ESSAYS ON CHOICES AND BELIEFS
IN UNCERTAINTY AND GAME THEORY

A thesis presented

by

Eddie Dekel

to

The Department of Economics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Economics

Harvard University
Cambridge, Massachusetts
May 1986

ABSTRACT

In this thesis I examine a number of issues in single- and multi-person decision theory. In Chapter 1 an implicit expected utility representation of preferences is characterized by replacing the controversial independence axiom with the betweenness axiom. Many results and analytical tools from expected utility theory can be extended to the representation derived, which is, moreover, consistent with empirical evidence which violates expected utility maximization. Chapter 2 focuses on the relationship between portfolio diversification and risk aversion for general (non-linear) preferences over probability distributions.

In Chapters 3 and 4 the concept of common knowledge is examined. First an equivalence between a Bayesian definition (in terms of beliefs) and an informational definition (in terms of σ-fields) is proven in a general framework (which allows for null events in the join of the information partitions). In Chapter 4 the approach of hierarchies of beliefs is developed. A *type* of an individual is an infinite hierarchy of beliefs -- over some space S, over others' beliefs over S, and so on. I show that a coherent type determines a belief over S and other individuals' types, and that common knowledge of coherency is needed to "close" this model. In Chapter 3, the information structure (probabilities and σ-fields) is assumed to be common knowledge. In Chapter 4 this assumption is shown to be without loss of generality if the underlying space is expanded to include the type spaces and common

knowledge of coherency is satisfied.

Chapter 5 shows how common knowledge restrictions on beliefs can be related to equilibrium concepts in games. I prove that the set of correlated rationalizable payoff vectors in a game is equal to the set of payoff vectors from the a posteriori equilibria (a special type of subjective correlated equilibria) of that game, and similarly independent rationalizability is equivalent to mixed a posteriori equilibria.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

It is always problematic to make claims about the implications of a counterfactual assumption, however I can definitely say that the path of my studies leading to this thesis would have been much more difficult, if not unpassable, without the support of my advisors Jerry Green and Andreu Mas-Colell. Working as a research and teaching assistant for both was an enjoyable and beneficial experience. Some parts of this dissertation were motivated during a summer as a research assistant for Jerry Green, while the seeds for other parts were planted in his course on Uncertainty and Information. Moreover, he was always interested in discussing a wide variety of questions and supported the different directions of research they led to. Andreu Mas-Colell was also extremely accessible and interested in talking about any idea, at all stages of its development. My conversations with both were invaluable. In addition to my advisors I benefitted from my contact with Mark Machina and David Kreps. Mark suffered through lengthy discussions of ideas and problems over long distance phone lines -- these talks were helpful and enjoyable. I also greatly appreciate having had feedback from David Kreps.

I cannot express how important and valuable are my friendship and intellectual partnership with Adam Brandenburger. We have spent countless hours together discussing and working on ideas in decision theory and game theory.

I would also like to thank Bonnie Currier. Her willingness to help whenever necessary, and her technical support throughout my studies is greatly appreciated.

Financial support from the Sloan Foundation for this year, and the Chiles Fellowship last year, and partial support from NSF grants SES-8207338 and IST-8310118 are gratefully acknowledged.

My parents and brother have always supported me -- to them I dedicate this work. There are no words to say how important my wife, Ayelet Tabak, is for my work -- her comments from a perspective quite different from my own, have lead me to new thoughts on many issues.

INTRODUCTION

The essays in this thesis examine the implications that certain assumptions about the beliefs and behavior of individuals have on the choices these individuals make when faced with decisions under uncertainty. Different frameworks for analyzing choice under uncertainty, and several different questions are studied. Chapters 1 and 2 focus on single person decision theory when the probabilities are objectively given, in particular preferences over lotteries and assets. Chapters 3 and 4 examine the concept of common knowledge, fundamental in the area of multi-person decision theory. The standard framework of a given state space and information structure is used in Chapter 3, while Chapter 4 develops and applies the approach of hierarchies of beliefs. The relationship between restrictions on beliefs which are common knowledge and solution concepts in game theory is pursued in Chapter 5.

The axiomatic approach has been used extensively in the area of single person decision theory -- situations where the outcome of a person's action is influenced by "nature" (the "state of nature" and action jointly determine the outcome). The seminal result in this area is the theory of expected utility. Roughly speaking, continuity and order axioms which imply the existence of a preference function, are combined with the independence axiom to derive the linearity (in probabilities) of this function. The independence axiom is the most controversial of these axioms, in particular because of experiments in which behavior violating expected utility maximization is observed (eg. Allais and Hagen [1]). Chapter 1 proposes an alternative set of axioms where the independence axiom is replaced by the (weaker) betweenness axiom. The betweenness axiom is normatively appealing -- it only requires that receiving the lottery P if a coin comes up heads and lottery Q if it is tails is ranked (in terms of the preference ordering) between P and Q. The behavior implied by the axioms of Chapter 1 has many of the advantages of expected utility maximization. Preferences are represented by the solution of an implicit equation which bears a formal resemblance to an expected utility calculation, and a large class of results and methods of analysis from the theory of expected utility can be extended to the proposed framework (cf. Machina [10]). The experimental evidence mentioned above however is consistent with the axioms I use, and other reasons for choosing them are discussed. In Chapter 2 I examine the optimal choice of assets when the underlying preferences over probability measures are assumed to be representable by a continuous and increasing (in the sense of first order stochastic dominance) function on the space of distributions. Restrictions on these preferences are related to the quasiconcavity of the induced preferences over assets. The latter property is of interest since it implies diversification and convex valued asset demands. I show that risk aversion is necessary but not sufficient for this diversification property. Risk aversion together with quasiconcavity of the underlying preferences is sufficient, however the latter is not necessary and I derive a class of examples to clarify its role.

The area of multi-person decision theory (game theory) examines the problem of choices when individuals interact, in particular when the choice of one person affects the payoff to another and conversely. The axiomatic approach is much less developed in this area of research, and Chapters 3 - 5, which are based on Brandenburger and Dekel [6, 7, 8], are contributions in this direction. In Chapters 3 and 4 I look at the concept of common knowledge. The idea of common knowledge is central in game theory. For example, the analysis of a game starts by assuming that its structure is common knowledge. Intuitively speaking, two people 1 and 2 are said to have common knowledge of an event if both know it, 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on. Common knowledge was first given a formal definition in [3]. An important restriction on the information structure in [3] is that the join of 1 and 2's partition is assumed to consist of nonnull events. In Chapter 3 I provide a definition of common knowledge in the more general situation, which arises naturally in many decision problems (e.g. with infinite state spaces) where null events are permitted. The main result is an equivalence between a definition in terms of beliefs (where i knows A at ω means that i assigns A posterior probability 1 at ω) and a definition in terms of the σ-fields representing the information (where an individual knows an event if (s)he is informed that it -- or an event which implies it -- has occurred). The equivalence requires the conditional probabilities to be proper, which is an assumption that individuals correctly process their information, and the σ-fields to be completed in an appropriate manner, which is an assumption that the individual is (trivially) informed of events about which (s)he has no doubts -- that is regardless of the actual state of nature, (s)he believes with certainty that the event either occurred or did not occur.

The framework used in Chapter 3 is standard, there is a state space Ω and an information structure -- a σ-field and prior for each individual -- which for the purpose of interpretation are taken to be common knowledge. An alternative approach would start by noting that the individuals face some common space of uncertainty S. In order to determine his/her optimal decision, each individual must have a belief (probability measure) over S. But if other individuals' beliefs affect his/her decision, then (s)he must have a belief over others' beliefs over S. This leads to each person having an infinite hierarchy of beliefs (over S, over others' beliefs over S and so on) which is called his/her type (see also Boge and Eisele [5], Harsanyi [9] and Mertens and Zamir [11]). In Chapter 4 I develop this second approach and show how it relates to the standard framework. The notions of belief-closed sets and coherency are examined. Then I show that a definition of common knowledge of an event in terms of the hierarchy is equivalent to the definitions in Chapter 3 (under the assumption of common knowledge of coherency). So, assuming common knowledge of coherency, the two approaches are equivalent. In fact, as noted in [3], when the information structure is not common knowledge we must expand the state space of the framework used in Chapter 3, since otherwise the interpretations of the results are invalid. Chapter 4 shows that the expanded state space needed is precisely the product of the underlying state space with the type spaces restricted to satisfy common knowledge of coherency. This argument implies that assuming common knowledge of the information structure is equivalent to common knowledge of coherency.

Chapter 5 examines the problem of how individuals will play a game. The focus is on the relationship between behavioral assumptions and solution concepts. The fundamental solution concept for noncooperative games is that of a Nash equilibrium [12]. Aumann [2] proposed the idea of objective and subjective correlated equilibrium as an extension of Nash equilibrium to allow for correlation between the players' randomizations and for subjectivity in the players probability assessments. Chapter 5 starts with the solution concept of rationalizability (suggested by Bernheim [4] and Pearce [13]), since this is what is implied by the basic decision theoretic analysis of a game. However it is more closely related to an equilibrium approach than one might think. Rationalizable payoffs and payoffs from a posteriori equilibria -- a refinement of subjective correlated equilibria -- are proven equivalent. The analysis is developed for correlated and independent rationalizability. In proving the equivalence of the latter with mixed a posteriori equilibria the issue of how to correctly define "mixed" strategies in this framework is examined. Slight variations on the definitions, which differ only in how players update on null events, lead to quite different solution concepts.

References

[1] M. Allais and O. Hagen, Eds: *Expected Utility Hypotheses and the Allais Paradox.* Dordrecht, Holland, D. Reidel Publishing Co. 1979.

[2] R. Aumann: "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics,* Vol. 1, (1974), 67-96.

[3] R. Aumann: "Agreeing to Disagree," *The Annals of Statistics,* Vol. 4 (1976), 1236-1239.

[4] D. Bernheim: "Rationalizable Strategic Behavior," *Econometrica,* Vol. 52 (1984), 1007-1028.

[5] W. Boge and Th. Eisele: "On Solutions of Bayesian Games," *International Journal of Game Theory,* Vol. 8 (1979), 193-215.

[6] A. Brandenburger and E. Dekel: "Common Knowledge with Probability 1," Research Paper No. 796R, Graduate School of Business, Stanford University, Sept. 1985.

[7] A. Brandenburger and E. Dekel: "Hierarchies of Beliefs and Common Knowledge," Research Paper No. 841, Graduate School of Business, Stanford University, Sept. 1985.

[8] A. Brandenbuger and E. Dekel: "Rationalizability and Correlated Equilibria," Mimeograph, Dept. of Economics, Harvard University, Aug. 1985.

[9] J. Harsanyi: "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science,* Vol. 14 (1967-68), 159-189, 320-334, 486-502.

[10] M.J. Machina: "'Expected Utility' Analysis without the Independence Axiom," *Econometrica,* Vol. 50 (1982) 277-323.

[11] J.-F. Mertens and S. Zamir: "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory,* Vol. 14 (1985), 1-29.

[12] J. Nash: "Non-Cooperative Games," *Annals of Mathematics,* Vol. 54 (1951), 286-295.

[13] D. Pearce: "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica,* Vol. 52 (1984), 1029-1050.

CHAPTER 1

IMPLICIT EXPECTED UTILITY PREFERENCES: WEAKENING THE INDEPENDENCE AXIOM

1. Introduction

This chapter provides an implicit representation for an axiomatic characterization of preferences under uncertainty. Essentially only the controversial independence axiom is changed to the substantially weaker betweenness axiom (Chew [2]), keeping ordering, monotonicity and continuity type axioms. The betweenness axiom requires that indifference sets be convex, i.e. if an individual is indifferent between two lotteries, then any probability mixture of these two is equally good. This characterization is of interest for a number of reasons. The betweenness axiom is appealing from a normative viewpoint but is compatible with behavior which is not permitted in expected utility, such as the Allais paradox. It also provides a useful behavioral approach since it is the weakest form under which preferences are both quasiconcave and quasiconvex. Quasiconcavity is necessary for the proof of existence of a Nash equilibrium since if preferences are strictly quasiconvex anywhere then a mixed strategy is worse than one of the pure strategies used with positive probability in that mixed strategy. Furthermore, quasiconcavity together with risk aversion are sufficient conditions for continuity of asset demands (while risk aversion alone is not sufficient, see Chapter 2). Quasiconvexity on the other hand is necessary and sufficient for dynamic consistency of choices under uncertainty (see Green [10]). Thus, in order to guarantee the existence of a Nash equilibrium, dynamic consistency and continuous asset demands, we may want to impose quasiconcavity and quasiconvexity of preferences, giving betweenness -- without making the additional restrictions necessary for expected utility.

The chapter begins by presenting the axioms and characterization, discussing recent literature, and proving the representation. Two approaches are taken, one with a weak continuity axiom provides a representation for all simple probability measures (those whose support is a finite subset of the outcome set). The second imposes a stronger form of continuity which suffices both to extend the results to the set of all distributions and also implies that the utility function in the representation is continuous. Then an example is constructed to show that preferences may satisfy the axioms yet not have any differentiable preference functional, even when preferences are over the simplex (and trivially continuous). This implies that the generalization of local monotonicity and risk aversion to global conclusions as proven in Machina [12] might not hold for all preferences of the type discussed here. However, an alternative and intuitive extension of local properties is demonstrated by examining the slopes of the indifference hyperplanes.

2. Axiomatic Characterization

There is an underlying compact metric space W which is the space of outcomes of lotteries, representing for example monetary outcomes or commodity bundles. Preferences, $\succsim$, are defined on the space of all probability measures (D) or simple probability measures (D_s) on the Borel field of W. Convex subsets of D and D_s could also be dealt with, the details are not presented. From these preferences define the induced strict preference, $\succ$, and indifference, $\sim$, relations. Preferences over D and D_s also induce preferences over W, where for any $w, w' \in W$, w is preferred to w' if the measure assigning probability 1 to w is preferred to the measure assigning probability 1 to w'. These measures will be denoted w, w' and this preference relation is written as $\succsim$ where no confusion should result, and the context will clarify whether $w \in W$ (the outcome) or $w \in D$ (the degenerate measure) is implied. For any w, w' in W the measure which assigns probability α to w and $1-\alpha$ to w' is denoted $(\alpha, w; (1-\alpha), w')$.

The following axioms will be used (where P, Q, R are measures in D and $\bar{w}$, $\underline{w}$, w' and w'' are outcomes in W).

A1: (a) $\succ$ is a weak order ($\succsim$ is complete, $\succ$ is asymmetric, and both $\succ$ and $\sim$ are transitive).

(b) There exist best and worst elements in D_s which are the sure outcomes denoted by $\bar{w}$ and $\underline{w}$. (These are not necessarily unique.)

A2: Solvability: If $P \succ Q \succ R$, then there exists an $\alpha \in (0,1)$ such that $\alpha P + (1-\alpha)R \sim Q$.

A3: Monotonicity: If $w = \underline{w}$ or $w = \bar{w}$ and $w' \succ w''$ (resp. $w' \sim w''$) then $(\alpha, w'; (1-\alpha), w) \succ (\alpha, w''; (1-\alpha), w)$ for every $\alpha \in (0,1)$ (resp. $(\alpha, w'; (1-\alpha), w) \sim (\alpha, w''; (1-\alpha), w)$ for every $\alpha \in [0,1]$).

A4: Betweenness: If $P \succ Q$ (resp. $P \sim Q$), then $P \succ \alpha P + (1-\alpha)Q \succ Q$ for every $\alpha \in (0,1)$ (resp. $P \sim \alpha P + (1-\alpha)Q$ for every $\alpha \in [0,1]$).

PROPOSITION 1: Preferences over D_s satisfy A1 - A4 if and only if there exists a function u: $W \times [0,1] \rightarrow \mathbb{R}$ increasing in the preference ordering on W, and continuous in the second argument such that $P \succ Q$ (resp. $P \sim Q$) <=> $V[P] > V[Q]$ (resp. $V[P] = V[Q]$), where $V[F]$ is defined implicitly as the unique $v \in [0,1]$ that solves:

(*) $\int u(w, v)dF(w) = vu(\bar{w}, v) + (1-v)u(\underline{w}, v)$.

Furthermore $u(w, v)$ is unique up to positive affine transformations which are continuous functions of v. A particular transformation exists setting $u(\underline{w}, v) = 0$ and $u(\bar{w}, v) = 1$ for every v, giving the simpler representation:

(**) $\int u(w, V[F])dF(w) = V[F]$.

The uniqueness characterization of u in Proposition 1 is a natural extension of the result in expected utility theory that the Bernoulli utility function is unique up to affine transformations to the framework developed in this chapter. To clarify this generalization let $V[F]$ be uniquely defined from

u by (*) and let $\hat{V}[F]$ be uniquely defined by (*) when $\hat{u}$ replaces u. I say that u is unique up to positive affine transformations which are continuous functions of v when $V[F]$ and $\hat{V}[F]$ represent the same preferences if and only if $\hat{u}(w, v) = a(v)u(w, v) + b(v)$ for some $a(v)$ positive continuous and $b(v)$ continuous.

Monotonicity (A3) is a weaker axiom than the standard first order stochastic dominance axioms (cf. [2, Property 3]). However, as will be seen in Section 4, A1 - A4 are sufficient to prove that the preferences are first order stochastic dominance preserving. In the appendix I provide a characterization which does not assume A3. This characterization is similar to Proposition 1, except that $u(w, v)$ is not necessarily increasing in w (see Section 3.A).

The characterization in Proposition 1 is an implicit expected utility representation, and the similarity of equation (**) to an expected utility calculation suggests that results from the theory of expected utility can be extended to the framework of this chapter. A general result along these lines, based on Epstein's observation [15] that many properties of an optimal choice depend on the indifference curve through that choice and not on the whole indifference map, can be derived. Let U be the set of real valued functions on W, $\bar{U}$ a subset of U and $\bar{D}$ a subset of D. Consider any proposition in expected utility theory of the following form: if $u(\cdot) \in \bar{U}$ then the distribution $F \in D$ which maximizes $\int u(w)dF(w)$ is in $\bar{D}$. For example, if u is concave then F is not second order stochastically dominated, and if u also has positive third derivative then the optimal F isn't third order stochastically dominated. This proposition can be extended to implicit expected utility preferences as follows: if $u(\cdot, v) \in \bar{U}$ for every v then the F which maximizes (**) is in $\bar{D}$. This claim follows from Corollary 1 in [15]. So if $u(\cdot, v)$ has negative second derivative and positive third derivative with respect to the first argument for every v, then the optimal F is not third order stochastically dominated.

Proposition 1 is related to recent axiomatic work in non-linear utility theory, in particular Chew [2], and Fishburn [6, 7, 8]. There are two distinct approaches in this research, depending on whether transitivity of preferences is assumed [2, 7] or not [6, 8]. It is common in both cases to use a type of symmetry axiom which imposes restrictions on how indifference sets relate to one another, while the betweenness axiom imposes convexity on each indifference set (see the indifference sets in the probability simplices in Figure 1). Of course, the additional restriction provides stronger results, essentially guaranteeing the skew-symmetry of a bilinear function $\phi: D \times D \rightarrow \mathbb{R}$ which represents preferences by $\phi(P, Q) > 0$ if and only if $P \succ Q$ [6]. With transitivity ϕ can be decomposed [7] and a weighted expected utility decomposition has been analyzed [2, 3].

The results closest to my work are those of Chew [3] and Fishburn [7]. In [3] Chew has independently provided an implicit weighted utility characterization of preferences satisfying weak order, continuity and substitution axioms which, taken together, are equivalent to A1(a), A2 and A4. Preferences are represented by the solution of an implicit equation which has the form of a weighted utility function (cf. [2]) rather than the implicit expected utility structure in Proposition 1. Fishburn also drops A1(b) and compactness of W, assuming instead countable boundedness (there exists a countable subset $\bar{D}$ of D such that for every $P \in D$ there is $Q, Q' \in \bar{D}$ with $Q \succsim P \succsim Q'$). Other than this his axioms are equivalent to A1, A2 and A4 (Continuity in [7] is A2 and Dominance is A4) giving [7, Theorem 1]:

Countable boundedness, A1(a), A2 and A4 hold iff there exists a function $f: D \to \mathbb{R}$ s.t. $P \succ Q$ iff $f(P) > f(Q)$ and $f(\alpha P + (1-\alpha)Q)$ is continuous and increasing (constant) in α if $P \succ Q$ ($P \sim Q$).

The representation in this chapter is a more refined functional form, closer in structure to expected utility, admits a simple analysis of risk aversion and dominance and has a simple uniqueness characterization.

