Consequentialism and Bayesian Rationality in Normal Form Games

Peter J. Hammond

Department of Economics, Stanford University, CA 94305-6072, U.S.A.
e-mail: hammond@leland.stanford.edu

1 Introduction

In single-person decision theory, Bayesian rationality requires the agent first to attach subjective probabilities to each uncertain event, and then to maximize the expected value of a von Neumann–Morgenstern utility function (or NMUF) that is unique up to a cardinal equivalence class. When the agent receives new information, it also requires subjective probabilities to be revised according to Bayes' rule.

In social choice theory and ethics, Harsanyi (1953, 1955, 1975a, 1975b, 1976, 1978) has consistently advocated Bayesian rationality as a normative standard, despite frequent criticism and suggestions for alternatives. In game theory, however, Bayesian rationality is almost universally accepted, not only as a normative standard, but also in models intended to describe players' actual behaviour. Here too Harsanyi (1966, 1967–8, 1977a, b, 1980, 1982a, b, 1983a, b) has been a consistent advocate. In particular, his work on games of incomplete information suggests that one should introduce extra states of nature in order to accommodate other players' types, especially their payoff functions and beliefs. Later work by Bernheim (1984), Pearce (1984), Tan and Werlang (1988), and others emphasizes how subjective probabilities may be applied fruitfully to other players' strategic behaviour as well.

In the past I have tried to meet the social choice theorists' understandable criticisms of the Bayesian rationality hypothesis. To do so, I have found it helpful to consider normative standards of behaviour in single-person decision trees. In particular, it has been useful to formulate a surprisingly powerful "consequentialist" hypothesis. This requires the set of possible consequences of behaviour in any single-person decision tree to depend only on the feasible set of consequences in that tree. In other words, behaviour must reveal a consequence choice function mapping feasible sets into choice subsets.

Previous work (Hammond, 1988a, b, 1997a, b) has applied this consequentialist hypothesis to dynamically consistent behaviour in an (almost) unrestricted domain of finite decision trees. The only restriction is that objective probabilities must all be positive at any chance node. Then, provided that behaviour is continuous as objective probabilities vary, provided that there is state independence, and provided also that Anscombe and Aumann's (1963) reversal of order axiom is satisfied, it follows that behaviour must be Bayesian rational. Moreover, null events are excluded, so strictly positive subjective probabilities must be attached to all states of the world. Of course, these arguments do not really justify Bayesian rationality; they merely indicate that critics and proponents of alternative theories should go beyond discussions of simple "one-shot" problems and explain how to make sequential decisions in trees.

The aim of this paper and some associated work (Hammond, 1997c) is to extend the consequentialist hypothesis from single-person decision trees to n-person games. The appropriate extension appears to be the *consequentialist normal form invariance hypothesis*, with antecedents in von Neumann and Morgenstern (1944, 1953). Yet the results concerning consequentialist behaviour in single-person decision trees rely on being able to consider, if not necessarily a completely unrestricted domain of decision trees with a fixed set of states of the world, then at least one that is rich enough. In particular, a player i's preferences over the random consequences of two different strategies are revealed by forcing i to choose between just those two strategies. Now, when such alterations in the options available to an agent occur in a single-person decision tree, there is no reason to believe that nature's exogenous "choice" will change. But as Mariotti (1996) and Battigalli (1996) have pointed out, when such changes apply to an n-person game, they typically affect that player's anticipated behaviour in a way that makes other players want to change their strategies. Then, of course, it is illegitimate to treat these other players' strategies as exogenous.

In order to surmount this difficulty, Battigalli's (1996) comment on Mariotti's paper suggests introducing, for each player i whose subjective probabilities are to be determined, one extra player i^* who is an exact copy or clone of i. Player i^* faces a variable opportunity to bet on how players other than i will play the original game, but is unable to affect the consequences available to all the other players, including i. With this useful and ingenious device, player i^* can be faced with each possible single-person decision tree in turn. This allows i^*'s subjective probabilities over strategy profiles for players other than i to be inferred. Moreover, they apply to i^*'s behaviour when facing a single-person decision problem equivalent to that which i faces in the game itself. Because i^* is an exact copy of i, it follows that i's behaviour in the original game matches i^*'s in this equivalent single-person decision problem; in particular, i will maximize subjective expected utility using i^*'s subjective probabilities.

