterwards.

To demand that the $b_i(\cdot, x)$ be independent for distinct x—even for a fixed i—would mean that we must have a nondenumerable number of mutually independent bounded random variables on the same sample space; and except for the degenerate case in which almost all of these are constants, this is in fact impossible (when the phrase "sample space" is used in our restricted sense, which corresponds to the intuitive idea of random device). Indeed, suppose $\{b_x\}$ is a nondenumerable family of bounded nonconstant mutually independent random variables on the sample space Ω, and let $c_x = b_x - \text{mean}\,(b_x)$. Then from independence it follows that the c_x are mutually orthogonal in the Hilbert space $L^2(\Omega)$, and from the fact that the b_x are nonconstant it follows that the c_x do not vanish identically. So we have a nondenumerable number of nonzero mutually orthogonal members of $L^2(\Omega)$, and hence $L^2(\Omega)$ has nondenumerable dimension. But its dimension is known to be denumerable, so our contention is proved.

It may seem that this makes any genuine analogue of Kuhn's theorem in the continuous case impossible. The difference, however, is illusory, and there is no real loss of strength in our theorem. We have seen that the $b_i(\cdot, x)$ must necessarily be correlated as x ranges over X_i, simply because of the cardinality of X_i. However, this correlation is entirely irrelevant to the game, and cannot affect the payoff in any way. In fact, *the payoff distribution depends only on the distributions of the individual* $b_i(\cdot, x)$, *and not on any of the joint distributions* (this follows from Section 8). In other words, the $b_i(\cdot, x)$ are correlated (for fixed i and varying x) not because this correlation is necessary to mimic the effect of the given mixed strategy m; in fact, if this were necessary (as it may be when the game is not of perfect recall), this kind of in-stage correlation could not accomplish it,[18] and we would have to resort to interstage correlation. The correlation is rather in the way of being an irrelevant mathematical accident.

Yet another way of saying this is that as long as they have the proper distributions, the $b_i(\cdot, x)$ can be chosen in any way we please, without any regard to each other, except that in the end b_i must be simultaneously measurable in both ω and x. Though the $b_i(\cdot, x)$ must be correlated, what form the correlation takes is of no concern to us.

Finally, we remark that it would have been possible to define behavior strategies as functions that take members of X_i into distributions over Y_i, in a manner directly analogous to that of [K_1]. With that approach, the independence assumptions would have been implicit in the formula for the payoff to a behavior strategy, and the question under discussion would not have arisen at all. We prefer our approach because it underscores the natural relation between mixed and behavior strategies.

Bibliography

[A_1] Aumann, R. J., "Acceptable points in general cooperative *n*-person games," *Annals of Mathematics Study* 40 (Princeton, 1959) pp. 287–324 [Chapter 20].

[A_2] Aumann, R. J., "The core of a cooperative game without side payments," *Trans. Amer. Math. Soc.* 98 (1961), pp. 539–552 [Chapter 39].

[A_3] Aumann, R. J., "Borel structures for function spaces," *Illinois J. of Math.* 5 (1961), pp. 614–630 [Chapter 66].

[A_4] Aumann, R. J., "On choosing a function at random," *Ergodic Theory*, Academic Press, 1963, pp. 1–20 [Chapter 67].

[F] Fox, M., "Some zero sum two-person games with moves in the unit interval," *Pac. J. Math.* 10 (1960), pp. 1235–1242.

[H] Halmos, P. R., *Measure Theory*, Van Nostrand, 1950.

[I] Isbell, J. R., "Finitary games," *Annals of Mathematics Study* 39 (Princeton, 1957), pp. 79–96.

18. Cf. the italicized statement above.

[K_1] Kuhn, H. W., "Extensive games and the problem of information," *Annals of Mathematics Study* 28 (Princeton, 1953), pp. 193–216.

[K_2] Kuhn, H. W., Lectures on the Theory of Games, Princeton University Lecture Notes, 1952 (mimeographed).

[M] Mackey, G. W., "Borel structures in groups and their duals," *Trans. Amer. Math. Soc.* 85 (1957), pp. 283–295.

[Mc] McKinsey, J.C.C., *Introduction to the Theory of Games*, McGraw-Hill, 1952.

[N–M] Von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton University Press, 1944, Third Edition, 1953.

[S] Sierpinski, W., *Hypothèse du continu*, Chelsea, 1956.

[V] Ville, J., "Sur la théorie générale des jeux où intervient l'habilité des joueurs," note in *Traité du Calcul des Probabilites et de ses Applications* by E. Borel, Tome IV, Fascicule II, "Applications aux jeux de hasard," (1938), pp. 105–113.

[W] Wolfe, P., editor, Report of an informal conference on recent developments in the theory of games, Princeton University, 1955 (mimeographed).