Proposition 1 bears a formal resemblence to Fishburn's implicit characterization of a certainty equivalent functional $m: D \to \mathbb{R}$ [8]. $m(\cdot)$ is defined from $\int \phi(x, m(P))dP(x) = 0$ where ϕ is a skew symmetric, monotone function and W is an interval of the real line. However the cancellation axiom in [8] is of the symmetry class, thus ϕ is skew symmetric while u may not be. Note that when W is restricted to a compact interval of $\mathbb{R}$, I can use Proposition 1 to provide a mean value representation as follows. Given $u(w, v)$, normalized so that $u(\overline{w}, \cdot) = 1$, $u(\underline{w}, \cdot) = 0$, define $p(w)$ as the unique p which satisfies $w \sim (p, \underline{w}; (1-p), \overline{w})$ and define $c(w, w') = u(w, p(w')) - p(w')$. The certainty equivalent $M[F] \equiv \{w \in D \mid w \sim F\}$ satisfies $u(M[F], V[F]) = V[F]$ by (*) so we have $\int c(w, M[F]) dF(w) = \int \{u(w, V[F]) - V[F]\} dF(w) = 0$. This shows how a generalized mean value without symmetry axioms can be derived using the approach of this chapter. (Note that c may or may not be skew symmetric depending on whether or not the cancellation axiom is satisfied.)

Before going through the constructive proof, it is worthwhile to consider the intuition of the representation. A4 implies that indifference sets are convex. Since thick indifference sets are ruled out (by A4), we are left with indifference sets as hyperplanes. Recall that preferences of the expected utility type have parallel hyperplanes for indifference sets. Imagine now that given the indifference hyperplane, say $H(v)$, through the lottery $(v, \overline{w}; (1-v), \underline{w})$, we ignore all the other indifference sets and construct instead a collection of parallel hyperplanes. These can be taken to represent preferences satisfying the expected utility hypothesis and therefore there exists a function u_v (the subscript indicating the original hyperplane $H(v)$) which satisfies $\int u_v(w) dF(w)$ equals the expected utility evaluation of F. If we set, as we are free to do with expected utility preferences, $u_v(\overline{w}) = 1$ and $u_v(\underline{w}) = 0$ then for $F = (v, \overline{w}; (1-v), \underline{w})$ we have $u_v(\overline{w})v + u_v(\underline{w})(1-v) = v$. Thus for any $F' \in H(v)$, which is an indifference set both for the original preferences and these artificial expected utility preferences, we know that $\int u_v(\cdot)dF'(\cdot) = v$. Doing this for indifference hyperplanes through points $(v, \bar{w}; (1-v), \underline{w})$ for every $v \in (0,1)$ we get a collection of functions $u_v(w)$ which is exactly $u(w, v)$. The intuition of examining the expected utility extension of a given indifference hyperplane lies behind most of the subsequent results. A number of the proofs are done using the characterization (**). This is not restrictive and is only a choice of normalization.

PROOF OF PROPOSITION 1: I will choose a normalization and prove the existence of a representation such as (**), and the uniqueness result will extend this to representations of the form (*). For any p and w with $w \neq \underline{w}$, $w \neq \bar{w}$ and $p \in (0,1)$ the lottery $(p, \bar{w}; (1-p), \underline{w})$ is either: (i) strictly preferred to w; (ii) strictly worse than w; or (iii) indifferent to w. By solvability find: (i) a $\beta \in (0,1)$ s.t. $(\beta, \bar{w}; (1-\beta), w) \sim (p, \bar{w}; (1-p), \underline{w})$, or (ii) a $\gamma \in (0,1)$ s.t. $(\gamma, \underline{w}; (1-\gamma), w) \sim (p, \bar{w}; (1-p), \underline{w})$. In case (i) set $u(w, p) = (p-\beta)/(1-\beta)$, in case (ii) $u(w, p) = p/(1-\gamma)$, and in case (iii) $u(w, p) = p$. If $w = \bar{w}$ or $\underline{w}$ then for every p set $u(\bar{w}, p) = 1$, and $u(\underline{w}, p) = 0$. Since $u(w, v)$ will be shown to be continuous on the open interval $(0,1)$, extend the definition of $u(w, v)$ to the closed interval by continuity. Diagramatically (see Figure 2) what has been done is: (a) construct the intersection of the indifference set through $(p, \bar{w}; (1-p), \underline{w})$ with the two dimensional simplex with vertices $(\underline{w}, w, \bar{w})$; (b) find the line parallel to this intersection going through the w

vertex, this is the dashed line in the diagram; (c) define the point at which that parallel line meets the $(\underline{w}, \overline{w})$ edge of the simplex as $u(w, p)$. This is exactly the value that expected utility preferences parallel to the hyperplane through $(p, \overline{w}; (1-p), \underline{w})$ would have assigned to the sure outcome w (if the values of $\overline{w}$ and $\underline{w}$ were normalized to 1 and 0).

The proof that (**) actually represents the preferences when using the constructed u will proceed in five steps:

(1) Assigning a value to lotteries $(p, \overline{w}; (1-p), \underline{w})$.

(2) Considering lotteries on the edges of $(\underline{w}, w, \overline{w})$ simplices.

(3) Considering other two-outcome lotteries.

(4) Lotteries in a $(\underline{w}, w, \overline{w})$ simplex.

(5) General simple lotteries.

(1) Let $V[p, \overline{w}; (1-p), \underline{w}] = p$, which is obviously consistent with the preference ordering of such lotteries. Substituting in (**) gives $pu(\overline{w}, p) + (1-p)u(\underline{w}, p) = p$ as required.

(2) Consider $(\beta, \underline{w}; (1-\beta), w) \sim (p, \overline{w}; (1-p), \underline{w})$. By the previous step it is sufficient to show that $(1-\beta)u(w, p) + \beta u(\underline{w}, p) = p$. By construction $u(w, p) = p/(1-\beta)$ so $(1-\beta)p/(1-\beta) + \beta \cdot 0 = p$ as desired. A similar proof holds for lotteries on the $(w, \overline{w})$ edge.

(3) This stage in the proof shows that if a lottery of the type $(\alpha, w'; (1-\alpha), w'')$ with $w', w'' \in W$, is indifferent to $(p, \overline{w}; (1-p), \underline{w})$, then

$\alpha u(w', p) + (1-\alpha)u(w'', p) = p$. Since the proof is a simple but lengthy geometric analysis it is provided in Appendix B.

(4) Given $Q = (\underline{q}, \underline{w}; q, w; \bar{q}, \bar{w}) \sim (p, \underline{w}; (1-p), \bar{w})$ with $\underline{q} + q + \bar{q} = 1$, find t and α, as in Figure 3, such that:

$$Q = \alpha(p, \bar{w}; (1-p), \underline{w}) + (1-\alpha)(t, \bar{w}; (1-t), w)$$

and

$$Q \sim (p, \bar{w}; (1-p), \underline{w}) \sim (t, \bar{w}; (1-t), w).$$

I need to show that: $u(\underline{w}, p)\underline{q} + u(w, p)q + u(\bar{w}, p)\bar{q} = p$. By the decomposition of Q, $\underline{q} = \alpha(1-p)$, $q = (1-\alpha)(1-t)$, and $\bar{q} = \alpha p + (1-\alpha)t$. Thus, $u(\underline{w}, p)\underline{q} + u(w, p)q + u(\bar{w}, p)\bar{q} = \alpha[(1-p)u(\underline{w}, p) + pu(\bar{w}, p)] + (1-\alpha)[(1-t)u(w, p) + tu(\bar{w}, p)] = \alpha p + (1-\alpha)p = p$ since the first square brackets equal p by step (1) and the latter square brackets equal p by step (2).

(5) Given a simple measure P which assigns positive weights $p_1, \ldots, p_n$ to $w_1, \ldots, w_n$ and is indifferent to $(p, \bar{w}; (1-p), \underline{w})$ it is necessary to prove that $\sum_{i=1}^{n} p_i u(w_i, p) = p$. Consider the simplex $\Delta \subset D$ which includes all measures over $w_1, \ldots, w_n$. The intersection of the indifference hyperplane through P with Δ (this intersection is denoted by H) is a compact convex subset of Δ, thus any point $h \in H$ can be written as a finite convex combination of extreme points of H. Therefore, $P = \sum_{i=1}^{m} \lambda_j Q_j$ where $Q_j \in H$ is a lottery assigning probability q_j to w' and $(1-q_j)$ to w'' (these are extreme points of H, where $w', w'' \in \Delta$). By step (3) above $\int u(w, p) dQ_j = p$. Therefore

$\int u(w, p)dP = \int u(w, p)d(\Sigma\lambda_j Q_j) = \Sigma\lambda_j \int u(w, p)dQ_j = \Sigma\lambda_j p = p$. □

I now present the continuous representation theorem (Proposition 2) since the proofs of the properties and uniqueness results are identical for both representations and are provided in Section 3. Proposition 1 showed that there is a characterization similar to expected utility even when the independence axiom is weakened, for all simple measures on a compact consequence space. In order to get an integral representation theorem for more general measures we need more assumptions (just as in expected utility theory -- see Fishburn [5, Chapter 3]). Rather than attempt to provide equivalent theorems for all possible extension results, only one approach of special interest is considered. It allows for a continuous "local utility" function $u(w, v)$ by assuming that preferences are continuous. This is in the spirit of Grandmont [9], adapted to the more general approach of this chapter.

A2': Continuity: The sets $\{P \in D: P \succsim P^*\}$ and $\{P \in D: P^* \succsim P\}$ for all $P^* \in D$ are closed (in the topology of weak convergence).

PROPOSITION 2: Preferences over D satisfy A1(a), A2', A3, A4 if and only if there exists u: $W \times [0,1] \rightarrow \mathbb{R}$ increasing in the preference ordering of W, continuous in both its arguments, such that $P \succ Q$ (resp. $P \sim Q$) <=> $V[P] > V[Q]$ (resp. $V[P] = V[Q]$) where $V[F]$ is defined implicitly as the unique $v \in [0,1]$ that solves:

$$(*) \quad \int u(w, v)dF(w) = vu(\overline{w}, v) + (1-v)u(\underline{w}, v).$$

Furthermore, $u(w, v)$ is unique up to positive affine transformations which are continuous functions of v.

PROOF: A1(b) is implied by compactness of D and A2'. Parts (1) - (4) of the proof are as before and only part (5) changes as below, where $\int u(w, p)dQ(w)$ is a continuous linear function of $Q \in D$ since the constructed $u(w, p)$ is continuous by A2'.

(5') Given a lottery $F \sim (p, \bar{w}; (1-p), \underline{w})$ I need to show that $\int u(w, p)dF(w) = p$.

By Choquet's theorem ([14], pp. 19, 20) there exists a probability measure, say ν, on the indifference hyperplane H which includes F, such that ν represents F and is supported by the extreme points of H. An extreme point, say h, of H is one which can be represented only by the measure which assigns 1 to all Borel sets of H which include h, zero elsewhere. In this case the extreme points are those on simplex edges, i.e. of the type $(p, w'; (1-p), w'')$. Thus the continuous linear function $U(Q) \equiv \int u(w, p)dQ(w)$ for $Q \in D$ satisfies $U(F) = \int_H U(\cdot)d\nu(\cdot)$ where $\nu(L-S) = 0$ and S is the set of all distributions on the indifference hyperplane which are also on simplex edges. Since for each s in S, $U(s) = p$ (by (2) and (3) above) this shows that $U(F) = p$. □

3. Properties of the Characterization

A. $u(w, v)$ is increasing in w.

The proof that $u(w, v)$ is increasing in w relies on monotonicity. For any $v \in [0,1]$ and $w \succ w'$ consider $P \equiv (v, \overline{w}; (1-v), \underline{w})$. If $w \succ P \succ w'$ then $u(w, v) > v > u(w', v)$ by the construction of u. If $w \succ w' \succ P$ then by A2 find β and β' such that $(\beta, w; (1-\beta), \underline{w}) \sim P$ and $(\beta', w'; (1-\beta'), \underline{w}) \sim P$. Now, β' $> \beta$ since otherwise $P \sim (\beta, w; (1-\beta), \underline{w}) \succ (\beta, w'; (1-\beta), \underline{w}) \succsim (\beta', w'; (1-\beta'), \underline{w}) \sim P$ (where the strict prefernce follows from A3 and the weak preference can be derived using A4). Thus $u(w, v) = v/\beta$ is greater than $u(w', v) = v/\beta'$. The proof for the case when $P \succ w \succ w'$ is similar.

B. Uniqueness of $u(w, v)$ up to continuous positive affine transformations.

The proof of uniqueness up to affine transformations includes two steps. First I show that for any function $g(w, v)$ increasing in w, no preference functional other than those assigning value p to lotteries $F_p \equiv (p, \overline{w}; (1-p), \underline{w})$ are represented by (*). Consider these distributions F_p and a possible preference funtional $H[F_p]$. By substituting into (*), $g(\overline{w}, H[F_p])p + g(\underline{w}, H[F_p])(1-p) = g(\overline{w}, H[F_p])H[F_p] + g(\underline{w}, H[F_p])(1-H[F_p])$. And this implies that $p = H[F_p]$, since $g(\overline{w}, H[F_p]) \neq g(\underline{w}, H[F_p])$ by assumption.

The next stage asks whether, for a fixed preference function V, there exist transformations of u for which $V[F]$ is the solution to (*). Obviously $V[F]$ still solves (*) if we take a positive continuous affine transformation of

u. These are now shown to be the only acceptable transformations. Assume $\int h(w,p)dF(w) = h(\underline{w},p)(1-p) + h(\overline{w},p)p$ and $\int u(w,p)dF(w) = p$ so that h correctly solves $V[F] = p$. Define $b(p) = h(\underline{w},p)$, $a(p) = [h(\overline{w},p) - h(\underline{w},p)]$, and $g(w,p) = a(p)u(w,p) + b(p)$. I now show that $g(w,p) = h(w,p)$, so any solution $h(w,p)$ which solves (**) is a generalized affine transformation of $u(w,p)$. For any $F \sim (p,\overline{w};(1-p),\underline{w})$, $\int g(w,p)dF(w) = \int[a(p)u(w,p) + b(p)]dF(w) = [h(\overline{w},p) - h(\underline{w},p)] \int u(w,p)dF(w) + h(\underline{w},p) = [h(\overline{w},p) - h(\underline{w},p)]p + h(\underline{w},p) = \int h(w,p)dF(w)$. Now consider $F_w = (\beta,w;(1-\beta),\underline{w})$ or $F_w = (\gamma,w;(1-\gamma),\overline{w})$ such that $F_w \sim F$, one of which exists by solvability. Then either $h(w,p)\beta + h(\underline{w},p)(1-\beta) = g(w,p)\beta + g(\underline{w},p)(1-\beta)$ or $h(w,p)\gamma + h(\overline{w},p)(1-\gamma) = g(w,p)\gamma + g(\overline{w},p)(1-\gamma)$ but since by construction $h(\overline{w},p) = g(\overline{w},p)$ and $h(\underline{w},p) = g(\underline{w},p)$ this implies $h(w,p) = g(w,p)$.

C. Uniqueness of the implicit solution.

Since $V[F]$ is defined implicitly, it is necessary to show that the solution to the implicit function is unique. This is done by considering the expected utility extension of these preferences. Assume the solution to (**) is not unique, i.e. in addition to the correct solution v, there exists $\hat{v} \in [0,1]$, $\hat{v} \neq v$ such that $\int u(w,v)dF(w) = v$, $\int u(w,\hat{v})dF(w) = \hat{v}$, and $\int u(w,\hat{v})d\hat{F}(w) = \hat{v}$ where $\hat{v}$ is the correct solution for $\hat{F}$. (Solutions where $\hat{v} \notin [0,1]$ can be ignored since, even if they solve equation (**) they lie outside the range of permissible values -- recall that $v \in [0,1]$.) Holding $\hat{v}$ constant consider

$\hat{u}(w) = u(w, \hat{v})$ as a Bernoulli utility function which defines expected utility preferences through $\hat{F}$ and $(\hat{v}, \overline{w}; (1-\hat{v}), \underline{w})$ but not through F (the latter by assumption that $\hat{v}$ is not the correct solution for F). However, by assumption also $\int \hat{u}(w)dF(w) = \hat{v} = \int \hat{u}(w)d\hat{F}(w)$ implying that F and $\hat{F}$ do lie in the same indifference hyperplane.

D. For every $w \in W$, $u(w, v)$ is continuous in v on the open interval $(0,1)$.

First fix w not indifferent to $\overline{w}$, and consider the simplex with vertices $\underline{w}$, w, and $\overline{w}$. Let $B(w) = \{v \in [0,1] \mid (v, \overline{w}; (1-v), \underline{w}) \succsim w\}$ and note that $B(w)$ is a closed interval from some $\overline{v}$ to 1. The function $\beta(v)$ (which was used in the constuction of $u(w, v)$) is defined as the solution of $(\beta, \overline{w}; (1-\beta), w) \sim (v, \overline{w}; (1-v), \underline{w})$ for any $v \in B(w)$. Clearly $\beta(\overline{v}) = 0$, $\beta(1) = 1$, and $\beta(\cdot)$ is an increasing function (otherwise two indifference lines in the simplex will cross). I now show that $\beta(\cdot)$ is continuous. If not then there exists $v_n \uparrow v$ with $\beta(v) > \lim\beta(v_n)$. Choose $\hat{\beta}$ satisfying $\beta(v) > \hat{\beta} > \lim\beta(v_n)$ and let $\hat{v} = \beta^{-1}(\hat{\beta})$. $\hat{v} = v - \epsilon$ for some $\epsilon > 0$ since $(\hat{\beta}, \overline{w}; (1-\hat{\beta}), w) \sim (\hat{v}, \overline{w}; (1-\hat{v}), \underline{w}) \prec (v, \overline{w}; (1-v), \underline{w}) \sim (\beta(v), \overline{w}; (1-\beta(v)), w)$. But $\hat{v} > v_n$ for every n since $(\hat{v}, \overline{w}; (1-\hat{v}), \underline{w}) \succ (v_n, \overline{w}; (1-v_n), \underline{w})$. And this contradicts $v_n \uparrow v$, so $\beta(\cdot)$ is continuous. Recall that by construction $u(w, v)$ is equal to $\frac{v-\beta(v)}{1-\beta(v)}$ for $v \in (\overline{v}, 1)$ and equal to $\overline{v}$ when $v = \overline{v}$, so $u(w, v)$ is continuous for $v \in [\overline{v}, 1)$. A similar proof shows that $u(w, v)$ is continuous for $v \in (0, \overline{v}]$. For $w \sim \overline{w}$ (resp. $w \sim \underline{w}$) monotonicity implies that $u(w, v) = u(\overline{w}, v) = 1$ (resp. $u(w, v) = u(\underline{w}, v) = 0$).

Continuity is necessary to avoid indifference sets which do not separate the simplex into two disconnected sets. For example, in the simplex with vertices $\underline{w}$, w and $\bar{w}$, if $u(w, v) = 1/4$ for $v < 1/2$ and $u(w, v) = 1/2$ for $v \geq 1/2$ there would be indifference lines which end inside of the simplex.

4. Extending Local Properties

This section relates this representation of preferences to Machina's [12] work on non-expected utility preferences. If all preferences satisfying A1 - A4 had a preference functional $U[F]$ which was everywhere Frechet differentiable then Machina's extension results (that local monotonicity and risk aversion everywhere in D imply global monotonicity and risk aversion) would go through. It is shown below that this is not true, in particular it is shown how to construct a set of indifference lines in the simplex which have no differentiable preference functional representation. However, an alternative approach to extending local results is presented. Rather than examine the first order approximation to the preference functional (which may not be smooth), if the indifference sets are smooth manifolds than the first order approximation to an indifference curve can be taken, and extended to parallel hyperplanes, giving an expected utility approximation. For preferences considered in this chapter such an extension is simple. It is shown that for $W \subset \mathbb{R}$ if $u(w, \cdot)$ is increasing in w, then P first order stochastically dominates Q if and only if $V[P] > V[Q]$. (Note that the property shown in the previous section is that $u(w, v)$ is increasing in w with respect to the preference order on W. Any conclusion on stochastic dominance requires that this induced order is the natural order on the reals, i.e. $w \succ w'$ if and only if $w > w'$.) Furthermore it is proven that if $u(w, v)$ is concave in w for every v then the individual is averse to mean preserving increases in risk. These two properties do follow from the discussion after Proposition 1 which explained why a class of results in expected utility theory extends to the implicit expected utility framework. However the direct proofs of these properties, while brief, further clarify the relationship between the expected utility representation and this chapter.

Let $f\colon [0, \frac{1}{2}] \to [0, a]$ for some $a \in (\frac{1}{2}, 1)$ be continuous, strictly increasing with derivative zero a.e. (see Billingsley [1, Example 31.1]). Define the indifference sets in a simplex as in Figure 4. Let V be a functional representing these preferences, so that $V[1, f(y)] = V(0, y)$ for $y \in [0, \frac{1}{2}]$ where the second argument indicates the distance along the simple edge from the lower vertex, and the first argument indicates which edge (1 for the lower sloped edge, and 0 for the vertical edge). If V is differentiable then $V_2(0, y) = V[1, f(y)]f'(y)$ wherever f is differentiable. Hence $V_2(0, y) = 0$ a.e., implying that $V(0, y)$ is constant for $y \in [0, \frac{1}{2})$, (note that $V(0, \cdot)$ is absolutely continuous since it is everywhere differentiable and monotone). However, since these preferences are by assumption strictly increasing along the vertical edge of the simplex, we have a contradiction. Therefore V can not be differentiable.

PROPERTY 1: The following statements are equivalent:

(a) For any $P, Q \in D$ if P stochastically dominates Q then P is preferred to Q.

(b) $u(w, v)$ is increasing in w.