Hence, Battigalli's device can be used to provide a consequentialist justification for behaviour in n-person games to be Bayesian rational. There is a need, however, to attach strictly positive probabilities to all other players' strategies which are not ruled out as completely impossible and so irrelevant to the game. This suggests that strictly positive probabilities should be attached to all other players' rationalizable strategies, at least — i.e., to all those that are not removed by iterative deletion of strictly dominated strategies.

In the remainder of the paper, Section 2 reviews in somewhat more detail the earlier results characterizing consequentialist behaviour in single-person decision trees. Then Section 3 defines a "consequentialist" n-person normal game form G, in which payoffs are replaced by personal consequences, together with associated families $\mathcal{G}_i$. For each $i \in N$, these consist of game forms $G(i, T)$ which differ from G in having an extra player i^* who is an exact copy of i and faces the decision tree T. Next, Section 4 introduces three different kinds of "consequentialist" player type in order to describe each player's behaviour in every family $\mathcal{G}_i$. Last, Section 5 explains why the consequentialist hypotheses imply that there should be subjective probabilities for each player that are attached to the profiles of all other players' strategies.

2 Consequentialist Single-Person Decision Theory: A Brief Review

In single-person decision theory, the basic "consequentialist" hypothesis requires that actions should be evaluated purely on the basis of their consequences. More specifically, if two decision trees face the decision maker with identical feasible sets of consequences, then behaviour in those trees should generate, or reveal as "chosen", identical sets of consequences in the two trees. This consequentialist hypothesis has strong implications when applied to an unrestricted domain of finite decision trees involving a given *consequence domain* Y, on which behaviour is required to satisfy a mild "dynamic consistency" requirement.

The first and simplest result applies to finite decision trees T in the domain $\mathcal{T}_1(Y)$ which contain decision nodes and also terminal nodes having consequences in the specified domain Y. In any tree $T \in \mathcal{T}_1(Y)$, behaviour has sure consequences in the domain Y. On the unrestricted domain $\mathcal{T}_1(Y)$, consequentialist dynamically consistent behaviour must be *ordinal* in the sense that it both reveals and maximizes a (complete and transitive) weak preference ordering R over the feasible set of consequences. This can be proved either indirectly by means of Arrow's (1959) characterization of ordinal choice, or directly — see Hammond (1977, 1988b, 1997a) for more details. Moreover, ordinality is a complete characterization of behaviour satisfying the three "consequentialist" axioms.

A second result applies to finite decision trees T in the domain $\mathcal{T}_2(Y)$ which, in addition to decision nodes and terminal nodes having consequences in the specified domain Y, also contain chance nodes at which strictly positive objective probabilities are specified.[1] In any such tree, behaviour has random consequences in the domain $\Delta(Y)$ of simple (finitely supported) lotteries on Y. Then, on the unrestricted domain $\mathcal{T}_2(Y)$, consequentialist dynamically consistent behaviour not only maximizes a weak preference ordering R over the feasible set of consequence lotteries; the independence axiom due to Marschak (1950) and Samuelson (1952) is also satisfied. Again, this is a complete characterization of behaviour satisfying the three axioms. If behaviour is also required to vary continuously as objective probabilities vary, in the sense of generating a closed graph behaviour correspondence from probabilities to decisions, then R must also satisfy a familiar continuity axiom ensuring the existence of a unique cardinal equivalence class of NMUFs on Y, all of whose expected values represent R.

A final third result is required in order to justify subjective expected utility maximization. It is necessary to consider a new class of finite decision trees T in the family of domains $\mathcal{T}_3(E, Y)$ ($\emptyset \neq E \subset S$), where Y is the consequence domain, S is a finite set of possible uncertain *states* of the world, and each non-empty $E \subset S$ respresents an *event*. Trees in each domain $\mathcal{T}_3(E, Y)$ contain "natural" nodes where nature refines a partition of the subset of E that corresponds to the set of states which remain possible. These nodes are in addition to the decision nodes, to the terminal nodes having consequences in the specified domain Y, and to the chance nodes at which strictly positive objective probabilities are specified. In any tree $T \in \mathcal{T}_3(E, Y)$, behaviour has random consequences in the domain $\Delta(Y^E)$ of simple lotteries on the Cartesian product space $Y^E = \prod_{s \in E} Y_s$ whose

[1] Allowing zero probabilities at chance nodes yields the unacceptably strong conclusion that all lotteries in $\Delta(Y)$ should be indifferent.

members $y^E = \langle Y_s \rangle_{s \in E}$ specify a function $s \mapsto y_s$ mapping states of the world $s \in E$ to consequences $y_s \in Y_s = Y$. Of course, each such y^E is what Savage (1954) calls an "act".