PROOF: Assume that P first order stochastically dominates Q, while p and q which solve $\int u(w,p)\mathrm{d}P(w) = p$ and $\int u(w,q)\mathrm{d}Q(w) = q$ satisfy $p < q$. The indifference hyperplane through Q separates D into two convex sets:

$$\mu \in U \implies \int u(w,q)\mathrm{d}\mu(w) \geq \int u(w,q)\mathrm{d}Q(w) = q$$
$$\nu \in L \implies \int u(w,q)\mathrm{d}\nu(w) \leq \int u(w,q)\mathrm{d}Q(w) = q.$$

Since P stochastically dominates Q, $P \in U$. If $p < q$ then $(p,\overline{w};(1-p),\underline{w}) \in L$ since $pu(\overline{w},q) + (1-p)u(\underline{w},q) = p < q$. So the convex indifference surface through P and $(p,\overline{w};(1-p),\underline{w})$ lies both above and below the separating hyperplane, thus two indifference sets intersect, obviously leading to a contradiction. The converse is straightforward. □

PROPERTY 2: Concavity of $u(w,v)$ in w implies risk aversion (in the sense that the individual is weakly averse to mean preserving increases in risk).

PROOF: Assume that $u(w,v)$ is concave in w for every v, and that G differs from F by a mean preserving increase in risk. Hence $\int u(w,v)\mathrm{d}[G(w)-F(w)] < 0$. Let p and q solve $\int u(w,p)\mathrm{d}F(w) = p$ and $\int u(w,q)\mathrm{d}G(w) = q$. Then $q = \int u(w,q)\mathrm{d}[F(w) + (G(w)-F(w))] < \int u(w,q)\mathrm{d}F(w)$ so F lies above the indifference hyperplane through q. If $q > p$ then $(p,\overline{w};(1-p),\underline{w})$ lies below the indifference hyperplane through q. But by betweenness the indifference set which includes $F \sim (p,\overline{w};(1-p),\underline{w})$ is convex, intersecting the separating indifference hyperplane through q, leading to a contradiction; hence $q < p$. □

Appendix

A. The Representation Theorem Without Monotonicity

PROPOSITION A.1: Preferences over D_s (resp. D) satisfy A1, A2 and A4 (resp. A1(a), A2' and A4) if and only if there exists $u(\cdot,\cdot)$: $W \times [0,1] \rightarrow \mathbb{R}$, continuous in the second argument (resp. continuous in both arguments), such that $P \succ Q \iff V[P] > V[Q]$ and $P \sim Q \iff V[P] = V[Q]$ where $V[F]$ is defined implicitly as the unique $v \in [0,1]$ that solves:

$$(*) \quad \int u(w, v)dF(w) = vu(\overline{w}, v) + (1-v)u(\underline{w}, v).$$

Furthermore, $u(w, v)$ is unique up to positive affine transformations which are continuous in v.

The proof follows essentially the same lines as the proofs of Propositions 1 and 2. However, monotonicity of $u(w, v)$ in w cannot be proven without A3 (see Section 3.A).

B. Step 3 in Proving the Representation Theorem

For any distribution $R = (\theta, w'; (1-\theta), w'')$ consider the three dimensional simplex with vertices $(\underline{w}, \overline{w}, w', w'')$, where without loss of generality assume $w'' \succ w'$. By A2 find p such that $R \sim (p, \overline{w}; (1-p), \underline{w})$, where this last lottery is the point B in the three-dimensional simplex.

Construct the indifference hyperplane through R, (see Figure 5). I want to show that $\theta u(w', p) + (1-\theta)u(w'', p) = p$. Define C as the lottery

$(\gamma, \underline{w}; (1-\gamma), w'') \sim B$, and E as the degenerate lottery w''. Thus, $CE = \gamma$ and $u(w'', p) = p/(1-\gamma)$ by case (ii) on page 9, with w replaced by w''. Using trigonometric identities based on equilateral triangles with edges of length normalized to 1, using also the lengths γ, θ, and $1-\theta$ it will be shown that $u(w', p) = \frac{\theta-\gamma}{\theta} \cdot \frac{p}{1-\gamma}$ giving the desired result. Take a parallel shift of the indifference plane through R, D and C such that the new plane intersects w'. This new plane intersects the $(\underline{w}, w'')$ edge at point C' and the $(\underline{w}, \overline{w})$ edge at B'. Recall that by definition $u(w', p)$ is the length of the segment between B' and $\underline{w}$. Therefore $u(w', p)$ can be found from the length of RC in triangle REC, where $RE = \theta$, $CE = \gamma$, and $\measuredangle\, REC = 60^\circ$, giving $RC = (\theta^2+\gamma^2-\gamma\theta)^{1/2}$. Then $\cos\measuredangle\, RCE = (2\gamma-\theta)/2(\theta^2+\gamma^2-\gamma\theta)^{1/2}$, and $\cos\measuredangle\, CRE = 2(\theta-\gamma)/2(\theta^2+\gamma^2-\gamma\theta)^{1/2}$. By examining the trapezoid with corners (w', R, C, C') one can see that length $CC' = (1-\theta)\gamma/\theta$ and thus length $\underline{w}C' = (\theta-\gamma)/\theta$. Looking now at the $(\overline{w}, \underline{w}, w'')$ simplex, $\sin\measuredangle\, \underline{w}BC = \sin 60(1-\gamma)/BC$ and $\sin\measuredangle\, BC\underline{w} = p(\sin 60)/BC$. Thus, since $u(w', p) = B'\underline{w} = \underline{w}C'(\sin BC\underline{w})/(\sin \underline{w}BC)$, it has been shown that $u(w', p) = (\theta-\gamma)p/\theta(1-\gamma)$ as desired. □

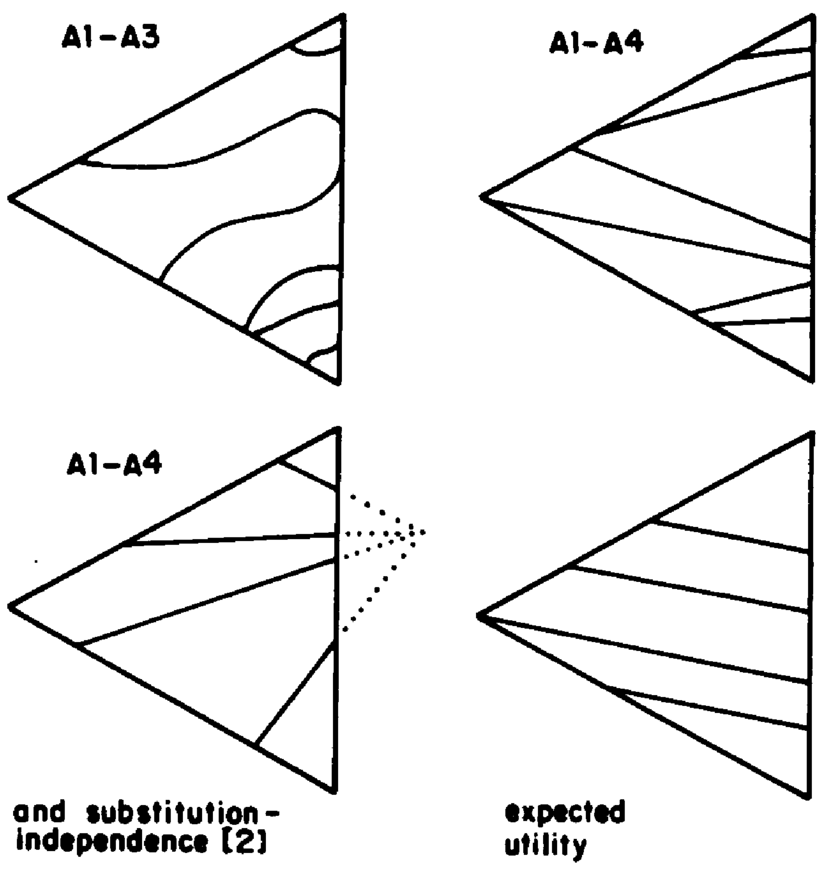
A1-A3
A1-A4
A1-A4
and substitution-
independence [2]
expected
utility

Figure 1

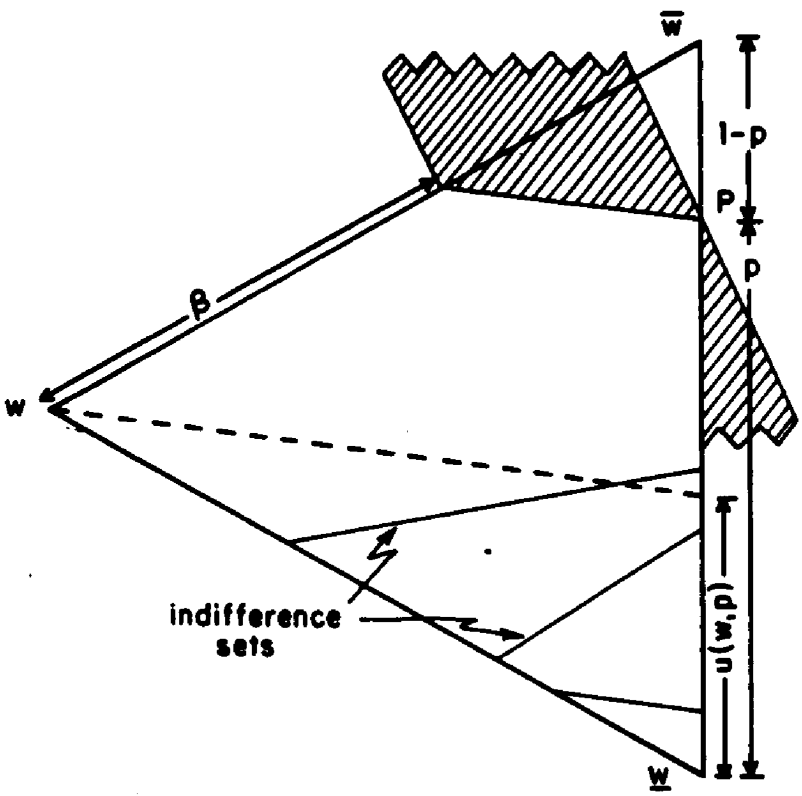

Figure 2

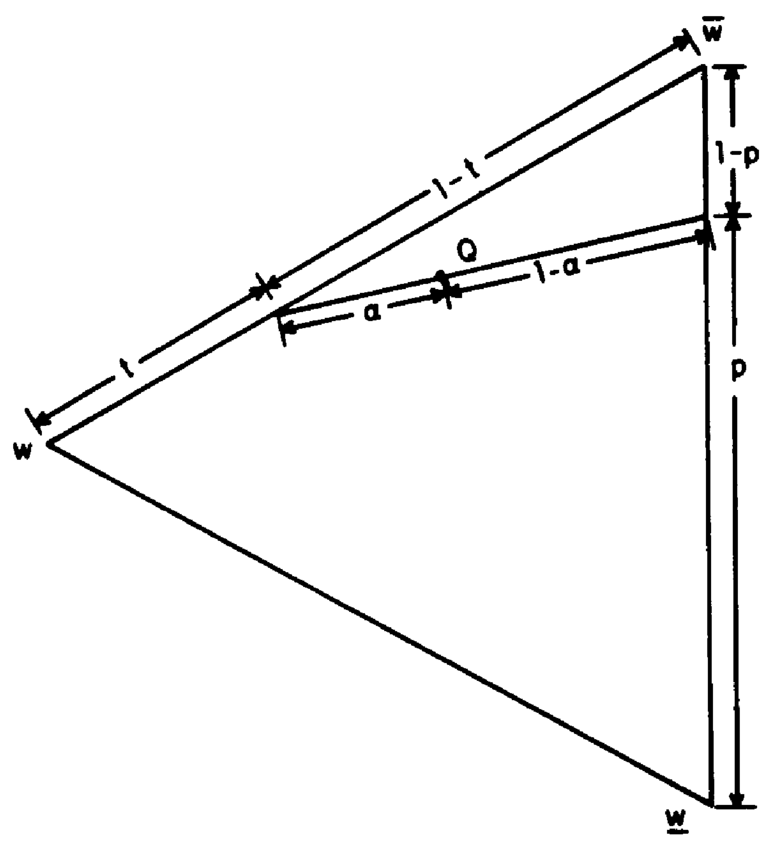

Figure 3

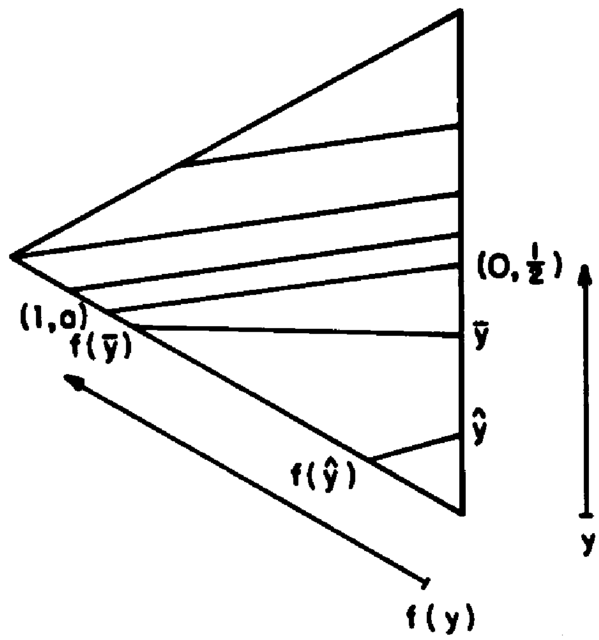

Figure 4

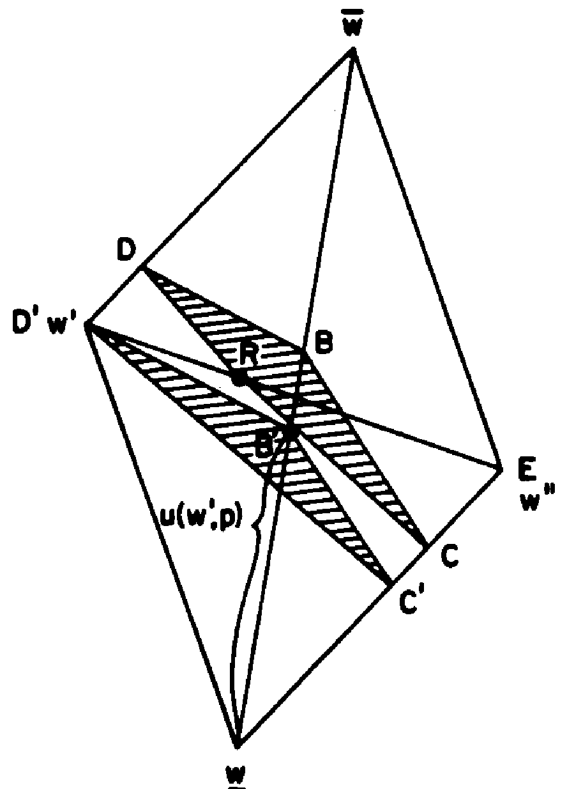

Figure 5

References

[1] P. Billingsley: *Probability and Measure*, New York, Wiley, 1979.

[2] S.H. Chew: A Generalization of the Quasi-Linear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox, *Econometrica*, Vol. 57 (1983), 1065-1092.

[3] S.H. Chew: "Weighted and Semi Weighted Utility Theories, M-Estimators and Non-Demand Revelation of Second Price Auctions for an Uncertain Auctioned Object," Mimeograph, Johns Hopkins University, June 1985.

[4] E. Dekel: "Asset Demands Without the Independence Axiom," Mimeograph, Harvard University, 1984.

[5] P.C. Fishburn: *The Foundations of Expected Utility*, Dordrecht,Reidel, 1982.

[6] P.C. Fishburn: Nontransitive Measurable Utility, *Journal of Mathematical Psychology*, Vol. 26 (1982), 31-67.

[7] P.C. Fishburn: Transitive Measurable Utility, *Journal of Economic Theory*, Vol. 31 (1983), 293-317.

[8] P.C. Fishburn: "Implicit Mean Value and Certainty Equivalence," Mimeograph, AT&T Bell Laboratories, 1984.

[9] J.M. Grandmont: Continuity Properties of a Von Neumann-Morgenstern Utility, *Journal of Economic Theory*, Vol. 4 (1972), 45-57.

[10] J. Green, "Dynamic Consistency and Non-linear Utility Theory," H.I.E.R. Discussion Paper No. 1053, April 1984.

[11] D. Kreps and E.L. Porteus: Temporal Von Neumann-Morgenstern and Induced Preferences, *Journal of Economic Theory*, Vol. 20 (1979), 81-109.

[12] M.J. Machina: 'Expected Utility' Analysis Without the Independence Axiom, *Econometrica*, Vol. 50 (1982), 277-323.

[13] M.J. Machina: Temporal Risk and the Nature of Induced Preferences, *Journal of Economic Theory*, 33 (1984), 199-231.

[14] R.R. Phelps: *Lectures on Choquet's Theorem*, Princeton, NJ, D. van Nostrand Co. Inc., 1966.

[15] L.G. Epstein: "Implicitly Additive Utility and the Nature of Optimal Economic Growth," Mimeograph, University of Toronto, Jan. 1986.

CHAPTER 2

RISK AVERSION AND PORTFOLIO DIVERSIFICATION

This chapter examines the relationship between properties of a preference function over distributions, V, and the quasiconcavity of the induced preferences over random variables. The primitive for the analysis of preferences under uncertainty is generally V. However an important application is to markets for assets (random variables). Quasiconcavity of the preferences over assets is of interest for the technical reason that it implies that the demand correspondence for assets will be convex valued. Moreover, it is equivalent to a preference for portfolio diversification. If preferences are linear in probabilities then concavity of the utility function, i.e. risk aversion, is equivalent to three behavioral properties: (i) a preference for substituting the mean of any risky outcome for the outcome, (ii) an aversion to mean preserving increases in risk and (iii) a preference for portfolio diversification. Machina [2] and Chew and Mao [1] have shown that for a general class of preferences the first two properties are equivalent. However, in general the third is not equivalent to the other two for non-expected utility preferences. After developing the framework for analysis I provide a class of examples which show that preferences can exhibit property (ii) while exhibiting an aversion to diversification. On the other hand (iii) is shown to imply (i). I go on to prove that quasiconcavity of V together with risk aversion (in the sense of (i) and (ii)) is sufficient to guarantee portfolio diversification. Quasiconcavity of V is shown to be not necessary by constructing a class of examples. These examples help clarify the role of quasiconcavity by demonstrating how it can be relaxed if the risk aversion is sufficient.

Let $V: D \to \mathbb{R}$ be a preference function over the space of probability distributions on $[0, 1]$, which is continuous in the topology of weak convergence and is consistent with first order stochastic dominance. The random variables x^i ($i \geq 1$) on the probability space $([0, 1], B, \lambda)$ (where B is the Borel field on the unit interval and λ is the Lebesgue measure) have probability distributions $F(x^i; \cdot)$ which are also denoted F^i. Also, for any n assets x^i, $i=1,\ldots, n$ define the diversified asset x^α by $x^\alpha(s) \equiv \sum \alpha^i x^i(s)$ for every s, where $\alpha^i \geq 0$ and $\sum \alpha^i = 1$. F^α denotes the distribution $F(x^\alpha; \cdot)$ induced by the diversification, while $\alpha \cdot F$ is the convex combination of the distributions, i.e. $\alpha \cdot F \equiv \sum \alpha^i F^i$.

DEFINITION 1: V exhibits risk aversion if:

(i) $V(F) \geq V(G)$ whenever G is a mean preserving spread of F, or

(ii) $V[p\tilde{F} + (1-p)F] \leq V[p\tilde{F} + (1-p)\delta_{E(F)}]$.

(δ_c denotes the distribution with point mass at c, and E is the expectation operator.) These two properties are equivalent for preferences which are consistent with first order stochastic dominance and continuous [1, 2]. If V is Fréchet differentiable then they are also equivalent to concavity of the local utility function $u(\cdot, F)$ [2]. The local utility function satisfies $\int u(\cdot, F) d(\tilde{F}-F)$

$= \Psi(\tilde{F}-F, F)$ where the latter is the Fréchet differential of V at F in the direction of $\tilde{F}$. Roughly speaking $u(\cdot,F)$ is the Bernoulli utility function of the linear approximation to V at F. (The approximation exists by the assumption of differentiability, and its linearity implies that the expected utility axioms are satisfied so a utility function exists.)

DEFINITION 2: V exhibits divérsification if for any $n \geq 1$ and any random variables x^i, $i=1,\ldots,n$:

$$V(F^1) = \cdots = V(F^n) \text{ implies } V(F^\alpha) \geq V(F^1) \ \forall\alpha$$

I now show that although the equivalence of the definitions of risk aversion in terms of (i) and (ii) extends to non-linear preferences, a similar extension of the equivalence of property (iii) fails. To do so I construct a Fréchet differentiable preference function, with concave local utility functions for which there exist assets x^1, x^2 and an α such that $V(F^\alpha) < V(F^1) = V(F^2)$.

THEOREM 1: There exist V which do exhibit risk aversion but do not exhibit diversification.

PROOF: First choose any two assets x^1 and x^2 with different means where neither second order stochastically dominates (SSD) the other. Assume that $E(F^1) > E(F^2)$. Now choose an increasing and concave ν such that $\int\nu dF^1 < \int\nu dF^2$. ($\nu$ exists since F^1 does not SSD F^2). Affinely normalize ν so that the first integral equals $E(F^2)$ and the second integral equals $E(F^1)$. Clearly $\int dF^1 > \int dF^\alpha > \int dF^2$. For α^1 sufficiently close to 1: $\int\nu dF^1 < \int\nu dF^\alpha < \int\nu dF^2$, where the first inequality follows from concavity of ν and the second from continuity of ν. Choose an $\bar{\alpha}^1$ sufficiently close to 1 for which the last inequality holds and then choose an increasing and differentiable g such that: $g[E(F^1)] + g[E(F^2)] > g(\int \nu dF^{\bar{\alpha}}) + g(\int dF^{\bar{\alpha}})$, where $\bar{\alpha} = (\bar{\alpha}^1, 1-\bar{\alpha}^1)$. Let $V(F) = g(\int \nu dF) + g(\int dF)$. By construction V does not exhibit diversification ($F^{\bar{\alpha}}$ is less preferred than F^1 which is indifferent to F^2). On the other hand the local utility functions of V are $u(\theta, F) = g'(\int \nu dF)\nu(\theta) + g'(\int dF)\theta$ and are concave by construction so V exhibits risk aversion. □

In [4] Tobin discussed diversification and risk aversion for the case of preferences over means and variances of distributions, $U(\mu, \sigma^2)$. A risk averter is defined as having a positive tradeoff between these two moments, that is an upward sloping indifference curve, and a "plunger" (i.e. non-diversifier) is a risk averter with quasiconvex preferences (over (μ, σ^2)). Tobin noted that: "If the category defined as *plungers* ... exists at all, their indifference curves must be determined by some process other than those described in 3.3" [4, p. 77], where Section 3.3 derived mean-variance preferences from expected utility preferences with either normal distributions or quadratic Bernoulli utility functions. Theorem 3 below shows that a necessary condition for plungers is quasiconvexity of V in F (although this isn't sufficient as in the mean-variance case). The example above showed how preferences exhibiting risk aversion and plunging can be derived from general preferences over distributions.

It has just been shown that the sufficiency of risk aversion for diversification in the case of expected utility preferences does not extend to more general preferences. However the reverse implication, that is the necessity of risk aversion, does extend to general V's.

THEOREM 2: If V does not exhibit risk aversion then V does not exhibit diversification.

PROOF: If V does not exhibit risk aversion then there exist F, F and t such that $V[(1-t)\tilde{F} + tF] < V[(1-t)\tilde{F} + t\delta_{E(F)}]$. First assume that F is a simple distribution (i.e. with finite support) which assigns rational probabilities p_k to outcomes θ_k. Rewrite F as an equal probability distribution assigning probability $1/m$ to $\bar{\theta}_1,\ldots,\bar{\theta}_m$. Let y be a random variable with the distribution $\tilde{F}$. Now for $k = 1,\ldots,m$ define the following assets:

$$x^k(s) = \bar{\theta}_{[i+k]} \text{ if } s \in (t\frac{i-1}{m}, t\frac{i}{m}]$$

$$x^k(s) = y(\frac{s-t}{1-t}) \text{ if } s > t$$

(Where $[i+k] = i+k$ modulo m.) For each k, x^k clearly has the distribution $(1-t)F + tF$ (s is greater than t with probability $1-t$ and then x^k is just y contracted from the unit interval to $(t, 1]$ and for $s \le t$, x^k is a permutation of the random variable which assigns $\bar{\theta}_i$ to the i'th interval of measure $t\frac{1}{m}$). On the other hand x^α for $\alpha = (1/m,\ldots,1/m)$ has the distribution $(1-t)\tilde{F}+t\delta_{E(F)}$. To conclude note that $V(F^k) = V[(1-t)\tilde{F} + tF] < V[(1-t)\tilde{F} + t\delta_{E(F)}] = V(F^\alpha)$. If the p_k aren't rational then a similar construction gives assets with distributions arbitrarily close to F, which is sufficient since V is continuous by assumption. Similarly if F is not simple then consider a sequence of simple distributions $F_n \uparrow F$, where for n sufficiently large $V[(1-t)\tilde{F} + tF_n] < V[(1-t)\tilde{F} + t\delta_{E(F_n)}]$ by continuity of V. □

Since risk aversion is not a sufficient condition for quasiconcavity of the induced preferences over assets, and the latter is an important assumption for the analysis of asset markets, I now turn to find conditions which imply this property (and hence diversification). The condition required is quasiconcavity of V in addition to risk aversion. However, quasiconcavity of V is not necessary for diversification, as the example following Theorem 3 shows. That example does help clarify the role of quasiconcavity of V, since in it quasiconcavity can be relaxed only by putting a lower bound on the risk aversion.

THEOREM 3: If V is quasiconcave in F and V exhibits risk aversion then V exhibits diversification.

PROOF: I first show that if V exhibits risk aversion then $V[F^{\alpha}] \geq V[\alpha \cdot F]$. Let u be an arbitrary concave Bernoulli utility function. Then $\int u dF^{\alpha} = \int u[\sum \alpha^i x^i(s)] d\lambda(s) \geq \sum \alpha^i \int u[x^i(s)] d\lambda(s) = \sum \alpha^i \int u dF^i = \int u d(\alpha \cdot F)$. But since u is an arbitrary concave function and $E(F^{\alpha}) = E(\alpha \cdot F)$ it follows from Rothschild and Stiglitz [3] that $\alpha \cdot F$ is a mean preserving spread of F^{α}, so $V(F^{\alpha}) \geq V(\alpha \cdot F)$. Now note that quasiconcavity of V implies that $V(\alpha \cdot F) \geq$

$\min_i \{V(F^i)\}$ which together with the preceding observation implies diversification. □

The proof shows that $V(F^\alpha) \geq V(\alpha \cdot F) \geq V(F^i)$, where the first inequality follows from risk aversion, and the second from quasiconcavity of V in F. These two inequalities are important for understanding the results of this paper. The necessity of risk aversion was shown by finding F^i's such that the second inequality held with equality because the assets had the same distribution, while the first was reversed from lack of risk aversion (since the equal proportion diversification among the assets gave the expected value of the distribution). That risk aversion alone was not sufficient was demonstrated by finding a case where the reversal of the second inequality through lack of quasiconcavity of V in F was "stricter" than the first inequality (which remained correct because of risk aversion). Finally I will now show that quasiconcavity is not necessary by constructing a slightly non-quasiconcave example where the risk aversion inequality is always "stricter" than the reversal of the second (quasiconcavity) inequality.

Given a concave increasing ν (with strictly negative second derivative), define $V(F) = g(\int \nu dF) + g(\int dF)$, where g satisfies $0 < g''(c) \leq \inf_{\theta \in [0,1]} |-\nu''(\theta)|$ and $g'(c) \geq [\sup_{\theta \in [0,1]} \nu'(\theta)]^2 + 1 > 0$ for all c in the range of $\int dF$ and $\int \nu dF$. The following two Lemmas imply that this V exhibits diversification even though it is not quasiconcave in F. Note that the convexity of V in F will depend on the convexity of g. On the other hand the risk aversion coefficient for the local utility functions of V is equal to $\frac{\nu'}{1+\nu''}$. And this is bound from below by $\frac{g''}{g'}(1+\nu') \geq \frac{g''}{g'}$.

LEMMA 1: V is convex.

PROOF: $V[\alpha F^1 + (1-\alpha)F^2] = g[\alpha \int dF^1 + (1-\alpha)\int dF^2] + g[\alpha \int \nu dF^1 + (1-\alpha)\int \nu dF^1] \leq \alpha(g[\int dF^2] + g[\int \nu dF^2]) + (1-\alpha)(g[\int dF^1] + g[\int \nu dF^2]) = \alpha V(F^1) + (1-\alpha)V(F^2)$. □

LEMMA 2: V exhibits diversification.

PROOF: Let $V(F^1) = V(F^2)$, and I will show that this implies $V(F^\alpha) \geq V(F^1)$, for $\alpha = (\alpha^1, 1-\alpha^1)$ with $\alpha^1 \in [0, 1]$. Consider $H(\alpha^1) \equiv V(F^\alpha)$ as a function of α^1. Since by assumption $H(0) = H(1) = V(F^1)$ it is sufficient to show that H'' ≤ 0.

$$
\begin{aligned}
H' &= g'(\textstyle\int dF^\alpha)\int(x^1(s) - x^2(s))d\lambda(s) \\
&\quad + g'(\textstyle\int \nu dF^\alpha)\int[\nu'(x^\alpha(s))(x^1(s) - x^2(s))]d\lambda(s). \\
H'' &= g''(\textstyle\int dF^\alpha)\{\int[x^1(s) - x^2(s)]d\lambda(s)\}^2 \\
&\quad + g''(\textstyle\int \nu dF^\alpha)\{\int[\nu'(x^\alpha(s))(x^1(s) - x^2(s))]d\lambda(s)\}^2 \\
&\quad + g'(\textstyle\int \nu dF^\alpha)\int[\nu''(x^\alpha(s))(x^1(s) - x^2(s))^2]d\lambda(s) \\
&\quad \leq A\{g''(\textstyle\int dF^\alpha) + B^2 g''(\int \nu dF^\alpha) + Cg'(\int \nu dF^\alpha)\}
\end{aligned}
$$

where $A = \int[x^1(s) - x^2(s)]^2 d\lambda(s)$, $B = \sup_{\theta \in [0,1]} \nu'(\theta)$ and $C = \sup_{\theta \in [0,1]} \nu''(\theta)$. Recall that $0 < g'' \leq -C$ so the last line is in fact less than or equal to: $AC[-1 - B^2 + g'(\int \nu dF^\alpha)]$. However, $g' \geq 1 + B^2$ so the last expression is in fact non-positive. □

References

[1] S.H. Chew and M.H. Mao: "A Schur-Concave Characterization of Risk Aversion for Nonlinear, Nonsmooth Continuous Preferences," Mimeograph, Johns Hopkins University, Aug. 1985.

[2] M.J. Machina: "'Expected Utility' Analysis without the Independence Axiom," *Econometrica*, Vol. 50 (1982), 277-323.

[3] M. Rothschild and J. Stiglitz: "Increasing Risk: I. A Definition," *Journal of Economic Theory*, Vol. 2, (1970), 225-243.

[4] J. Tobin: "Liquidity Preference as Behavior Toward Risk," *Review of Economic Studies*, Vol. 25, (1957-8), 65-86.

CHAPTER 3

COMMON KNOWLEDGE WITH PROBABILITY 1

1. Introduction

The idea of common knowledge is central in game theory and the economics of uncertainty and information. For example, the noncooperative analysis of a game (with complete information) starts with the assumption that the structure of the game is common knowledge among the players. Intuitively speaking, two people 1 and 2 are said to have common knowledge of an event if both know it, 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on.

Common knowledge was first given a formal definition by Aumann in [1]. In [1], 1 and 2's private information is represented by a pair of partitions of some state space Ω. Individual i is said to know an event A at some state of the world ω if the member of i's information partition that contains ω is itself contained in A. Using this definition of what it means to know an event, Aumann shows that an event A is common knowledge at ω if and only if A contains the member of the meet (finest common coarsening) of 1 and 2's partitions that contains ω.

An important restriction on the information structure in [1] is that the join (coarsest common refinement) of 1 and 2's partitions is assumed to consist of nonnull events. But many decision problems naturally call for a general -- possibly infinite -- state space, in which case null events (in the join) must be permitted. In this chapter I define common knowledge in this more general situation. (Of course, this definition coincides with Aumann's when his applies.) I start with the observation that according to Bayesian decision theory, the basic definition of knowledge must be in terms of beliefs. That is, to say a person knows an event A at some state ω means that (s)he assigns A posterior probability one at ω. Having defined what it means for someone to know an event A at ω, I can go on to define common knowledge of A at ω. (Nielsen [9] provides a different definition of common knowledge on an infinite state space using Boolean σ-algebras.) The main result proven in this chapter is an equivalence between a definition of common knowledge in terms of beliefs and a definition in terms of the σ-fields representing 1 and 2's information.

Apart from the intrinsic interest in defining common knowledge on an infinite state space, there is one fundamental issue which can only be addressed if one is able to define common knowledge on an infinite Ω. In both [1] and this chapter, the information partitions and priors (i.e. the information structure on Ω) are assumed to be common knowledge in an informal sense. I say "in an informal sense" because the information structure is not an event in Ω and a formal mathematical definition of common knowledge applies only to events in Ω. Of course, any mathematical theorems one proves on common knowledge -- such as the equivalence result in this chapter -- are true whether or not one assumes that the information structure is common knowledge, since the theorems hold regardless of interpretation. However, to interpret the theorems as statements about common knowledge, it is necessary to make the assumption that the information structure is common knowledge. As argued by Aumann in [1], if this assumption is not satisfied then the state space can (and should) be expanded. In Chapter 4 I find the expanded state space such that if common knowledge is defined on this space, then the assumption that the information structure is common knowledge is without loss of generality. And the point is that this expanded state space is infinite even if the underlying state space is finite. (The expanded state space in Chapter 4 is the product of the underlying space of uncertainty, S say, and the spaces of all possible "types" of 1 and 2. Following Harsanyi [5] and Mertens and Zamir [7], a type of person 1 (resp. 2) is an infinite hierarchy of beliefs -- over S, over 2's (resp. 1's) beliefs over S, and so on. The type spaces are infinite even if S is finite.)

Bayesian decision theory suggests that the basic definition of common knowledge should be in terms of beliefs -- a person knows an event means that (s)he assigns it (posterior) probability one. Another approach would be to say that a person knows an event if (s)he is informed that it occurs -- this is a definition in terms of an information partition/σ-field. The main result I prove in this chapter is an equivalence between the definition of common knowledge in terms of beliefs (probability measures) and a definition in terms of the intersection (meet) of 1 and 2's σ-fields. To obtain this result, these two approaches have to be reconciled. This is achieved by assuming that the probability measures are regular and proper (see Definition 2.2) and that the σ-fields are completed in a suitable manner (see Definition 3.2).

Recall that Bayes' rule says nothing about how individual i updates his/her beliefs over Ω if informed that a null event occurs. Regularity says that i must have a belief over Ω even if informed of some null partition cell h. But it is quite possible for i to ignore the fact that (s)he was informed that the true state lies in h, i.e. to assign positive (posterior) probability to states outside of h. Properness requires that after being informed of any partition cell h, i assigns (posterior) probability one to h even if h has prior probability zero. The use of the term "properness" originated in [3] (see also [2]). In [2] and [3] it is argued that an intuitively satisfactory theory of probability should involve proper regular conditional probabilities.

To see the role of completion of the σ-fields, suppose that i has no information, i.e has the trivial partition $\{\Omega\}$, but i assigns probability one to some proper subset A of Ω. Then the only event of which i is informed is Ω, but i knows A according to the definition in terms of beliefs. i's partition should be refined by adding in the events to which (s)he assigns probability one or -- what amounts to the same thing -- the events to which i assigns probability zero. That is, i's partition must be completed. Actually, the issue of completion is rather more delicate than this argument suggests. What is needed is "posterior completion," which is different from the standard definitions of completion in probability theory.

The organization of the rest of the chapter is as follows. Section 2 begins with a review of Aumann's definition of common knowledge, and then the definition in terms of beliefs (Definition 2.1) is introduced. I go on to motivate the restriction that the conditional probabilities be proper, and using properness show that Definition 2.1 is implied by an "informational" definition in terms of σ-fields (Proposition 2.1). In Section 3 I start by showing that the converse to Proposition 2.1 is false, and then explain how posterior completion of the σ-fields rectifies this situation. I conclude with the main equivalence result (Proposition 3.3).

2. Common Knowledge in Terms of Beliefs

I begin with a review of Aumann's definition of common knowledge [1]. There is a measurable space (Ω, F), where Ω is the space of states of the world and F is a σ-field of subsets of Ω. Individual i's (i=1,2) information about the state of the world is represented by a partition H^i of Ω. 1 and 2 have a common prior which assigns positive probability to every event in the join $H^1 \vee H^2$ of 1 and 2's partitions.

Consider an event $A \in F$ and a state of the world $\omega \in \Omega$. i is said to know A at ω if $H^i(\omega) \subset A$, where $H^i(\omega)$ is the member of H^i that contains ω. An event A is said to be common knowledge at some state ω if 1 knows A at ω, 2 knows A at ω, 1 knows 2 knows A at ω, 2 knows 1 knows A at ω, and so on. Aumann shows that A is common knowledge at ω if and only if $(H^1 \wedge H^2)(\omega) \subset A$ where $(H^1 \wedge H^2)(\omega)$ is the member of the meet of H^1 and H^2 that contains ω.

As argued in the Introduction, many decision problems naturally call for an infinite state space Ω, in which case null events (in the join) must be allowed. So start again with a measurable space (Ω, F). Individual i (i=1,2) has a sub σ-field F^i of F and a prior P^i. (I consider two individuals. All the results generalize immediately to the case of n individuals.) For each i, fix a version of a regular conditional probability given F^i, that is, a function $Q^i: F \times \Omega \rightarrow [0,1]$ such that:

(1) for each $A \in F$, $Q^i(A, \cdot)$ is a version of $P^i(A \mid F^i)$;

(2) for each $\omega \in \Omega$, $Q^i(\cdot, \omega)$ is a probability measure on F.

(This can be done if, for example, Ω is a complete separable metric space and F is the Borel field on Ω.) This completes the description of the information structure. The elements of this structure -- the state space (Ω, F), the σ-fields F^1, F^2, and the conditional probabilities Q^1, Q^2 -- are assumed to be common knowledge among 1 and 2. In particular, notice that the conditional probabilities Q^1, Q^2 must be specified and hence must be common knowledge -- it is not enough for just the priors P^1, P^2 to be common knowledge. As stated in the Introduction, it is shown in Chapter 4 that this assumption that the information structure is common knowledge is without loss of generality.

Consider an event $A \in F$ and a state of the world $\omega \in \Omega$. I define "i knows A at ω " to mean $Q^i(A, \omega) = 1$. The event that i knows A, to be denoted $K^i(A)$, is then the set of ω's such that i knows A at ω:

$$K^i(A) = \{\omega \mid Q^i(A, \omega) = 1\}.$$

$K^i(\cdot)$ is a function from F to F^i. The following properties of $K^i(\cdot)$ show that this function captures some aspects of one's intuitive notion of what "to know" means. (The proofs are straightforward and are omitted.)

(P1) For any $A \in F$, $K^i(A) \in F^i$.

(P2) For any $A, B \in F$, if $A \subset B$, $[i]$, then $K^i(A) \subset K^i(B)$.

(P3) For any $A_1, A_2, \ldots \in F$, $K^i(\cap_{n=1}^{\infty} A_n) = \cap_{n=1}^{\infty} K^i(A_n)$.

"$A \subset B$, $[i]$" means that for every $\omega \in \Omega$, $Q^i(A - B, \omega) = 0$. So if $A \subset B, [i]$, then i's posterior belief at every state of the world ω is that B happens whenever A happens. (P2) says that in this case if i knows A then i knows B. Note that $A \subset B$, [i] implies that $P^i(A - B) = 0$ (which says that i's prior belief is that B happens whenever A happens), but not conversely. (P2) has the form of most of the subsequent results in this chapter in that it uses the conditionals Q^i and not the prior P^i. (P3) says that i knows A_1 and A_2 and ... if and only if i knows A_1 and i knows A_2 and ... .

Now consider the event: 1 knows A, 1 knows 2 knows A, 1 knows 2 knows 1 knows A, and so on. Call this event L^1A. Formally:

$$L^1A = K^1A \cap K^1K^2A \cap K^1K^2K^1A \cap \dots .$$

(Note that by (P1) all sets of the type $K^1K^2...A$ lie in F^1. Therefore L^1A, being a countable intersection of such sets, also lies in F^1.) Let L^2A denote the corresponding event: 2 knows A, 2 knows 1 knows A, and so on.

DEFINITION 2.1: An event $A \in F$ is common knowledge at a state of the world $\omega \in \Omega$ if $\omega \in L^1A \cap L^2A$.

Definition 2.1 formalizes the notion of common knowledge using 1 and 2's beliefs, i.e. their posteriors Q^1, Q^2, as Bayesian decision theory dictates. An "informational" approach would suggest that common knowledge can be defined using the σ-fields F^1, F^2, i.e. using 1 and 2's private information. I now want to relate Definition 2.1 to an "informational" definition. To see the first issue which arises, consider Figure 1. The dotted ovals belong to 1's partition H^1. The heavy line ovals belong to 2's partition H^2. P^1 assigns

positive probability to $\{\omega_1\}, \{\omega_2\}$, and $\{\omega_5\}$, while $P^1(\{\omega_3, \omega_4\}) = 0$. If 1 observes $\{\omega_3, \omega_4\}$ then 1 "ignores" this information so that the conditional probability $P^1(\cdot \mid \{\omega_3, \omega_4\})$ is equal to $P^1(\cdot)$. P^2 assigns positive probability to every ω. Consider the state ω_1. It looks like A should be common knowledge at ω_1 since the member of the meet of H^1 and H^2 that contains ω_1 is contained in A. But A is not common knowledge at ω_1 in the sense of Definition 2.1. To see this, note that whereas $\omega_1 \in K^1A$ and $\omega_1 \in K^2A$, $\omega_1 \notin K^2K^1A$ since after observing $\{\omega_3, \omega_4\}$ 1 does not assign probability one to A.