Now, given the family $\mathcal{T}_3(E, Y)$ $(\emptyset \neq E \subset S)$ of unrestricted domains, consequentialist dynamically consistent behaviour in any tree $T \in \mathcal{T}_3(E, Y)$ must not only maximize a conditional weak preference ordering R_E over the feasible set of consequence lotteries in $\Delta(Y^E)$. In addition, R_E must satisfy the Marschak–Samuelson independence axiom. Moreover, different the orderings R_E $(\emptyset \neq E \subset S)$ must together satisfy Anscombe and Aumann's (1963) extension to "random acts" of the sure thing principle originally formulated by Savage (1954). As before, this is a complete characterization of behaviour satisfying the three axioms. If behaviour is also continuous as objective probabilities vary, then for each event E there must exist a unique cardinal equivalence class of NMUFs on Y^E whose expected values all represent R_E. Moreover, under two additional axioms like those used by Anscombe and Aumann (1963) — namely reversal of order and state independence — there must exist a unique and strictly positive family of subjective conditional probabilities $P(s|E)$ $(s \in E \subset S)$ satisfying Bayes' rule, together with a unique cardinal equivalence class of state-independent NMUFs on Y whose subjective conditionally expected values all represent each ordering R_E. In this sense, when supplemented by mild additional conditions, consequentialism implies the Bayesian rationality hypothesis.

3 A Family of Consequentialist Normal Game Forms

It would seem highly desirable to have a single integrated theory of normative decisions which applies to all n-person games, and which reduces to consequentialist single-person decision theory in the special case of one-person games "against nature". Accordingly, it appears natural to formulate the *consequentialist normal form invariance hypothesis*. This requires that, whenever two game forms have identical or "equivalent" normal forms, any player's strategic behaviour in those two games should give rise to identical sets of consequences that are revealed as chosen. The hypothesis is an obvious adaptation of a claim that figured so prominently in von Neumann and Morgenstern's book (1944, 1953). Obviously, it is also a natural extension to extensive form games of the consequentialist hypothesis that was previously advanced for decision trees. The aim of this paper is to sketch the main implications of this hypothesis, to be discussed elsewhere in more detail later on.

The first step in such a theory is to formulate an appropriate extension of single-person decision trees to n-person "consequentialist extensive game forms". These are like decision trees and game forms, but they differ from orthodox extensive form games in that they have consequences rather than payoffs attached to terminal nodes. This seems entirely appropriate for a theory in which the existence of a payoff function, in the form of a von Neumann–Morgenstern utility function, should be a major result rather than a questionable assumption.

However, to save space, I shall proceed directly to the definition of an associated *consequentialist normal game form*. This consists of a list

$$G = \langle N, Y^N, S^N, \phi^N \rangle$$

Here N denotes the finite set of *players*. Each player $i \in N$ is assumed to have a *personal consequence domain* Y_i which is one component of the Cartesian product $Y^N := \prod_{i \in N} Y_i$ of *consequence profiles*, and also a finite strategy space S_i making up one component of the Cartesian product $S^N := \prod_{i \in N} S_i$ of *strategy profiles*. Finally, $\phi^N : S^N \to \Delta(Y^N)$ is the *outcome function* determining the random consequence profile that results from each possible strategy profile. For each $i \in N$, let $\phi_i(s^N) \in \Delta(Y_i)$ denote the marginal distribution on Y_i that is induced by $\phi^N(s^N)$ on Y^N.