The example uses a finite Ω. It would seem that with a finite Ω the difficulty above only arises if the priors P^1, P^2 are not mutually absolutely continuous (i.e. have different null events), since otherwise one can simply throw out all the null events. (In fact, throwing them out would be the "wrong" procedure. Adding them in would be better -- see the discussion of completion below.) One can view the example in two ways: first, the possibility of non-mutually absolutely continuous priors is not excluded; second, it is illustrative of the difficulties which arise in the infinite case.

In the example, if 1 is informed of $\{\omega_3, \omega_4\}$, (s)he ignores this information. To rule out this somewhat implausible situation 1 should assign (posterior) probability one to *any* partition cell, even if that cell has P^i-probability zero. Another justification for imposing this restriction is that if it does not hold, 1 may not know his/her own beliefs. For example, suppose that when informed of $\{\omega_3, \omega_4\}$, 1 assigns posterior probability 0.5 to $\{\omega_1, \omega_2\}$ (and hence to the belief $P^1(\cdot \mid \{\omega_1, \omega_2\})$) and posterior probability 0.5 to $\{\omega_5\}$ (and hence to the belief $P^1(\cdot \mid \{\omega_5\})$). The general version of the restriction on conditional probabilities which is needed is called properness (see [2], [3]).

DEFINITION 2.2: Q^i is proper if for each $\omega \in \Omega$, $Q^i(B, \omega) = 1_B(\omega)$ for every $B \in F^i$.

Rather surprisingly, the standard assumptions -- (Ω, F) Borel, F^i countably generated -- do not suffice to guarantee the existence of a proper regular conditional probability. (Theorem 1 in [3] provides a necessary and sufficient condition for the existence of a proper version.) Nevertheless, for the purposes of this chapter there is no loss of generality in assuming that Q^i is proper. This is because on the expanded state space in Chapter 4, properness is automatically satisfied provided the underlying state space is complete separable metric (see Chapter 4, Section 5). The assumption of properness is also made in the literature on (extensive form) refinements of Nash equilibrium. It is implicit in the intuition behind subgame perfection and in the definition of a sequential equilibrium (Kreps and Wilson [6]). It is explicit in Myerson's definition of a conditional probability system ([8, p.21]). Given properness, the function $K^i(\cdot)$ defined earlier can be shown to satisfy the following properties in addition to (P1)-(P3).

(P4) For any $A \in F$, $K^iA \subset A$, $[i]$.

(P5) For any $B \in F^i$, $K^iB = B$.

Under the assumption that Q^1, Q^2 are proper I can now prove a one-way implication between the definition of common knowledge in terms of beliefs (Definition 2.1) and an "informational" definition in terms of the σ-fields F^1 and F^2.

PROPOSITION 2.1: Suppose there is a set B in the meet $F^1 \wedge F^2$ with $\omega \in B$ and $B \subset A$, $[i]$, $i=1,2$. Then A is common knowledge at ω.

PROOF: If $B \subset A$, $[1]$, then $K^1 A \subset K^1 A$ by (P2). But by (P5) $B \subset K^1 B$, so $B \subset K^1 A$. Hence $\omega \in K^1 A$. Similarly, $K^2 B \subset K^2 A$ by (P2) and $B \subset K^2 B$ by (P5). So $B \subset K^2 A$ and thus $B \subset K^1 A \subset K^1 K^2 A$ by (P5) and (P2). Hence $\omega \in K^1 K^2 A$. Continuing in this way shows that $\omega \in L^1 A$. A similar argument shows that $\omega \in L^2 A$. □

In the next section I start with an example to demonstrate that the converse to Proposition 2.1 is false. I go on to show how to complete 1 and 2's σ-fields in order to obtain an equivalence between Definition 2.1 and an "informational" definition.

3. Common Knowledge in Terms of σ-fields

I start this section by asking whether the converse to Proposition 2.1 holds. That is, is it true that if A is common knowledge at ω then there is a set $B \in F^1 \wedge F^2$ with $\omega \in B$ and $B \subset A$, $[i]$, $i=1,2$? The answer is no, as the following example shows. Consider Figure 2 where 1's (resp. 2's) information partition consists of the three vertical (resp. horizontal) strips. 1 and 2 have the same prior P on Ω. P assigns probability zero to $\{\omega_k\}$, $k = 2, 4, 6, 8$, and positive probability elsewhere. $K^1A = \{\omega_2, \omega_5, \omega_8\}$. $K^2A = \{\omega_4, \omega_5, \omega_6\}$. So $K^1K^2A = \{\omega_2, \omega_5, \omega_8\} = K^1A$. Continuing in this way, one can see that $L^1A = \{\omega_2, \omega_5, \omega_8\}$. Similarly, $L^2A = \{\omega_4, \omega_5, \omega_6\}$. So $\omega_5 \in L^1A \cap L^2A$, i.e. A is common knowledge at ω_5 in the sense of Definition 2.1. On the other hand, the meet of 1 and 2's information partitions is just the trivial partition $\{\Omega\}$. Hence this example shows that the converse to Proposition 2.1 is false since although $\omega_5 \in \Omega$, it is not true that $\Omega \subset A$, $[i]$, $i=1,2$.

It looks like the way to deal with the problem in the example is to throw out the null events $\{\omega_k\}$, $k = 2, 4, 6, 8$. But clearly this cannot be the right intuition for a general (infinite) Ω. Instead let's do the opposite -- add in the null events to i's partition. That is, add to i's partition all the events which a priori i knows will not happen. So 1's refined partition is $\{\{\omega_1, \omega_7\}, \{\omega_4\}, \{\omega_2\}, \{\omega_5\}, \{\omega_8\}, \{\omega_3, \omega_9\}, \{\omega_6\}\}$. Similarly, 2's refined partition is $\{\{\omega_1, \omega_3\}, \{\omega_2\}, \{\omega_4\}, \{\omega_5\}, \{\omega_6\}, \{\omega_7, \omega_9\}, \{\omega_8\}\}$. Now there is a set B in the meet of 1 and 2's partitions such that $\omega_5 \in B$ and $B \subset A$, $[i]$, $i=1,2$, namely the set $\{\omega_5\}$. The general version of this procedure is to complete i's sub σ-field [4, p. 31 Exercise 20].

DEFINITION 3.1: The completion of F^i is the σ-field $\hat{F}^i$ generated by F^i and the class of sets

$$\hat{N}^i = \{G \in F \mid P^i(G) = 0\}.$$

PROPOSITION 3.1: Suppose A is common knowledge at ω. Then there is a set B in the meet $\hat{F}^1 \wedge \hat{F}^2$ of the completed σ-fields such that $\omega \in B$ and $B \subset A$, $[i]$, i=1,2.

The proof of Proposition 3.1, which is a partial converse to Proposition 2.1, is omitted since it is a trivial corollary to Proposition 3.2 proved below.

Our goal is to prove an equivalence result between Definition 2.1 and a definition of common knowledge in terms of F^1 and F^2. But Propositions 2.1 and 3.1 are not strict converses. There are two reasons for this. To see the first difficulty, consider the following variation on the previous example. The information structure is unchanged except that now 1's prior P^1 assigns probability zero to $\{\omega_k\}$, $k = 2, 4, 5, 6, 8$, and positive probability elsewhere. Since the middle vertical strip $\{\omega_2, \omega_5, \omega_8\}$ is now P^1-null, 1's posterior probability $P^1(\cdot \mid \{\omega_2, \omega_5, \omega_8\})$ is no longer determined by Bayes' rule. Set $P^1(\{\omega_2\} \mid \{\omega_2, \omega_5, \omega_8\}) = P^1(\{\omega_5\} \mid \{\omega_2, \omega_5, \omega_8\}) = 0.5$. There is still a set B in the meet of 1 and 2's completed partitions such that $\omega_5 \in B$ and $B \subset A$, $[i]$, i=1,2, namely the set $\{\omega_5\}$ as before. But now there is no ω at which 1 knows A (i.e. assigns posterior probability one to A), so clearly A is not common knowledge at ω_5 (or any other ω) in the sense of Definition 2.1. What has gone wrong here is that too many events have been added to 1's partition. Only those events to which 1 assigns posterior probability zero at *every* state of the world should have been added. In this case, 1's refined partition would include $\{\omega_2, \omega_5\}$ but not $\{\omega_2\}$ and $\{\omega_5\}$ separately. I call the general version of this procedure "posterior completion."

DEFINITION 3.2: The posterior completion of F^i is the σ-field F^i generated by F^i and the class of sets

$$\bar{N}^i = \{G \in F \mid Q^i(G, \omega) = 0 \text{ for every } \omega \in \Omega\}.$$

Clearly, $\bar{N}^i \subset \hat{N}^i$ and so $F^i \subset \hat{F}^i$. Hence the posterior completion of F^i is coarser than the completion of F^i used in probability theory (Definition 3.1). By a standard method of argument on completion [4, p.31 Exercise 20], F^i can be written as

$$F^i = \{G \in F \mid G \,\Delta\, B \in \bar{N}^i \text{ for some } B \in F^i\}$$

where $G \,\Delta\, B = (G - B) \cup (B - G)$. Lemma 3.1 provides an alternative characterization of F^i which will be more convenient. This second characterization relies on Q^i being proper, which the first did not.

LEMMA 3.1: $F^i = \{G \in F \mid \text{for every } \omega \in \Omega,\ Q^i(G, \omega) = 0 \text{ or } 1\}$.

PROOF: First suppose that $G \,\Delta\, B \in \bar{N}^i$ for some $B \in F^i$. For any ω, $Q^i(G, \omega) = Q^i(G - B, \omega) + Q^i(G \cap B, \omega)$. But $G - B \subset G \,\Delta\, B$ so

$G \Delta B \in \mathcal{N}^i$ implies $Q^i(G-B, \omega) = 0$. That is, $Q^i(G, \omega) = Q^i(G \cap B, \omega)$. Furthermore, $Q^i(G \cap B, \omega) = Q^i(B, \omega) - Q^i(B-G, \omega)$. And $Q^i(B-G, \omega) = 0$ since $G \Delta B \in \mathcal{N}^i$, $Q^i(B, \omega) = 1_B(\omega)$ since Q^i is proper. So $Q^i(G, \omega) = 1_B(\omega)$, i.e. $Q^i(G, \omega) = 1$ or 0 according as ω does or does not lie in B.

Conversely, suppose for every ω, $Q^i(G, \omega) = 0$ or 1. Set $B = K^i G$. Thus for every ω, $Q^i(B-G, \omega) = 0$ by (P4). But $\Omega - B = K^i(\Omega - G)$ so $G - B = G \cap (\Omega - B) = G \cap K^i(\Omega - G) = K^i(\Omega - G) - (\Omega - G)$. That is, for every ω, $Q^i(G-B, \omega) = Q^i[K^i(\Omega-G) - (\Omega - G), \omega] = 0$ by (P4). So I have shown that for every ω, $Q^i(B-G, \omega) = Q^i(G-B, \omega) = 0$. That is, $G \Delta B \in \mathcal{N}^i$. □

Lemma 3.1 says that i's posterior completed σ-field F^i contains all the events G in the underlying σ-field F such that, whatever state of the world occurs, i knows either G or the complement of G. As stated above, F^i is coarser than $\hat{F}^i$. Hence the following proposition is an improvement on Proposition 3.1.

PROPOSITION 3.2: Suppose A is common knowledge at ω. Then there is a set B in the meet $F^1 \wedge F^2$ of the posterior completed σ-fields such that $\omega \in B$ and $B \subset A$, [i], i=1,2.

PROOF: Set $B = L^1A \cap L^2A$. Clearly $L^1A \cap L^2A \subset L^1A \subset K^1A$, and (P4) says that $K^1A \subset A$, [1]. So $L^1A \cap L^2A \subset A$, [1]. I now want to show that $L^1A \cap L^2A \in F^1$, i.e. that for every ω, $Q^1(L^1A \cap L^2A, \omega) = 0$ or 1 (using the characterization in Lemma 3.1). By definition if $\omega \in K^1(L^1A \cap L^2A)$,

$Q^1(L^1A \cap L^2A, \omega) = 1$. So to prove that $L^1A \cap L^2A \in F^1$, it will be enough to show that if $\omega \in \Omega - K^1(L^1A \cap L^2A)$, then $Q^1(L^1A \cap L^2A, \omega) = 0$. Now $K^1(A \cap L^2A \cap L^1A) = K^1(A \cap L^2A) \cap K^1L^1A = L^1A$ by (P3) and the definitions of L^1A, L^2A. But $K^1(A \cap L^2A \cap L^1A) \subset K^1(L^1A \cap L^2A)$ by (P2). Hence $L^1A \cap L^2A \subset K^1(L^1A \cap L^2A)$. If $\omega \in \Omega - K^1(L^1A \cap L^2A)$, then $Q^1[\Omega - K^1(L^1A \cap L^2A), \omega] = 1$ by properness. That is, if $\omega \in \Omega - K^1(L^1A \cap L^2A)$, then $Q^1(L^1A \cap L^2A, \omega) = 0$ as required. Similar arguments establish that $L^1A \cap L^2A \subset A$, [2], and $L^1A \cap L^2A \in F^2$. □

There is one more difficulty to overcome before I can state the equivalence result. The remaining problem is that when the σ-fields are completed one gets events in the meet which are believed never to happen. To see this, refer back to the example in Figure 2. $\{\omega_4\}$ is a member of both 1 and 2's posterior completed σ-fields, hence it lies in the meet. But even if ω_4 happens, 1 and 2 both assign posterior probability zero to $\{\omega_4\}$. So certainly $\{\omega_4\}$ is not common knowledge at any state of the world in the sense of Definition 2.1. In order to rule out such situations one should consider only "nonnull" members of the meet. (I say an event $G \in F$ is nonnull at ω if $Q^i(G, \omega) > 0$, $i=1,2$.)

PROPOSITION 3.3: A is common knowledge at ω if and only if there is a set $B \in F^1 \wedge F^2$ which is nonnull at ω with $\omega \in B$ and $B \subset A$, [i], $i=1,2$.

PROOF: *Only if:* Set $B = L^1A \cap L^2A$ and proceed as in the proof of Proposition 3.2. The only additional step is to show that B is nonnull at ω. But this follows immediately from $L^1A \cap L^2A \subset K^i(L^1A \cap L^2A)$, i=1,2, which was shown in the course of that proof.

If: Suppose $\omega \in B$ and $B \subset A$, $[i]$, i=1,2, where B is a member of the meet which is nonnull at ω. Let $B' = B \cap K^iB$. I claim that $\omega \in B' \subset K^iB'$. To see this, first note from Lemma 3.1 that if B is nonnull at ω, then $\omega \in K^iB$. Hence $\omega \in B'$. Second, $K^iB' = K^iB$ by (P1), (P3) and (P5). So $B' \subset K^iB'$. The proof now follows exactly the lines of the proof of Proposition 2.1 -- replacing the set B there with the set B'. □

Note that the definitions and results in this chapter (with the exception of Definition 3.1 and Proposition 3.1 which were later replaced with Definition 3.2 and Proposition 3.2) depend only on the conditionals Q^i and not the priors P^i. In particular, the definition of common knowledge (Definition 2.1) and the completion of the σ-fields (Definition 3.2) must both be stated in terms of the conditionals in order to obtain the main equivalence result Proposition 3.3.

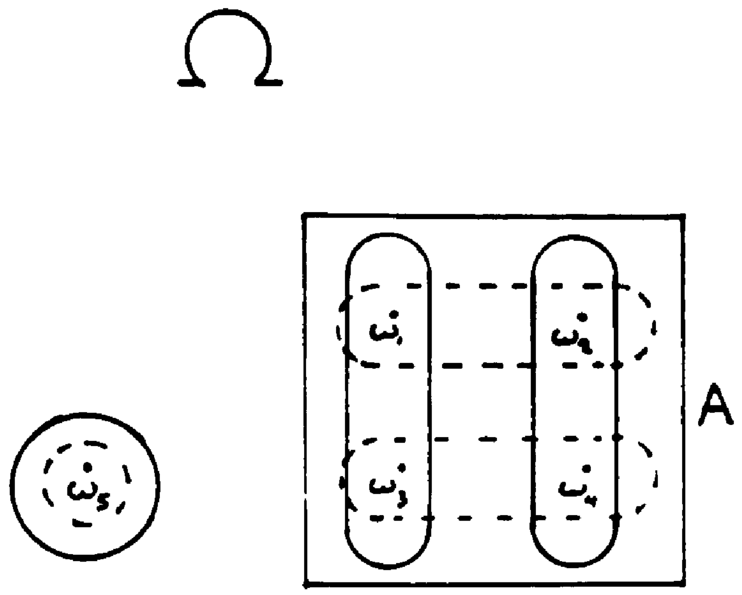

Figure 1

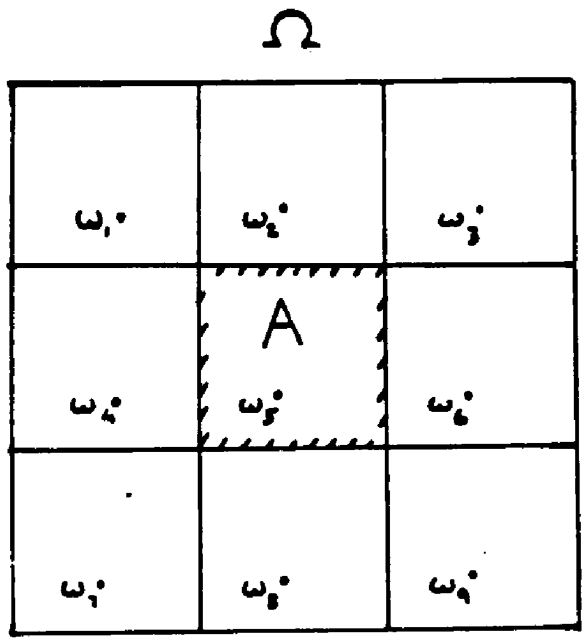

Figure 2

References

[1] R. Aumann: "Agreeing to Disagree," *The Annals of Statistics,* Vol. 4 (1976), 1236-1239.

[2] D. Blackwell and L. Dubins: "On Existence and Non-existence of Proper, Regular, Conditional Distributions," *The Annals of Probability,* Vol. 3 (1975), 741-752.

[3] D.Blackwell and C. Ryll-Nardzewski: "Non-existence of Everywhere Proper Conditional Distributions," *Annals of Mathematical Statistics,* Vol.34 (1963), 223-225.

[4] K. L. Chung: *A Course in Probability Theory.* New York: Academic Press, Second Edition 1974.

[5] J. Harsanyi: "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science,* Vol. 14 (1967-68), 159-189, 320-334, 486-502.

[6] D. Kreps and R. Wilson: "Sequential Equilibria," *Econometrica,* Vol. 50 (1982), 863-894.

[7] J.-F. Mertens and S. Zamir: "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory,* Vol. 14 (1985), 1-29.

[8] R. Myerson: "Sequential Correlated Equilibria of Multistage Games," Research Paper 590, J.L. Kellogg Graduate School of Management, Northwestern University, April 1984.

[9] L. Nielson: "Common Knowledge, Communication, and Convergence of Beliefs," *Mathematical Social Sciences,* Vol. 8 (1984), 1-14.

CHAPTER 4

HIERARCHIES OF BELIEFS AND COMMON KNOWLEDGE

1. Introduction

Consider an n-person decision problem in which each individual faces some (common) space of uncertainty S. In order to determine his/her optimal decision, each individual must have a belief (probability measure) over the space S. But if other individuals' beliefs affect their decisions, then each individual must also have beliefs over everyone else's beliefs over S. This argument leads to each individual having an infinite hierarchy of beliefs (over S, over others' beliefs over S, and so on). Thus each individual is characterized by an infinite hierarchy of beliefs, which is called his/her *type* (cf. Harsanyi [5]).

The first question asked in this chapter is: when does a hierarchy of beliefs of individual i determine a belief over the underlying space S and the set of the other individuals' types? Proposition 2.1 says that this happens provided i's type is coherent (defined below). The idea is that any coherent hierarchy of beliefs can be mapped to a probability measure over S and other individuals' types. Furthermore, this map is one-to-one and onto so any belief over S and the others' types comes from a unique coherent hierarchy.

I now provide a definition of coherency. Consider the hierarchy of beliefs of an individual i. One must allow for the possibility that i's beliefs over S

and i's beliefs over other individuals' beliefs over S are correlated. So a complete specification of i's hierarchy of beliefs must include with i's beliefs over S, i's *joint* beliefs over S and others' beliefs over S, and so on. To say that i's hierarchy of beliefs is coherent means that the marginal on S of i's joint belief on S and others' beliefs on S is equal to i's belief on S, with similar restrictions on i's higher order beliefs.

Of course, each individual only knows his/her own type and not the types of the others. So it seems that one needs a "second level" hierarchy of beliefs -- each individual must also have beliefs over the types of the others, over others' beliefs over types, and so on. As stated above, Proposition 2.1 shows that a coherent first level hierarchy of beliefs, i.e. a coherent type, of individual i determines i's beliefs over others' types. But a coherent type of i does not necessarily determine i's beliefs over others' beliefs over types. This happens if i believes some individual j may be of a type which is not coherent -- for then i thinks it possible that j's type does not determine a belief over S and others' types. A product of type sets $\prod_{i=1}^{n} T^i$ is *belief-closed* if for every i, T^i is a set of coherent types of individual i and each type in T^i assigns probability 1 to the others' types lying in $\prod_{j\neq i} T^j$. If i's type lies in T^i then i knows that the others are of coherent types, so now i's type does determine i's beliefs over others' beliefs over types. In fact, it is clear that in a belief-closed set common knowledge of coherency is satisfied, so that all of i's beliefs in the "second level" hierarchy are determined by i's type. It is in this sense that the model of beliefs is "closed" when one looks at belief-closed sets. In Section 3 it is shown that a union of belief-closed sets is belief-closed, so that there is a largest (in terms of set inclusion) belief-closed set. Proposition 3.1 shows that this is the belief-closed set satisfying common knowledge of coherency (and nothing more). It is also equivalent to the universal type space constructed in Mertens and Zamir [6, Theorem 2.9].

In Chapter 3 a generalization of Aumann's definition of common knowledge was presented. There, as in [1], the information structure on S was assumed to be common knowledge in an informal sense. I say "in an informal sense" because the information structure is not an event in S and of course the mathematical definition of common knowledge in Chapter 3 and in [1] applies only to events in S. In [1] Aumann argued that if the assumption that the information structure is common knowledge is not satisfied, then the state space S should be expanded. In Section 5 of this paper it is shown that the appropriate expanded state space is $S \times T \times T$ where $T \times T$ is the belief-closed set satisfying (just) common knowledge of coherency. Proposition 5.1 shows that if common knowledge is defined on this expanded space as in Chapter 3, then the assumption that the information structure is common knowledge entails no loss of generality. In effect, the assumption that the information structure is common knowledge is replaced with the assumption of common knowledge of coherency.

The problem of constructing hierarchies of beliefs has also been studied in Boge and Eisele [3] and Mertens and Zamir [6]. The former are specifically interested in solving games with incomplete information. Mertens and Zamir implicitly incorporate the assumption of common knowledge of coherency within their construction of hierarchies of beliefs. The approach in this chapter distinguishes between: (1) showing when a hierarchy of beliefs determines a belief over S and others' types (which only requires coherency); and (2) closing the model using belief-closed sets (which satisfy common knowledge of coherency). Making common knowledge of coherency an explicit mathematical assumption has certain advantages. First, as stated above, common knowledge of coherency plays a central role in understanding the assumption of common knowledge of an information structure. Second, this formulation uses somewhat simpler mathematical arguments than those in [6]. (At a technical level, the assumption in [6] that S is compact is replaced with the assumption that S is a complete separable metric space.)

2. Types and Coherency

In this section the hierarchies of beliefs are constructed. I define the type of an individual, and show when a hierarchy of beliefs determines a belief over S and other individuals' types.

Consider two individuals i and j who face some common space of uncertainty S. (I consider the two-person case. The generalization to n individuals is straightforward.) I assume that S is a complete separable metric (Polish) space. According to Bayesian decision theory, i and j must each have a belief (probability measure) over the space S. For any metric space Z let $\Delta(Z)$ denote the space of probability measures on the Borel field of Z, endowed with the weak topology. i and j's first-order beliefs (beliefs over S) are then elements of $\Delta(S)$. But since i does not know j's beliefs over S, i must also have a (second-order) belief over S and j's possible first-order beliefs. So i's second-order belief is an element of $\Delta[S \times \Delta(S)]$. Similarly for j. Formally, define the spaces:

$$
\begin{aligned}
X_0 &= S \\
X_1 &= X_0 \times \Delta(X_0) \\
X_2 &= X_1 \times \Delta(X_1) = X_0 \times \Delta(X_0) \times \Delta(X_1) \\
&\vdots \\
X_n &= X_{n-1} \times \Delta(X_{n-1}) = X_0 \times \Delta(X_0) \times \Delta(X_1) \times \ldots \times \Delta(X_{n-1}) \\
&\vdots
\end{aligned}
$$

A hierarchy of beliefs for i is an element $(\delta_1^i, \delta_2^i, \delta_3^i, \ldots) \in \Delta(X_0) \times \Delta(X_1) \times \Delta(X_2) \times \ldots$, where δ_1^i is i's belief over S; δ_2^i is i's joint belief over S and j's first-order belief; δ_3^i is i's joint belief over S, j's first-order belief, and j's second-order belief; and so on. Similarly, a hierarchy of beliefs for j is an element $(\delta_1^j, \delta_2^j, \ldots) \in \Delta(X_0) \times \Delta(X_1) \times \ldots$, where δ_1^j is j's belief over S; δ_2^j is j's joint belief over S and i's first-order belief; and so on.

A *type* t^i of individual i is just a hierarchy of beliefs $t^i = (\delta_1^i, \delta_2^i, \ldots) \in \Delta(X_0) \times \Delta(X_1) \times \ldots$. Similarly for individual j. Let $T_0 = \Delta(X_0) \times \Delta(X_1) \times \ldots$ be the space of all possible types of individual i or j.

The question asked in this section is: when does i's hierarchy of beliefs determine a belief over the underlying space S and the space of j's possible types?

DEFINITION 2.1: A type $t = (\delta_1, \delta_2, \ldots) \in T_0$ is coherent if for every $n \geq 2$, $\text{marg}_{X_{n-2}} \delta_n = \delta_{n-1}$, where $\text{marg}_{X_{n-2}} \delta_n$ is the marginal on X_{n-2} of the probability measure δ_n.

Recall that δ_n is a measure on the space $X_{n-1} = X_{n-2} \times \Delta(X_{n-2})$, while δ_{n-1} is a measure on X_{n-2}. Coherency requires that the two measures agree on X_{n-2}. Notice that what I've called a coherent sequence of probability measures is called a consistent sequence in the theory of stochastic processes. (I use the term coherent to avoid confusion with Harsanyi's use of the term consistent, which means something different. The examples in Section 4 should help clarify the distinction.) This analogy with the theory of stochastic processes will be used to prove the result of this section. Let T_1 be the set of all coherent types.

PROPOSITION 2.1: There is a homeomorphism $f: T_1 \rightarrow \Delta(S \times T_0)$.

Proposition 2.1 will be an easy consequence of Kolmogorov's Existence Theorem. Kolmogorov's theorem says, roughly speaking, th[illegible] [illegible]iven any coherent sequence of probability measures there is a unique measure on the infinite product space (which is $S \times T_0$ in the present context) with the prescribed marginals. In other words, Kolmogorov's theorem tells us that associated with the hierarchy of beliefs of any coherent type $(\delta_1, \delta_2, \ldots) \in T_1$ is a unique measure $\delta \in \Delta(S \times T_0)$ such that the marginal of δ on each X_{n-1} is exactly δ_n. On the other hand, any measure $\delta \in \Delta(S \times T_0)$ defines a coherent type $(\delta_1, \delta_2, \ldots) \in T_1$ (just set $\delta_n = \mathrm{marg}_{X_{n-1}} \delta$), and by uniqueness this type must be associated with δ. Hence the existence of an isomorphism $f: T_1 \rightarrow \Delta(S \times T_0)$ is essentially just a statement of Kolmogorov's theorem. That f is in fact a homeomorphism is a technical result. Formally, Proposition 2.1 is a consequence of the following lemma:

LEMMA 2.1: Suppose $\{Z_n\}_{n=0}^{\infty}$ is a collection of Polish spaces, and let

$$D = \{(\delta_1, \delta_2, \ldots) \mid \delta_n \in \Delta(Z_0 \times \ldots \times Z_{n-1}), \mathrm{marg}_{Z_0 \times \ldots \times Z_{n-2}} \delta_n = \delta_{n-1} \ \forall\, n\}.$$

Then there is a homeomorphism $f: D \rightarrow \Delta(\prod_{n=0}^{\infty} Z_n)$.

PROOF: (1) Consider any sequence $(\delta_1, \delta_2, \ldots) \in D$. By a generalized version of Kolmogorov's Existence Theorem ([4, p.68]), there is a unique measure

$\delta \in \Delta(\prod_{n=0}^{\infty} Z_n)$ such that $\text{marg}_{Z_0 \times \ldots \times Z_{n-1}} \delta = \delta_n$ for every n.

(2) It follows immediately that there is an isomorphism $f: D \rightarrow \Delta(\prod_{n=0}^{\infty} Z_n)$, where f maps any sequence $(\delta_1, \delta_2, \ldots) \in D$ to the unique $\delta \in \Delta(\prod_{n=0}^{\infty} Z_n)$ whose existence is guaranteed by (1). (Recall the discussion in the paragraph preceding this lemma.)

(3) For any $\delta \in \Delta(\prod_{n=0}^{\infty} Z_n)$, $f^{-1}(\delta) = (\text{marg}_{Z_0} \delta, \text{marg}_{Z_0 \times Z_1} \delta, \ldots)$. So f^{-1} is continuous since the mappings of δ into $\text{marg}_{Z_0 \times \ldots \times Z_{n-1}} \delta$ are all continuous. To see that f is continuous, consider a sequence $(\delta_1^r, \delta_2^r, \ldots) \in D$ converging to $(\delta_1, \delta_2, \ldots) \in D$, i.e. $\delta_n^r \Rightarrow \delta_n$ for every n. Let $\delta^r = f(\delta_1^r, \delta_2^r, \ldots)$, $\delta = (\delta_1, \delta_2, \ldots)$. I have to show that $\delta^r \Rightarrow \delta$. Now the cylinder sets form a convergence-determining class, that is, if $\delta^r(C) \rightarrow \delta(C)$ for every cylinder set C such that $\delta(\partial C) = 0$, then $\delta^r \Rightarrow \delta$ (see [2, p.22 Problem 7]). Since the values of δ^r, δ on the cylinder sets are given by the δ_n^r's, δ_n's respectively, it is clear that $\delta_n^r \Rightarrow \delta_n$ for every n implies that $\delta^r \Rightarrow \delta$. So f is continuous. □

To see why Proposition 2.1 follows from Lemma 2.1, set

$$Z_0 = X_0$$
$$Z_n = \Delta(X_{n-1}) \text{ for } n \geq 1$$

so $Z_0 \times \ldots \times Z_n = X_n$ and $\prod_{n=0}^{\infty} Z_n = S \times T_0$. If S is a Polish space then so is $\Delta(S)$ [4, p.73]. Hence the Z_n's will be Polish spaces provided S is. A coherent type $t \in T_1$ is just a sequence of probability measures $(\delta_1, \delta_2, \ldots) \in D$. So Lemma 2.1 implies that there is a homeomorphism between the set of coherent types T_1 and $\Delta(S \times T_0)$.

3. Common Knowledge of Coherency and Belief-Closed Subsets

In this section I define belief-closed subsets. It is shown that a union of belief-closed subsets is belief-closed, so that there is a largest (in terms of set inclusion) belief-closed subset. Proposition 3.1 shows that this set is equal to the belief-closed subset which satisfies just common knowledge of coherency.

DEFINITION 3.1: (cf. [6, Definition 2.15]) A subset $T^i \times T^j \subset T_1 \times T_1$ is belief-closed if

$$\text{for every } t^i \in T^i,\ f(t^i)(S \times T^j) = 1,$$
$$\text{for every } t^j \in T^j,\ f(t^j)(S \times T^i) = 1.$$

Definition 3.1 says that a subset $T^i \times T^j$ is belief-closed if every type t^i in T^i of individual i believes that j can only be of a type t^j in T^j, and similarly for individual j. Clearly, if $T^i \times T^j$ is belief-closed, then every type t^i in T^i of individual i also believes that j believes that i can only be of a type in T^i, and so on. Hence if every type in T^i or T^j satisfies a certain property (such as assigning certain probabilities to events in the underlying space of uncertainty S), then this property is common knowledge. In particular, it follows that any belief-closed set $T^i \times T^j$ satisfies common knowledge of coherency (since every type t^i in T^i or t^j in T^j is in T_1 and therefore coherent).

If the individuals' types t^i, t^j are in a belief-closed subset $T^i \times T^j$ then their types fully specify their beliefs. Of course a coherent type, say t^i, induces a belief over j's types. But if $T^i \times T^j$ is belief-closed, then i's beliefs over j's beliefs over i's type can also be calculated -- by examining the associated beliefs $\mathrm{marg}_{T^i} f(t^j)$ of every t^j which i considers possible. Similarly i's beliefs over j's beliefs over ... i's type can be calculated. These beliefs will induce a "second level" hierarchy, where the set of "second level" types thus calculated from all types in $T^i \times T^j$ will be belief-closed. And then one could calculate "third level" hierarchies, and so on. Clearly all beliefs are specified. It is in this sense that the model is closed by looking at belief-closed subsets.

The following properties should be noted: (1) if $T^i \times T^j$ is belief-closed, then so is $T^j \times T^i$; (2) if $T^i_\gamma \times T^j_\gamma$, $\gamma \in \Gamma$, is a collection of belief-closed sets, then $\bigcup_{\gamma\in\Gamma} T^i_\gamma \times \bigcup_{\gamma\in\Gamma} T^j_\gamma$ is "almost" belief-closed in the sense that for every $t^i \in \bigcup_{\gamma\in\Gamma} T^i_\gamma$ there is a subset of $S \times \bigcup_{\gamma\in\Gamma} T^j_\gamma$ to which $f(t^i)$ assigns probability 1 (in fact if $t^i \in T^i_\gamma$ it is the subset $S \times T^j_\gamma$), and similarly for j. (The slight qualification is needed since a priori there is no reason for $\bigcup_{\gamma\in\Gamma} T^j_\gamma$ to be a measurable set.) (1) and (2) imply that the component-by-component union of all belief-closed sets is symmetric, i.e. of the form $T^* \times T^*$ for some $T^* \subset T_1$, and is (almost) belief-closed.

I now start from the "other end." Rather than considering belief-closed subsets of $T_1 \times T_1$ and taking unions, I begin with $T_1 \times T_1$ and impose common knowledge of coherency. Formally, the sequence of sets T_k, $k \geq 2$ is defined by:

$$T_k = \{t \in T_1 \mid f(t)(S \times T_{k-1}) = 1\}.$$

(Notice that for T_k to be well-defined, T_{k-1} must be a Borel set. This is proved in Lemma A.1 in the Appendix.) Let $T = \cap_{k=1}^{\infty} T_k$. $T \times T$ is the subset of $T_1 \times T_1$ obtained by requiring all statements of the form "i knows j knows ... i's type is coherent" to be true. To say that i knows j's type is coherent means that if i's type is t^i, then the associated belief $f(t^i)$ assigns probability 1 to types t^j which are coherent. T_2 is the set of all such types for i. The interpretation of other T_k's is similar.

PROPOSITION 3.1: $T^* = T$.

PROOF: It is easy to check that $T = \{t \in T_1 \mid f(t)(S \times T) = 1\}$ so that $T \times T$ is belief-closed. To complete the proof it is enough to show that every belief-closed set $T^i \times T^j$ is a subset of $T \times T$. Without loss of generality a belief-closed set can be taken to be a symmetric set, say $T' \times T'$. (This follows from properties (1) and (2) stated above.) I have to show that $T' \subset T$. $T' \subset T_1$ by definition. And since for every $t \in T'$, $f(t)(S \times T') = 1$, it follows that $T' \subset T_2$. Clearly $T' \subset T_k$ for every k by induction, so $T' \subset \cap_{k=1}^{\infty} T_k = T$. □

$T^* \times T^*$ is the component-by-component union of all belief-closed sets, and $T \times T$ was constructed as the subset of $T_1 \times T_1$ which satisfies common knowledge of coherency. So Proposition 3.1 says that the largest belief-closed set is equal to the set satisfying just common knowledge of coherency.

PROPOSITION 3.2: T is homeomorphic to $\Delta(S \times T)$.

PROOF: From the proof of Proposition 3.1 $T = \{t \in T_1 | f(t)(S \times T) = 1\}$, or $f(T) = \{\delta \in \Delta(S \times T_0) | \delta(S \times T) = 1\}$ since f is onto. But the set on the right-hand side is homeomorphic to $\Delta(S \times T)$ (for any metric space Z and measurable subset W of Z, $\{\delta \in \Delta(Z) | \delta(W) = 1\}$ is homeomorphic to $\Delta(W)$), and $f(T)$ is homeomorphic to T. So T is homeomorphic to $\Delta(S \times T)$. □

I have argued that any belief-closed set closes the model of beliefs. But in Proposition 3.2, T cannot be replaced with any T' where $T' \times T'$ is belief-closed. Proposition 3.2 tells us that any type in T determines a belief over $S \times T$. Likewise, any type in T', where $T' \times T'$ is belief-closed, determine a belief over $S \times T'$. But in general it is not true that *any* belief over $S \times T'$ comes from a type in T'. If T' is not to "miss" any beliefs in this sense, then it must be equal to T (as implied by Proposition 3.2).

The set T is equivalent to the universal type space constructed in Mertens and Zamir [6, Theorem 2.9]. At a technical level, however, the assumption in [6] that S is compact has been replaced with the assumption that S is a Polish space.

4. Coherency and Consistency

In this section some simple examples of hierarchies of beliefs are provided. The examples should also help clarify the distinction between coherency and consistency (in Harsanyi's sense). Roughly speaking, a subset $C \times T^i \times T^j$ of $S \times T \times T$ is consistent if each individual's beliefs are determined by the conditional probability, given his/her private information, derived from some "prior" probability measure on $C \times T^i \times T^j$. A little more formally, $C \times T^i \times T^j$ is consistent if there is a probability measure P on $C \times T^i \times T^j$ such that for every $t^i \in T^i$, $f(t^i)$ is determined by the conditional P-probability given t^i, and similarly for j. Notice that if $C \times T^i \times T^j$ is consistent then $T^i \times T^j$ must be a belief-closed set. But the converse is not true as shown by Example 4. Let a and b be points in S. For any measurable space Z, let $\mu(z)$ denote the Dirac measure at the point $z \in Z$.

1. A coherent pair of hierarchies which constitutes a belief-closed set:

$$\begin{array}{ll}
\delta_1^i = \mu(a) & \delta_1^j = \mu(a) \\
\delta_2^i = \mu(a, \delta_1^j) & \delta_2^j = \mu(a, \delta_1^i) \\
\vdots & \vdots \\
\delta_n^i = \mu(a, \delta_1^j, \delta_2^j, \ldots, \delta_{n-1}^j) & \delta_n^j = \mu(a, \delta_1^i, \delta_2^i, \ldots, \delta_{n-1}^i) \\
\vdots & \vdots
\end{array}$$

This says that i assigns probability 1 to $\{a\}$, i assigns probability 1 to $\{a$ and j assigns probability 1 to $\{a\}\}$, and so on. Similarly for j. Let $t^i = (\delta_1^i, \delta_2^i, \ldots)$ and $t^j = (\delta_1^j, \delta_2^j, \ldots)$. Then i's hierarchy determines the belief $\mu(a, t^j)$ and j's hierarchy determines the belief $\mu(a, t^i)$. $\{(t^i, t^j)\}$ is a belief-closed set, and $\{(a, t^i, t^j)\}$ is consistent.

2. An incoherent hierarchy for j:

$$\hat{\delta}_1^j = \mu(b),$$
$$\hat{\delta}_n^j = \delta_n^j \text{ for } n \geq 2.$$

This says that j assigns probability 1 to $\{b\}$, but also that j assigns probability 1 to $\{a$ and i assigns probability 1 to $\{a\}\}$. Let $\hat{t}^j = (\hat{\delta}_1^j, \hat{\delta}_2^j, \ldots)$.

3. A coherent hierarchy for i which does not satisfy that i knows j is coherent, and hence cannot be part of any belief-closed set:

$$\hat{\delta}_1^i = \mu(a),$$
$$\hat{\delta}_n^i = \mu(a, \hat{\delta}_1^j, \hat{\delta}_2^j, \ldots, \hat{\delta}_{n-1}^j) \text{ for } n \geq 2.$$

This hierarchy determines the belief $\mu(a, \hat{t}^j)$.