Given any such game form G, it will also be necessary to consider a family $\mathcal{G} = \{G\} \cup (\cup_{i \in N} \mathcal{G}_i)$ of game forms derived from G, where

$$\mathcal{G}_i = \{\, G(i, T) \mid T \in \mathcal{T}_3(S_{-i}, Y_i) \,\}$$

That is, for each player $i \in N$ and tree $T \in \mathcal{T}_3(S_{-i}, Y_i)$, there is a corresponding game form in $G(i, T)$ in $\mathcal{G}_i$ specified by

$$G(i, T) = \langle \{i^*\} \cup N, Y_i \times Y^N, S^T \times S^N, \bar{\phi}^{\{i^*\} \cup N} \rangle$$

As explained in the introduction, this involves one extra player i^* who is a copy of player i. So player i^*'s consequence space, like i's, is Y_i. It is assumed that player i^* effectively faces a single-person decision tree $T \in \mathcal{T}_3(S_{-i}, Y_i)$, in which the set of possible states of nature is S_{-i}. The finite set of i^*'s strategies in tree T is denoted by S^T, and the outcome function is assumed to be $\phi^T : S^T \to \Delta(Y^{S_{-i}})$. In $G(i, T)$ the outcome function is given by

$$\bar{\phi}_{i^*}(s^T, s^N) := \phi^T(s^T) \quad \text{and} \quad \bar{\phi}_j(s^T, s^N) := \phi_j(s^N) \text{ (all } j \in N)$$

for all $(s^T, s^N) \in S^T \times S^N$. Note that, for each player $j \in N$, the random outcome in $G(i, T)$ of each strategy profile $s^N \in S^N$ that can be played in G is the same as the random outcome of s^N in G itself. In particular, this random outcome must independent both of T and of i^*'s choice of strategy $s^T \in S^T$ in T.

4 Players' Type Spaces

It may be useful to think of a game form as a book of rules, specifying what strategies players are allowed to choose, and what random consequence results from any allowable profile of strategic choices. So the family $\mathcal{G}$ of consequentialist game forms needs fleshing out with descriptions of players' preferences, beliefs, and behaviour. The Bayesian rationality hypothesis involves preferences represented by expected values of von Neumann–Morgenstern utility functions (NMUFs) attached to consequences. Also, beliefs take the the form of subjective probabilities attached jointly to combinations of other players' preferences, strategies, and beliefs. And behaviour should maximize subjectively expected utility. It has yet to be shown, however, that the consequentialist hypotheses imply such preferences, beliefs, and behaviour. To do so satisfactorily requires a framework for describing preferences, beliefs, and behaviour in game forms before the consequentialist hypotheses have been imposed. We shall postulate type spaces similar to those considered by Harsanyi (1967–8) in his discussion of games of incomplete information. However, each player will have three separate types, corresponding to preferences, beliefs, and behaviour respectively.

Indeed, since one cannot directly assume that preferences exist, it is necessary to consider instead, for each player $i \in N$, a *decision type* $d_i \in D_i$ which determines what is acceptable behaviour for i in any single-person finite decision tree $T \in \mathcal{T}_2(Y_i)$ having random consequences in $\Delta(Y_i)$. Of course, consequentialist normal form invariance implies the consequentialist hypotheses for single-person decision theory. So if continuity of behaviour is added to these hypotheses, we know already that each player $i \in N$ will have a unique cardinal equivalence class of NMUFs $v_i(y_i; d_i)$ on Y_i which are parametrized by their decision type d_i. Together, the list of all players' types forms a *decision type profile* $d^N \in D^N := \prod_{i \in N} D_i$.

As in orthodox game theory, each player $i \in N$ is assumed next to have beliefs or an *epistemic type* $e_i \in E_i$, with $E^N := \prod_{i \in N} E_i$ as the set of all possible *epistemic type profiles*. It will be a result rather than an assumption of the theory that all such beliefs can be represented by subjective probabilities on an appropriately defined space. For the moment, each $e_i \in E_i$ is assumed to determine parametrically player i's *strategic behaviour* in the form of a non-empty set $\sigma_i(G', d_i, e_i) \subset S_i$ defined for every game form $G' \in \mathcal{G}$ and each possible decision type d_i for player i. In orthodox game theory, $\sigma_i(G', d_i, e_i)$ is the set of i's "best responses" given the NMUF $v_i(y_i; d_i)$ and subjective probability beliefs over other players' strategies determined by e_i. The assumption that such a parameter e_i exists is without loss of generality because if necessary this parameter could be the correspondence $(G', d_i) \mapsto\!\!\!\to \sigma_i$ itself. Finally, it is also necessary to define $\sigma_{i^*}(G', d_i, e_i)$ for the copy i^* of player i in every game $G' \in \mathcal{G}_i$. Note that, because i^* is a copy of i, player i^*'s behaviour depends on i's type pair (d_i, e_i), as the above notation reflects.

From the normative point of view, each set $\sigma_i(G', d_i, e_i)$ already describes how i with decision type d_i and epistemic type e_i should play G'. However, in forming beliefs, it is not enough for player i (and also i^* if $G' \in \mathcal{G}_i$) to know the other players' sets $\sigma_j(G', d_j, e_j)$ $(j \in N \setminus \{i\})$; also relevant are the tie-breaking rules which the other players $j \in N \setminus \{i\}$ use to select one particular strategy s_j from the set $\sigma_j(G', d_j, e_j)$ whenever this set has more than one member. Accordingly, each player $i \in N$ is assumed to have in addition a *behaviour type* $b_i \in B_i$, with $B^N := \prod_{i \in N} B_i$ as the set of all possible *behaviour type profiles*. Each $b_i \in B_i$ is assumed to determine parametrically player i's *selection rule* yielding a single member $s_i(G', d_i, e_i, b_i) \in \sigma_i(G', d_i, e_i)$ of each strategic behaviour set. The assumption that b_i exists is without loss of generality because it could be the function $(G', d_i, e_i) \mapsto s_i$ itself. Note that player i^*'s behaviour type need not be specified because i^*'s behaviour has no effect on any other player.

To simplify notation in future, define for each player $i \in N$ a combined type space $\Theta_i := D_i \times E_i \times B_i$, whose members are triples $\theta_i := (d_i, e_i, b_i)$. Note that each player's selection rule can then be expressed as $s_i(G', \theta_i)$. Let $\Theta^N := D^N \times E^N \times B^N$ be the space of combined type profiles, with typical member $\theta^N := (d^N, e^N, b^N)$, and let $\Theta_{-i} := \prod_{j \in N \setminus \{i\}} \Theta_i$ denote the set of all possible types for players other than i. A complete epistemic type $e_i \in E_i$ should then describe in full player i's beliefs about the other players' types $\theta_{-i} \in \Theta_{-i}$, including their epistemic types e_{-i}. This creates a problem of circularity or infinite regress which is an inevitable and fundamental part of modern game theory.

5 Subjective Probabilities over Other Players' Strategies

First, given any game $G' = G(i, T) \in \mathcal{G}_i$, suppose that player i^* moves first, before any player $j \in N$, without these players knowing what i^* has chosen. Then after i^* has moved, G is effectively a "subgame of incomplete information". Now, given any player $j \in N$, and any combined type $\theta_j \in \Theta_j$ that player j may have, applying an obvious dynamic consistency hypothesis to the subgame G of $G' = G(i, T)$ yields the result that

$$\sigma_j(G', d_j, e_j) = \sigma_j(G, d_j, e_j) \quad \text{and} \quad s_j(G', \theta_j) = s_j(G, \theta_j)$$

In particular, for each $j \in N$, both $\sigma_j(G', d_j, e_j)$ and $s_j(G', \theta_j)$ are effectively independent of whatever player $i \in N$ is copied and of whatever tree $T \in \mathcal{T}_3(S_{-i}, Y_i)$ is given to the copy i^* of player i. So variations in i^*'s decision tree within the domain $\mathcal{T}_3(S_{-i}, Y_i)$ are possible without inducing changes in the behaviour of other players $j \in N$. This justifies applying the consequentialist hypotheses to the whole domain $\mathcal{T}_3(S_{-i}, Y_i)$ of single-person decision trees faced by i^* and so by i, while treating each $s_{-i} \in S_{-i}$ as a state of nature determined entirely outside the tree. So the usual arguments imply the existence of unique and strictly positive subjective probabilities $P_i(s_{-i})$ $(s_{-i} \in S_{-i})$ such that i^*'s decisions in trees $T \in \mathcal{T}_3(S_{-i}, Y_i)$ maximize the subjectively expected value of a von Neumann–Morgenstern utility function $v_i(y_i; d_i)$ parametrized by i's decision type $d_i \in D_i$.