4. A coherent pair of hierarchies which constitutes a belief-closed set:

$$\begin{array}{ll} \tilde{\delta}_1^i = \mu(a) & \tilde{\delta}_1^j = \mu(a) \\ \tilde{\delta}_2^i = \mu(a, \delta_1^j) & \tilde{\delta}_2^j = \mu(a, \delta_1^i) \\ \vdots & \vdots \\ \tilde{\delta}_n^i = \mu(a, \delta_1^j, \delta_2^j, \ldots, \delta_{n-1}^j) & \tilde{\delta}_n^j = \mu(a, \delta_1^i, \delta_2^i, \ldots, \delta_{n-1}^i) \\ \vdots & \vdots \end{array}$$

This says that i assigns probability 1 to $\{a\}$, i assigns probability 1 to $\{a$ and j assigns probability 1 to $\{b\}\}$, and so on; while j assigns probability 1 to $\{b\}$, j assigns probability 1 to $\{b$ and i assigns probability 1 to $\{a\}\}$, and so on. Let $\tilde{t}^i = (\tilde{\delta}_1^i, \tilde{\delta}_2^i, \ldots)$ and $\tilde{t}^j = (\tilde{\delta}_1^j, \tilde{\delta}_2^j, \ldots)$. Then the hierarchies determine the beliefs $\mu(a, \tilde{t}^j)$, $\mu(b, \tilde{t}^i)$. $\{(\tilde{t}^i, \tilde{t}^j)\}$ is a belief-closed set, but there is no $s \in S$ such that $(s, \tilde{t}^i, \tilde{t}^j)$ lies in a consistent subset of $S \times T \times T$.

5. Common Knowledge

In Chapter 3 a generalization of Aumann's definition [1] of common knowledge was presented. That definition will be referred to in this section as the "standard" definition. The assumption that the information structure (σ-fields, etc.) is common knowledge in an informal sense was necessary there (as in [1]) in order to justify the interpretation of the results. In particular note that $K^i K^j A$ is, by definition (since $K^j A$ is an event in Ω) the event "i knows $K^j A$." To replace the latter with "i knows j knows A" is implicitly to assume that i knows j's information structure.

This section looks at common knowledge of events E in the underlying state space S of Section 2. The question asked is what can be done if the information structure on S is not common knowledge. Aumann [1] argued that in this case the state space should be expanded. Proposition 5.1 below implies that if one takes the expanded state space to be $S \times T \times T$, and applies the standard definition of common knowledge to this expanded space, then the assumption that the information structure is common knowledge entails no loss of generality. Since the assumption that the information structure is common knowledge is an informal one, I will now clarify the sense in which it is implied by Proposition 5.1.

Even if the information structure on S is not common knowledge, one can easily provide a definition of common knowledge of an event E in S using types. To say E is common knowledge is to impose certain restrictions on i

and j's types, namely to require that i knows E (i.e. i assigns probability 1 to E), i knows j knows E, and so on, and similarly for j. Let $V(E) \subset T_0$ be the set of types of i or j which satisfy these restrictions (by symmetry it is the same set for both i and j). So E is common knowledge if $(t^i, t^j) \in V(E) \times V(E)$. It is important to note that $V(E)$ is a subset of T_0 but not in general of T. That is, it is quite possible for an event E to be common knowledge without common knowledge of coherency being satisfied. For example, if one sets $E = S$, then $V(E) = T_0$. To see this, recall from the definition of T_0 (see Section 2) that for any pair of types $(t^i, t^j) \in T_0 \times T_0$, i knows S, i knows j knows S, and so on, and similarly for j. (In fact, the only case in which common knowledge of E implies common knowledge of coherency is when E is a singleton.)

We have just seen how to define common knowledge of an event E in S in terms of types and without using an information structure. Proposition 5.1 says that if common knowledge of coherency is satisfied, then E is common knowledge in terms of types if and only if E is common knowledge according to the standard definition applied to $S \times T \times T$. (Of course strictly speaking E is not an event in the expanded space $S \times T \times T$, but E is naturally identified with $E \times T \times T$.) More precisely, Proposition 5.1 says that $(t^i, t^j) \in [V(E) \cap T] \times [V(E) \cap T]$ if and only if $(s, t^i, t^j) \in L^i(E \times T \times T) \cap L^j(E \times T \times T)$. The interpretation of $(t^i, t^j) \in V(E) \times V(E)$ is that i knows E, i knows j knows E, and so on, and similarly for j. So

Proposition 5.1 implies that if common knowledge of coherency is satisfied, then this is also the interpretation of $(s, t^i, t^j) \in L^i(E \times T \times T) \cap L^j(E \times T \times T)$. But to make this latter interpretation, one needs the information structure on $S \times T \times T$ to be common knowledge. Hence Proposition 5.1 implies that this assumption entails no loss of generality. In effect, the assumption that the information structure is common knowledge has been replaced with the assumption of common knowledge of coherency.

Notice that even if S is finite, $S \times T \times T$ is an infinite space. So Proposition 5.1 must be stated using the definition provided in Chapter 3 of common knowledge. Of course, to write down this definition one needs an information structure on $S \times T \times T$. Let F denote the Borel field of $S \times T \times T$. Since i knows his/her own type, the natural sub σ-field of F for i is $F^i = \{S \times B \times T \mid B \text{ a Borel subset of } T\}$. Similarly, the natural sub σ-field for j is $F^j = \{S \times T \times B \mid B \text{ a Borel subset of } T\}$. For any event $A \in F$, i's natural conditional probability of A at a state $\omega = (s, t^i, t^j)$ is $Q^i(A, \omega) = f(t^i)(A_{t^i})$ where A_{t^i} is the t^i-section of A. Only the conditional probabilities Q^i are specified since these are in fact all that are needed to define common knowledge as in Chapter 3 – i's prior P^i is irrelevant. Notice that Q^i is proper as required. For any event $A \in F$, j's natural conditional probability of A at $\omega = (s, t^i, t^j)$ is $Q^j(A, \omega) = f(t^j)(A_{t^j})$ where A_{t^j} is the t^j-section of A. Having defined the information structure on $S \times T \times T$, I can now define K^i, K^j and then L^i, L^j as in Chapter 3.

PROPOSITION 5.1: For any event E in S,

$$S \times [V(E) \cap T] \times [V(E) \cap T] = L^i(E \times T \times T) \cap L^j(E \times T \times T).$$

PROOF: $K^i(E \times T \times T) = \{\omega \mid Q^i(E \times T \times T, \omega) = 1\}$

$$= \{(s, t^i, t^j) \mid f(t^i)(E \times T) = 1\}$$
$$= S \times V_1(E) \times T$$

where $V_1(E) = \{t \in T \mid f(t)(E \times T) = 1\}$. Similarly, $K^j(E \times T \times T) = S \times T \times V_1(E)$. So

$$K^iK^j(E \times T \times T) = \{\omega \mid Q^i[K^j(E \times T \times T), \omega] = 1\}$$
$$= \{\omega \mid Q^i[S \times T \times V_1(E), \omega] = 1\}$$
$$= S \times V_2(E) \times T$$

where $V_2(E) = \{t \in T \mid f(t)[S \times V_1(E)] = 1\}$. Continuing in this fashion, for any $k \geq 2$ let $V_k(E) = \{t \in T \mid f(t)[S \times V_{k-1}(E)] = 1\}$. (Notice that for $V_k(E)$ to be well-defined, $V_{k-1}(E)$ must be a Borel subset of T. This is proved in Lemma A.2 in the Appendix.) Then

$$L^i(E \times T \times T) = K^i(E \times T \times T) \cap K^iK^j(E \times T \times T) \cap \ldots$$
$$= S \times \cap_{k=1}^{\infty} V_k(E) \times T.$$

But clearly $\cap_{k=1}^{\infty} V_k(E) = V(E) \cap T$, so $L^i(E \times T \times T) = S \times [V(E) \cap T] \times T$. Similarly, $L^j(E \times T \times T) = S \times T \times [V(E) \cap T]$. Therefore $L^i(E \times T \times T) \cap L^j(E \times T \times T) = S \times [V(E) \cap T] \times [V(E) \cap T]$. □

Appendix

This section deals with two measure-theoretic questions in the text. The first relates to the sequence of sets $\{T_k\}_{k=1}^{\infty}$ defined on p.8. I claimed that each T_k is a Borel set. This follows from:

LEMMA A.1: T_k, $k \geq 1$, is closed.

PROOF: First note that the set D defined in Lemma 2.1 is closed since the mappings taking δ_n into $\mathrm{marg}_{Z_0 \times \ldots \times Z_{n-2}} \delta_n$ are all continuous. From this and the remarks following Lemma 2.1 it follows that T_1 is closed. Now assume inductively that T_{k-1} is closed, and consider a sequence $t_m \rightarrow t$ where $t_m \in T_k \; \forall \, m$. Since f is continuous and $S \times T_{k-1}$ is closed by assumption, $\limsup_m f(t_m)(S \times T_{k-1}) \leq f(t)(S \times T_{k-1})$ (this is criterion (iii) of the Portmanteau Theorem in [2, p.11]). But by assumption $f(t_m)(S \times T_{k-1}) = 1 \; \forall \, m$ so $f(t)(S \times T_{k-1}) = 1$, i.e. $t \in T_k$. So T_k is closed. □

The second measure-theoretic question relates to the sequence of sets $\{V_k(E)\}_{k=1}^{\infty}$ defined in the proof of Proposition 5.1.

LEMMA A.2: $V_k(E)$, $k \geq 1$, is a Borel set.

PROOF: I first show that for any $A \in F$ (F is the Borel field of $S \times T \times T$) $f(t^j)(A_{t^j})$ is measurable as a function of t^j. (For notational reasons it will be easier to prove the result for j.) If $A = A' \times A''$ where $A' \subset S \times T$ and $A'' \subset T$ then $f(t^j)(A_{t^j}) = 1_{A''}(t^j) \cdot f(t^j)(A')$. Since f is continuous $f(t^j)(A')$ is lower semicontinuous if A' is open (this is criterion (iv) of the Portmanteau Theorem in [2, p.11]), and clearly $1_{A''}(t^j)$ is lower semicontinuous if A'' is open. So $f(t^j)(A_{t^j})$ is lower semicontinuous -- hence measurable -- on the π-system of open rectangles $A = A' \times A''$ which generates F. But it is straightforward to check that the class of sets A on which $f(t^j)(A_{t^j})$ is measurable must be a λ-system, hence by the π-λ Theorem, $f(t^j)(A_{t^j})$ is measurable for all $A \in \mathcal{F}$. It follows that $V_1(E) = \{t^j \in T \mid f(t^j)(E \times T) = 1\}$ is a Borel set. Now proceed by induction. □

References

[1] R. Aumann: "Agreeing to Disagree," *The Annals of Statistics,* Vol. 4 (1976), 1236-1239.

[2] P.Billingsley: *Convergence of Probability Measures.* New York, Wiley, 1968.

[3] W. Boge and Th. Eisele: "On Solutions of Bayesian Games," *International Journal of Game Theory,* Vol. 8 (1979), 193-215.

[4] C. Dellacherie and P.A. Meyer: *Probabilities and Potential,* New York, North Holland, Mathematics Studies 29, 1978.

[5] J. Harsanyi: "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science,* Vol. 14 (1967-68), 159-189, 320-334, 486-502.

[6] J.-F. Mertens and S. Zamir: "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory,* Vol. 14 (1985), 1-29.

CHAPTER 5

RATIONALIZABILITY AND CORRELATED EQUILIBRIUM

1. Introduction

The fundamental solution concept for noncooperative games is that of a Nash equilibrium [13]. Probably the most common view of Nash equilibrium is as a self-enforcing agreement. A game is envisaged as being preceded by a more or less explicit period of communication by the players. It is argued that if an agreement is reached to play a Nash equilibrium, then no player will have an incentive to violate it. Aumann [1] proposed the idea of objective and subjective correlated equilibrium as an extension of Nash equilibrium to allow for correlation between the players' randomizations and for subjectivity in the players' probability assessments.

The Nash equilibrium solution concept has been criticized from two opposing directions. On the one hand, the literature on refinements of Nash equilibrium (e.g. Myerson [11], Kreps and Wilson [9], Kohlberg and Mertens [8]) starts from the contention that not every Nash equilibrium can be viewed as a plausible agreed-upon way to play the game. On the other hand, Bernheim [3] and Pearce [14] have argued that Nash equilibrium is too restrictive in that it rules out behavior that does not contradict the rationality of the players. Bernheim and Pearce propose instead the concept of rationalizability as the logical consequence of assuming that the structure of the game and the rationality of the players (and nothing more) is common knowledge.

This chapter starts with the solution concept of rationalizability, since this is what is implied by the basic decision-theoretic analysis of a game. However, it is shown that rationalizability is more closely related to an equilibrium approach than one might at first think. An equivalence between rationalizable payoffs and payoffs from a posteriori equilibria -- a refinement of subjective correlated equilibria is proven. (A slightly specialized version of the results in this chapter appeared in [5].) Notice that the equivalence is stated in terms of payoffs and not strategies. There are two reasons for doing this. First, the equivalence is most readily stated in this form. Second, all that matters to a player in a game is what his/her expected payoff is -- (s)he does not care about the strategies played per se.

The equivalence results come in two parts depending on whether I start with "correlated" or "independent" rationalizability. The difference between correlated and independent rationalizability is that the second requires a player to believe that the others choose their actions independently, while the first does not. (Of course, the two versions of rationalizability coincide for two-person games.) Independent rationalizability is the appropriate concept if one thinks of the players in a "laboratory" situation: i.e. any correlating devices are explicitly modelled, the players are placed in separate rooms, and then are informed of the game they are to play. Correlated rationalizability seems more appropriate when the players are able to coordinate their actions via a large collection of correlating devices (such as sunspots) which are not explicitly modelled in the game but which are taken into account by allowing for correlated beliefs.

In Section 2 an equivalence between correlated rationalizable payoffs and payoffs from a posteriori equilibria is proven (Propositions 2.1 and 2.2). In view of this result one would also expect to be able to prove an equivalence between independent rationalizable payoffs and payoffs from "mixed" a posteriori equilibria. This intuition is correct, but the definition of mixed needed to prove the result is different from the standard notion of mixed strategies. Choosing the appropriate definition is quite delicate -- it turns out that there are several possible definitions depending on how players update their beliefs on null events. In Section 3 the appropriate definition of mixed in order to prove the equivalence result (Proposition 3.1) is presented. I go on to show that if a small change in this definition is made, then far from getting an equivalence with independent rationalizability, one gets an equivalence with Nash equilibrium (Propositions 4.1, 4.2).

The viewpoint of this chapter is that rationalizability is the solution concept implied by the "Bayesian rationality" of the players. It is then shown that there is a close relationship between rationalizability and a posteriori equilibrium. In a related paper [2], Aumann adopts a somewhat different notion of Bayesian rationality. In [2], Bayesian rationality is taken to mean that all players face some common state space Ω over which they have a common prior. There is a random variable mapping each state ω into a vector of actions for the players. Each player has a commonly known partition of Ω, and is assumed to choose an action to maximize his/her expected utility given his/her private information. Aumann shows that this set-up leads to an objective correlated equilibrium.

To obtain the results in this chapter and [2] within a unified framework, it is best to use types as in [5]. A type of a player is an infinite hierarchy of beliefs of that player -- over other players' actions, over other players' beliefs over other players' actions, and so on (cf. Chapter 4, Harsanyi [7], Mertens and Zamir [10]). If the only restriction on beliefs is common knowledge of rationality then by definition one gets rationalizability -- or according to the results in this chapter, a posteriori equilibrium. To get objective correlated equilibrium, two additional assumptions are needed. The first is consistency (in Harsanyi's sense), which corresponds to the assumption of common priors in [2]. The second is common knowledge of measurability (defined in [5]), which plays a similar role to the assumption in [2] of a commonly known random variable mapping states into actions.

2. Correlated Rationalizability and A Posteriori Equilibria

This section starts with a definition of the sets of correlated rationalizable strategies and payoffs in a game. The approach is based on that in [14]. However, unlike in [14], players are not allowed to select mixed strategies – doing so would not expand the set of rationalizable payoffs. Also, a player's beliefs over the actions of the other players is allowed to be correlated (cf. [14, p.1035]). In the next section the case in which these beliefs are independent is examined.

Consider an n-person game $\Gamma = \langle A^1,\ldots, A^n; u^1,\ldots, u^n\rangle$ where for each $i=1,\ldots,n$, A^i is player i's (finite) set of pure strategies and u^i is i's payoff function. For any finite set Y, let $\Delta(Y)$ denote the set of probability measures on Y. So $\Delta(A^i)$ (with typical element σ^i) denotes the set of mixed strategies of player i. The following notation will be used. Given sets $Y^1,\ldots, Y^n$, Y^{-i} denotes the set $Y^1\times\ldots\times Y^{i-1}\times Y^{i+1}\times\ldots\times Y^n$. I will also write y^{-i} for the element $(y^1,\ldots, y^{i-1}, y^{i+1}\ldots, y^n)\in Y^{-i}$.

DEFINITION 2.1: A subset $B^1\times\ldots\times B^n$ of $A^1\times\ldots\times A^n$ is a best reply set if for each i and every $a^i\in B^i$, there is a $\sigma\in\Delta(B^{-i})$ such that a^i is a best reply to σ. The set of *correlated rationalizable strategies* $R^1\times\ldots\times R^n$ is then the (finite) component by component union $(\bigcup_\alpha B^1_\alpha)\times\ldots\times(\bigcup_\alpha B^n_\alpha)$ of all best reply sets $B^1_\alpha\times\ldots\times B^n_\alpha$. For any $\sigma\in\Delta(R^{-i})$, i's maximal expected payoff against σ is a *correlated rationalizable payoff* to i. Let Π^i denote the set of all correlated rationalizable payoffs to i.

I now want to define an a posteriori equilibrium [1, Section 8] of the game Γ. I start by reviewing the definition of a correlated equilibrium of Γ, and then define an a posteriori equilibrium as a special type of correlated equilibrium. To define a correlated equilibrium of Γ, one must add to the basic description of the game a finite space Ω. The finiteness of Ω involves no loss of generality. Each player i has a prior P^i -- a probability measure on Ω -- and a partition H^i of Ω. A strategy of player i is an H^i-measurable map $f^i: \Omega \rightarrow A^i$. An n-tuple of strategies $(f^1,..., f^n)$ is a correlated equilibrium if for every i

$$\sum_{\omega \in \Omega} u^i[f^i(\omega), f^{-i}(\omega)] \, P^i(\{\omega\}) \geq \sum_{\omega \in \Omega} u^i[f'^i(\omega), f^{-i}(\omega)] \, P^i(\{\omega\})$$

for every strategy f'^i of i.

In this definition the players' strategies are only required to be ex ante optimal. In an a posteriori equilibrium the players' strategies must be optimal even after they have learned their private information. The following example motivates this distinction. Consider the game in Figure 1. Ω consists of two points ω_1, ω_2. Row is informed of the true state, Column has no private information. Row assigns (prior) probability 1 to ω_1, Column assigns probability ½ to ω_1, ½ to ω_2. The following strategies form a correlated equilibrium: Row plays U if informed that ω_1 happens, D if ω_2 happens; Column plays L. However, we (and Column) would not expect Row to play D if ω_2 happens (D is strongly dominated) unless Row is committed to do so ex ante. But such a commitment seems implausible. As in the refinement literature, one wants to require optimal behavior even on null events -- in this case after a move by Nature which is assigned prior probability zero. The definition of an a posteriori equilibrium (see below) is designed to rule out such situations. The unique a posteriori equilibrium of this game has Row playing U and hence Column playing R for sure.

To define an a posteriori equilibrium formally, the players' revised beliefs over Ω at every ω must be specified. In other words, for each player i and *every* $h^i \in H^i$, $P^i(\cdot \mid h^i)$ is required to be a probability measure on Ω and to be proper, i.e. to satisfy $P^i(h^i \mid h^i) = 1$ (cf. [4]). Of course if $P^i(h^i) > 0$, then by Bayes' rule $P^i(\cdot \mid h^i)$ automatically satisfies both requirements, but the point is that $P^i(\cdot \mid h^i)$ must satisfy them even if $P^i(h^i) = 0$. For each i, let $H^i(\omega)$ denote the cell in i's partition that contains ω.

DEFINITION 2.2: An n-tuple of strategies $(f^1,..., f^n)$ is an *a posteriori equilibrium* of Γ if for each i and every $\omega \in \Omega$:

$$\sum_{\omega' \in \Omega} u^i[f^i(\omega), f^{-i}(\omega')]\, P^i[\{\omega'\} \mid H^i(\omega)] \geq$$
$$\sum_{\omega' \in \Omega} u^i[a^i, f^{-i}(\omega')]\, P^i[\{\omega'\} \mid H^i(\omega)] \quad \forall\, a^i \in A^i.$$

Notice that by a change of variables, i's optimality condition requires that for every $\omega \in \Omega$:

$$\sum_{a^{-i} \in A^{-i}} u^i[f^i(\omega), a^{-i}]\, P^i[\{\omega' \mid f^{-i}(\omega') = a^{-i}\} \mid H^i(\omega)] \geq$$
$$\sum_{a^{-i} \in A^{-i}} u^i(a^i, a^{-i})\, P^i[\{\omega' \mid f^{-i}(\omega') = a^{-i}\} \mid H^i(\omega)] \quad \forall\, a^i \in A^i.$$

i's conditional expected payoff at any ω is called an *interim payoff* to i from the a posteriori equilibrium. i's *ex ante payoff* is the expectation of i's interim payoffs with respect to P^i.