It remains to consider player i's behaviour in the game form G itself. To do so, consider the special decision tree $T_i^G \in \mathcal{T}_3(S_{-i}, Y_i)$ in which the set of i^*'s strategies is S_i, equal to i's strategy set in G, and the random outcome of each strategy $s_i \in S_i$ is specified by

$$\phi_i^G(s_i)(y^{S_{-i}}) = \prod\nolimits_{s_{-i} \in S_{-i}} \phi(s_i, s_{-i})(y_{s_{-i}})$$

for all $y^{S_{-i}} = \langle y_{s_{-i}} \rangle_{s_{-i} \in S_{-i}} \in Y^{S_{-i}}$. Then, at least under Anscombe and Aumann's reversal of order axiom, both the strategy set S_i and the outcome function ϕ_i^G are exactly the same as in G itself. In this case T_i^G and G are consequentially equivalent from i's (or i^*'s) point of view, so consequentialism requires i's behaviour in G to match that of i^* in T_i^G or $G(i, T_i^G)$. This implies that $\sigma_{i^*}(G(i, T_i^G), d_i, e_i) = \sigma_i(G, d_i, e_i)$.

It follows that player i should behave according to the hypothesis of Bayesian rationality and choose $s_i \in S_i$ to maximize subjectively expected utility based on the subjective probabilities $P_i(s_{-i})$ $(s_{-i} \in S_{-i})$ that are appropriate for all decision trees in $\mathcal{T}_3(S_{-i}, Y_i)$. Really, one should write these probabilities as $P_i(s_{-i}, e_i)$ to indicate that they represent player i's epistemic type e_i and so characterize i's acceptable behaviour sets $\sigma_i(G', d_i, e_i)$ on the domain $\mathcal{G}$ of game forms, including G itself. This is the promised consequentialist justification of the Bayesian rationality hypothesis in normal form n-person games.

Acknowledgements

November 1996 revision of a paper presented to the SITE (Stanford Institute of Theoretical Economics) summer workshop on "Game Theory: Epistemic and Other Foundational Issues". The paper is based on an earlier presentation to the International Symposium in Honor of John C. Harsanyi on "Game Theory, Experience, Rationality: Foundations of the Social Sciences, Economics and Ethics" organized by the Institute Vienna Circle in Vienna, June 12–15, 1996. I am grateful to Marco Mariotti and Pierpaolo Battigalli for most helpful discussions of this and related work.

References

F.J. Anscombe and Robert J. Aumann (1963) "A Definition of Subjective Probability", in: *Annals of Mathematical Statistics* 34, pp. 199–205.

Kenneth J. Arrow (1959) "Rational Choice Functions and Orderings", in: *Economica* 26, pp. 121–127.

Pierpaolo Battigalli (1996) "Comment [on Mariotti (1996)]", in: Kenneth J. Arrow, Enrico Colombatto, Mark Perlman, and Christian Schmidt (Eds.) *The Rational Foundations of Economic Behaviour* (London: Macmillan), pp. 149–154.

B. Douglas Bernheim (1984) "Rationalizable Strategic Behavior", in: *Econometrica* 52, pp. 1007–1028.

B. Douglas Bernheim (1986) "Axiomatic Characterizations of Rational Choice in Strategic Environments", in: *Scandinavian Journal of Economics* 88, pp. 473–488.

Peter J. Hammond (1977) "Dynamic Restrictions on Metastatic Choice", in: *Economica* 44, pp. 337–350.

Peter J. Hammond (1988a) "Consequentialism and the Independence Axiom", in: Bertrand Munier (Ed.) *Risk, Decision, and Rationality: Proceedings of the 3rd International Conference on the Foundations and Applications of Utility, Risk, and Decision Theories.* Dordrecht: D. Reidel, pp. 503–516.

Peter J. Hammond (1988b) "Consequentialist Foundations for Expected Utility", in: *Theory and Decision* 25, pp. 25–78.

Peter J. Hammond (1997a) "Objective Expected Utility: A Consequentialist Perspective", in: Salvador Barberà, Peter J. Hammond and Christian Seidl (Eds.) *Handbook of Utility Theory.* Dordrecht: Kluwer Academic Publishers — in preparation.

Peter J. Hammond (1997b) "Subjective Expected Utility", in: Salvador Barberà, Peter J. Hammond and Christian Seidl (Eds.) *Handbook of Utility Theory*. Dordrecht: Kluwer Academic Publishers — in preparation.

Peter J. Hammond (1997c) "Utility Theory for Non-Cooperative Games", in: Salvador Barberà, Peter J. Hammond and Christian Seidl (Eds.) *Handbook of Utility Theory*. Dordrecht: Kluwer Academic Publishers — in preparation.

John C. Harsanyi (1953) "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", in: *Journal of Political Economy* 61, pp. 434–5; reprinted in Harsanyi (1976).

John C. Harsanyi (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", in: *Journal of Political Economy* 63, pp. 309–321; reprinted in Harsanyi (1976).

John C. Harsanyi (1966) "A General Theory of Rational Behavior in Game Situations", in: *Econometrica* 34, pp. 613–634.

John C. Harsanyi (1967–8) "Games with Incomplete Information Played by 'Bayesian' Players, I–III", in: *Management Science* 14, pp. 159–182, 320–334, 486–502; reprinted in Harsanyi (1982a), chs. 6–8.

John C. Harsanyi (1975a) "Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" in: *Theory and Decision* 6, pp. 311–32; reprinted in Harsanyi (1976).

John C. Harsanyi (1975b) "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory", in: *American Political Science Review* 69, pp. 594–606; reprinted in Harsanyi (1976).

John C. Harsanyi (1976) *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: D. Reidel.

John C. Harsanyi (1977a) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

John C. Harsanyi (1977b) "Advances in Understanding Rational Behavior", in: Robert E. Butts and Jaakko Hintikka (Eds.) *Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, Vol. II*. Dordrecht: D. Reidel, pp. 315–343; reprinted in Harsanyi (1976).

John C. Harsanyi (1978) "Bayesian Decision Theory and Utilitarian Ethics", in: *American Economic Review (Papers and Proceedings)* 68, pp. 223–8.

John C. Harsanyi (1980) "Uses of Bayesian Probability Models in Game Theory", in: D.H. Milnor (Ed.) *Science, Belief and Behaviour: Essays in Honour of R.B. Braithwaite*. Cambridge: Cambridge University Press, pp. 189–201; reprinted in Harsanyi (1982a), ch. 9.

John C. Harsanyi (1982a) *Papers in Game Theory*. Dordrecht: D. Reidel.

John C. Harsanyi (1982b) "Subjective Probability and the Theory of Games: Comments on Kadane and Larkey's Paper" and "Rejoinder to Professors Kadane and Larkey", in: *Management Science* 28, pp. 120–124 and 124–125.

John C. Harsanyi (1983a) "Uses of Subjective Probability in Game Theory", in: Bernt P. Stigum and Fred Wenstop (Eds.) *Foundations of Utility and Risk with Applications*. Dordrecht: D. Reidel, pp. 297–310.

John C. Harsanyi (1983b) "Bayesian Decision Theory, Subjective and Objective Probabilities, and Acceptance of Empirical Hypotheses", in: *Synthese* 57, pp. 341–365.

Marco Mariotti (1996) "The Decision-Theoretic Foundations of Game Theory" in: Kenneth J. Arrow, Enrico Colombatto, Mark Perlman, and Christian Schmidt (Eds.) *The Rational Foundations of Economic Behaviour*. London: Macmillan, ch. 6, pp. 133–148.

Jacob A. Marschak (1950) "Rational Behavior, Uncertain Prospects, and Measurable Utility", in: *Econometrica* 18, pp. 111–141.

David Pearce (1984) "Rationalizable Strategic Behavior and the Problem of Perfection", in: *Econometrica* 52, pp. 1029–1050.

Paul A. Samuelson (1952) "Probability, Utility, and the Independence Axiom", in: *Econometrica* 20, pp. 670–678.

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

Tommy C.-C. Tan and Sérgio R. da C. Werlang (1988) "The Bayesian Foundations of Solution Concepts of Games", in: *Journal of Economic Theory* 45, pp. 370–391.

John von Neumann and Oskar Morgenstern (1944, 1953) *Theory of Games and Economic Behavior (3rd. edn.)*. Princeton: Princeton University Press.