The basic equivalence result proven in this section (Proposition 2.1) is between correlated rationalizable payoffs and interim payoffs from a posteriori equilibria. The idea behind rationalizability is that (according to Bayesian decision theory) player i has a certain given belief over the actions of the other players, and this determines i's (maximal) expected payoff. On the other hand, at the ex ante stage in an a posteriori equilibrium i does not yet know what his/her belief over the other players' actions will be. This belief will be equal to i's conditional probability which is determined by i's information, i.e. it is i's belief at the interim stage. This is why the basic equivalence result is stated in terms of interim payoffs. In fact, because of convexity of the set of correlated rationalizable payoffs to i (Lemma 2.3) one can also prove an equivalence between correlated rationalizable payoffs and ex ante payoffs from a posteriori equilibria -- see Proposition 2.2.

PROPOSITION 2.1: $(\pi^1,\ldots,\pi^n) \in \Pi^1 \times\ldots\times \Pi^n$ if and only if there is an a posteriori equilibrium of Γ in which $(\pi^1,\ldots,\pi^n)$ is the vector of interim payoffs.

Proposition 2.1 is a consequence of the following two lemmas.

LEMMA 2.1: If $(\pi^1,\ldots,\pi^n) \in \Pi^1 \times\ldots\times \Pi^n$ then there is an a posteriori equilibrium of Γ in which $(\pi^1 \ldots \pi^n)$ is the vector of interim payoffs.

PROOF: A mediator (cf. Myerson [12]) randomly selects a joint action

$(a^1,\ldots,a^n)$ and recommends to each player i to play a^i. Since π^i is a correlated rationalizable payoff to i, there is an $\hat{a}^i \in R^i$ and a $\hat{\sigma} \in \Delta(R^{-i})$ such that $\hat{a}^i$ is a best reply to $\hat{\sigma}$ and π^i is i's expected payoff from playing $\hat{a}^i$ against σ. If i is recommended to play $\hat{a}^i$ than the conditional probability with which i believes the mediator chooses actions in R^{-i} is σ. Similarly, for any other a^i in R^i choose a $\sigma \in \Delta(R^{-i})$ to which a^i is a best reply. If i is recommended to play a^i then the conditional probability with which i believes the mediator chooses actions in R^{-i} is σ. With these conditional probabilities i will be willing to follow the mediator's recommendations, and when informed of $\hat{a}^i$, i's conditional expected payoff from this a posteriori equilibrium is π. □

Observe that if i assigns probability 1 to the recommendation $\hat{a}^i$ then i's ex ante expected payoff is also π^i, so the term "interim" payoffs in Lemma 2.1 can be replaced with "ex ante".

LEMMA 2.2: The vector of interim payoffs from an a posteriori equilibrium $(f^1,\ldots,f^n)$ of Γ is an element of $\Pi^1 \times \ldots \times \Pi^n$.

PROOF: Let $A^i_+ = \{a^i \in A^i \mid a^i = f^i(\omega) \text{ for some } \omega \in \Omega\}$. $A^1_+ \times \ldots \times A^n_+$ is a best reply set. To prove this, I show that for every i and $a^i \in A^i_+$ there is a $\sigma \in \Delta(A^{-i}_+)$ to which a^i is a best reply. Given such an a^i choose an ω so that $f^i(\omega) = a^i$. Since $(f^1,\ldots,f^n)$ is an a posteriori equilibrium, and only strategies $a^{-i} \in A^{-i}_+$ "enter into" the equilibrium, i's optimality condition at ω can be written as:

$$\sum_{a^{-i} \in A_+^{-i}} u^i(a^i, a^{-i})\, P^i[\{\omega' \mid f^{-i}(\omega')=a^{-i}\} \mid H^i(\omega)] \geq$$
$$\sum_{a^{-i} \in A_+^{-i}} u^i(\tilde{a}^i, a^{-i})\, P^i[\{\omega' \mid f^{-i}(\omega')=a^{-i}\} \mid H^i(\omega)] \quad \forall\, \tilde{a}^i \in A^i.$$

This says that a^i is a best reply to the strategy σ which assigns probability $P^i[\{\omega' \mid f^{-i}(\omega')=a^{-i}\} \mid H^i(\omega)]$ to each $a^{-i} \in A_+^{-i}$.

Now choose any $\bar{\omega} \in \Omega$. I show that i's conditional expected payoff at $\bar{\omega}$ is a correlated rationalizable payoff to i. i's conditional expected payoff at $\bar{\omega}$ is the expected payoff from playing $f^i(\bar{\omega})$ against the strategy which assigns probability $P^i[\{\omega' \mid f^{-i}(\omega')=a^{-i}\} \mid H^i(\bar{\omega})]$ to each $a^{-i} \in A_+^{-i}$. That is, i's conditional expected payoff at $\bar{\omega}$ is a correlated rationalizable payoff to i. □

I have shown that i's interim payoff, conditional on $H^i(\omega)$, is a correlated rationalizable payoff. Therefore i's ex ante payoff is a convex combination of correlated rationalizable payoffs to i. Lemma 2.3 below says that the set of correlated rationalizable payoffs to i is convex, so in fact Lemma 2.2 implies that a vector of ex ante payoffs from an a posteriori equilibrium is a vector of correlated rationalizable payoffs. Together with the observation following Lemma 2.1 this implies:

PROPOSITION 2.2: The set of ex ante expected payoff vectors from a posteriori equilibria of Γ is equal to $\Pi^1 \times \ldots \times \Pi^n$.

As argued above, Proposition 2.2 will be implied by:

LEMMA 2.3: The set of correlated rationalizable payoffs to j in Γ is a closed interval in $\mathbb{R}$.

PROOF: For any $\sigma \in \Delta(R^{-j})$, let $v^j(a^j, \sigma)$ be j's expected payoff from playing a^j against σ. The set of correlated rationalizable payoffs to j is $\{\max_{a^j \in A^j} v^j(a^j, \sigma) \mid \sigma \in \Delta(R^{-j})\}$. This set is the image of the continuous map $\sigma \rightarrow \max_{a^j \in A^j} v^j(a^j, \sigma)$, and is therefore compact, connected since the domain $\Delta(R^{-j})$ is compact, connected. □

3. Independent Rationalizability and Mixed A Posteriori Equilibria

I begin by defining the set of independent rationalizable payoffs (cf. [3], [14]). This is the subset of the set of correlated rationalizable payoffs obtained by restricting a player's beliefs over the actions of the other players to be independent. To see that the set of independent rationalizable payoffs is a proper subset of the set of correlated rationalizable payoffs, consider the game in Figure 2. Player 1 chooses the row, 2 the column, 3 the matrix. 0.7 is a correlated rationalizable payoff to 3 as follows. 3 believes 1, 2 play (U, L) with probability ½, (D, R) with probability ½ (to which B is the best reply). 1 believes 2 plays L with probability ½, R with probability ½, and 3 plays B (to which U and D are best replies). 2 believes 1 plays U with probability ½, D with probability ½, and 3 plays B (to which L and R are best replies). On the other hand, 1 is the unique independent rationalizable payoff to 3. To see this, first note that B is not a best reply to any pair of mixed strategies of 1, 2. Hence 1, 2 must assign probability 0 to 3 playing B. But then U, L strongly dominate D, R for 1, 2 respectively.

DEFINITION 3.1: A subset $B^1 \times \ldots \times B^n$ of $A^1 \times \ldots \times A^n$ is an independent best reply set if for each i and every $a^i \in B^i$, there is a $\sigma^{-i} \in \prod_{k \neq i} \Delta(B^k)$ such that a^i is a best reply to σ^{-i}. The set of ***independent rationalizable strategies*** $R^1 \times \ldots \times R^n$ is then the (finite) component by component union $(\bigcup_\alpha B_\alpha^1) \times \ldots \times (\bigcup_\alpha B_\alpha^n)$ of all best reply sets $B_\alpha^1 \times \ldots \times B_\alpha^n$. For any $\sigma^{-i} \in \prod_{k \neq i} \Delta(R^k)$,

i's maximal expected payoff against σ^{-i} is an *independent rationalizable payoff* to i. The set of independent rationalizable payoffs to i is denoted Π^i.

The results of the last section would suggest an equivalence between independent rationalizable payoffs and interim payoffs from "mixed" a posteriori equilibria (and if the set of independent rationalizable payoffs is convex that the equivalence holds for ex ante payoffs also). This intuition is correct, however choosing the right definition of mixed a posteriori equilibrium is quite subtle. One might expect "mixed" to mean independence of the players' partitions of Ω in terms of their priors. In fact a form of conditional independence (which is not implied by a definition in terms of priors) is needed.

DEFINITION 3.2: $H^1,\ldots,H^n$ are P^i- *prior independent* if $P^i(\bigcap_{k=1}^{n} h^k) = \prod_{k=1}^{n} P^i(h^k)$ for every $h^k \in H^k$, $k=1,\ldots,n$. $H^1,\ldots,H^{i-1},H^{i+1},\ldots,H^n$ are P^i- *conditionally independent given* H^i if for every $h^i \in H^i$, $P^i(\bigcap_{k\neq i} h^k \mid h^i) = \prod_{k\neq i} P^i(h^k \mid h^i)$ for every $h^k \in H^k$, $k\neq i$.

Prior independence is the standard definition of independent σ-fields ([6, p.61]). It is also the notion of independence used in [1] to define mixed strategies. The above definition of conditional independence is a strengthening of the standard definition of conditionally independent σ-fields ([6, p.306]) from an almost everywhere to an everywhere requirement. Conditional independence says that *whatever* information i receives, i believes that the other players choose their actions independently. Prior independence implies that if $P^i(h^i)>0$, then $P^i(\bigcap_{k\neq i} h^k \mid h^i) = \prod_{k\neq i} P^i(h^k)$. So prior independence implies that the equality in the definition of conditional independence is satisfied for non P^i-null h^i 's. (In fact one can see that it implies more than this for such h^i's -- see Section 4.) But prior independence says nothing about $P^i(\bigcap_{k\neq i} h^k \mid h^i)$ if h^i is P^i-null, so prior independence does not imply conditional independence. Nor does conditional independence imply prior independence. A mixed a posteriori equilibrium will mean an a posteriori equilibrium in which for every i, $H^1,\ldots, H^{i-1}, H^{i+1},\ldots, H^n$ are P^i—conditionally independent given H^i.

LEMMA 3.1: Given a vector $(\pi^1,\ldots, \pi^n) \in \Pi^1 \times\ldots\times \Pi^n$, there is a mixed a posteriori equilibrium of Γ in which $(\pi^1,\ldots, \pi^n)$ is a vector of interim and ex ante payoffs to the players.

The proof is like that of Lemma 2.1. A mediator randomly selects a joint action $(a^1,\ldots, a^n) \in R^1 \times\ldots\times R^n$ and recommends to each player to play a^i. Since π^i is an independent rationalizable payoff to i, there is an $\hat{a}^i \in R^i$ and a $\sigma^{-i} \in \prod_{j\neq i} \Delta(R^j)$ such that $\hat{a}^i$ is a best reply to σ^{-i} and π^i is i's expected payoff from playing $\hat{a}^i$ against σ^{-i}. If i is recommended to play $\hat{a}^i$ then the conditional probability with which i believes the mediator chooses actions in R^{-i} is σ^{-i}. Note that σ^{-i} is a product measure on R^{-i}. Continuing in this way, after any recommendation i's conditional probability on R^{-i} is a product measure. So the a posteriori equilibrium constructed is mixed, and π^i is the interim payoff to i. By letting i assign prior probability 1 to the mediator recommending $\hat{a}^i$ this is also the ex ante payoff to i. □

LEMMA 3.2: Consider a mixed a posteriori equilibrium $(f^1,\ldots,f^n)$ of Γ. The interim expected payoff to each player is an element of Π^i.

The proof is essentially the same as that of Lemma 2.1. Let $A_+^1 \times \ldots \times A_+^n =$ $\{a^i \in A^i \mid a^i = f^i(\omega) \text{ for some } \omega \in \Omega\}$. $A_+^1 \times \ldots \times A_+^n$ is an independent best reply set. This follows from essentially the same argument as before, noting that $P^i[\{\omega' \mid f^{-i}(\omega') = a^{-i}\} \mid H^i(\omega)] = \prod_{k \neq i} P^i[\{\omega' \mid f^k(\omega') = a^k\} \mid H^i(\omega)]$ because of conditional independence. It follows that for any $\bar{\omega} \in \Omega$, j's conditional expected payoff at $\bar{\omega}$ is an independent rationalizable payoff to j. □

Lemma 3.2 implies that in any a posteriori equilibria each players ex ante payoff will be a convex combination of independent rationalizable payoffs, while a trivial modification of Lemma 3.2 shows that Π^i is convex. These remarks and Lemmas 3.1 and 3.2 provide an analog to Propositions 2.1 and 2.2.

PROPOSITION 3.1: The set of interim and ex ante payoff vectors in the mixed a posteriori equilibria of Γ is equal to $\Pi^1 \times \ldots \times \Pi^n$.

In order to prove an equivalence with independent rationalizability, the appropriate definition of mixed a posteriori equilibrium involves conditional independence. Recall that conditional independence does not in general imply prior independence. Nevertheless, when considering mixed a posteriori equilibria there is a sense in which prior independence can be assumed without loss of generality. More precisely, the set of interim and ex ante payoffs from mixed a posteriori equilibria which satisfy prior independence are also equal to $\Pi^1 \times \ldots \times \Pi^n$. This is because the a posteriori equilibrium constructed in the proof of Lemma 3.1 satisfy prior independence. And requiring the a posteriori equilibria of Lemma 3.2 to satisfy also prior independence will leave the conclusion that the payoffs are elements of Π^i for each i as valid.

4. A Posteriori Equilibria and Nash Equilibria

In this section I make a small change in the definition of conditional independence. Now, far from proving an equivalence with independent rationalizability, it is shown that for two-person games a posteriori equilibria satisfying the modified form of conditional independence are in fact equivalent to Nash equilibria. This result is extended to games with more than two players under an additional assumption of concordant priors (Definition 4.2).

Recall that conditional independence says that after observing any partition cell, i believes that the other players choose their actions independently. Prior independence says that i believes with probability 1 that (s)he will not update his/her prior. However, even both assumptions allow for a situation where $P^i(\cdot \mid h^i) \neq P^i(\cdot \mid \bar{h}^i)$ if $P^i(h^i)\, P^i(\bar{h}^i) = 0$. This says that i *can* update his/her prior if a null event happens. A possible restriction on beliefs is that the players do not update on null events either.

DEFINITION 4.1: $H^1,\ldots,H^{i-1},H^{i+1},\ldots,H^n$ are *P^i—everywhere independent given H^i* if for every $h^i \in H^i$, $P^i(\bigcap_{k\neq i} h^k \mid h^i) = \prod_{k\neq i} P^i(h^k)$ for every $h^k \in H^k$, $k\neq i$.

Clearly, everywhere independence implies both conditional and prior independence. But the converse is false as the following proposition implies.

PROPOSITION 4.1: (cf. [1, Proposition 8.1]) Suppose Γ is a two-person game. Then the sets of interim and ex ante payoff vectors from the a posteriori equilibria of Γ which satisfy everywhere independence are both equal to the set of payoff vectors from the Nash equilibria of Γ.

(An a posteriori equilibrium of Γ satisfies everywhere independence if for every i, $H^1,\ldots,H^{i-1},H^{i+1},\ldots,H^n$ are P^i–everywhere independent given H^i.) The proof is omitted since Proposition 4.1 is a special case of Proposition 4.2 below. Note that in Proposition 4.1 one gets Nash payoffs without requiring the players' priors over Ω to coincide. This is because everywhere independence implies that each player's posterior over the other player's actions is common knowledge. This idea can be extended to n-person games by requiring any two players to agree about a third player's actions.

DEFINITION 4.2: $P^1,\ldots,P^n$ are *concordant* if for each i and j, $k \neq i$, $P^j(h^i) = P^k(h^i)$ for every $h^i \in H^i$.

The assumption of concordant priors is closely related to that of common priors. It differs from the latter in that player i's prior over events in H^i need not be the same as the (common) prior of the other players. Technically, concordant priors differ only slightly from common priors. however, assuming the former is more natural since i's prior over events in H^i has no decision-theoretic significance for the play of the game. In any case, under the assumption of concordant priors there is an n-person analog to Proposition 4.1.

PROPOSITION 4.2: The sets of interim and ex ante payoff vectors from the a posteriori equilibria of Γ which satisfy everywhere independence and have concordant priors are both equal to the set of expected payoff vectors from the Nash equilibria of Γ.

PROOF: Consider an a posteriori equilibrium $(f^1,\ldots,f^n)$ of Γ and let $A_+^1,\ldots,A_+^n$ be defined as in the proof of Lemma 2.2. On any $h^i \in H^i$, i's conditional expected payoff from playing a^i is

$$\sum_{a^{-i}\in A_+^{-i}} u^i(a^i, a^{-i})\, P^i[\{\omega \mid f^{-i}(\omega)=a^{-i}\} \mid h^i]$$

which is equal to

$$\sum_{a^{-i}\in A_+^{-i}} u^i(a^i, a^{-i}) \prod_{j\neq i} P^i[\{\omega \mid f^j(\omega)=a^j\}]$$

by everywhere independence. Write $P^i[\{\omega \mid f^j(\omega)=a^j\}] = \sigma^j(a^j)$ and let $\sigma^j \in \Delta(A_+^j)$ be the mixed strategy which assigns probability $\sigma^j(a^j)$ to each $a^j \in A_+^j$. Note that σ^j does not depend on i by the assumption of concordant priors. In other words, i's expected payoff from playing a^i is the expected payoff from playing a^i against the vector of mixed strategies σ^{-i}. Let $BR^i(\sigma^{-i})$ be the set of i's best replies against σ^{-i}. Then $A_+^i \subset BR^i(\sigma^{-i})$. Hence there are sets $A_+^1,\ldots,A_+^n$ and mixed strategies $\sigma^1 \in \Delta(A_+^1),\ldots,\sigma^n \in \Delta(A_+^n)$ such that $A_+^1 \subset BR^1(\sigma^{-1}),\ldots, A_+^n \subset BR^n(\sigma^{-n})$. So $(\sigma^1,\ldots,\sigma^n)$ is a Nash equilibrium. In other words, i's conditional expected payoff on any h^i -- and hence i's ex ante payoff -- is equal to i's expected payoff from a Nash equilibrium. The converse is straightforward. □

5. Concluding Remarks

This chapter began with the basic decision-theoretic concept of rationalizability. It was then shown that there is a close connection between rationalizability and a posteriori equilibrium. The formal results fell into two groups depending on whether I worked with correlated or independent rationalizability. The structure of the argument was the same in both cases. First, an a posteriori equilibrium with an interim payoff equal to a given rationalizable payoff was constructed, and then it was shown that any interim payoff from an a posteriori equilibrium is a rationalizable payoff. Second, the convexity of the set of rationalizable payoffs was used to extend the equivalence to ex ante payoffs. Two properties of the solution concepts which were derived in the course of the proofs are of some interest in their own right: first, the sets of correlated and independent rationalizable payoffs to a player are convex; second, the set of payoffs to a player from "mixed" a posteriori equilibria shifts from the set of independent rationalizable payoffs to the set of Nash payoffs depending on how players update their beliefs on null events.

	L	R
U	0 3	2 1
D	4 0	1 0

FIGURE 1

A

	L	R
U	1,1,1	1,0,1
D	0,1,0	0,0,0

B

	L	R
U	2,2,.7	0,0,0
D	0,0,0	2,2,.7

C

	L	R
U	1,1,0	1,0,0
D	0,1,1	0,0,1

FIGURE 2

References

[1] R. Aumann: "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, Vol. 1, (1974), 67-96.

[2] ------: "Correlated Equilibrium as an Expression of Bayesian Rationality," Unpublished, The Hebrew University, June 1985.

[3] D. Bernheim: "Rationalizable Strategic Behavior," *Econometrica*, Vol. 52 (1984), 1007-1028.

[4] D. Blackwell and L. Dubins: "On Existence and Non-existence of Proper, Regular, Conditional Distributions," *The Annals of Probability*, Vol. 3 (1975), 741-752.

[5] A. Brandenburger and E. Dekel: "Notes on a Bayesian Approach to Non-cooperative Solution Concepts," Unpublished, Graduate School of Business, Stanford University, March 1985.

[6] K. L. Chung: *A Course in Probability Theory*. New York: Academic Press, Second Edition 1974.

[7] J. Harsanyi: "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science*, Vol. 14 (1967-68), 159-189, 320-334, 486-502.

[8] E. Kohlberg and J.-F. Mertens: "On the Strategic Stability of Equilibria," Unpublished, Harvard Business School, August 1984.

[9] D. Kreps and R. Wilson: "Sequential Equilibria," *Econometrica*, Vol. 50 (1982), 863-894.

[10] J.-F. Mertens and S. Zamir: "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, Vol. 14 (1985), 1-29.

[11] R. Myerson: "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, Vol. 7 (1978), 73-80.

[12] ------: "Bayesian Equilibrium and Incentive-Compatibility: An Introduction," In L. Hurwicz, D. Schmeidler, H. Sonnenschein (eds.), Social Goals and Social Organization, Forthcoming.

[13] J. Nash: "Non-Cooperative Games," *Annals of Mathematics*, Vol. 54 (1951), 286-295.

[14] D. Pearce: Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, Vol. 52 (1984), 1029-1050.