Essays on Uncertainty in Economics

by

Peter Lewis Klibanoff

B.A., Harvard University (1990)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Author ....
Department of Economics
May 12, 1994

Certified by. ....
Drew Fudenberg
Professor of Economics, Harvard University
Thesis Supervisor

Certified by..
Jean Tirole
Professor of Economics
Thesis Supervisor

Accepted by..
Olivier J. Blanchard
Chairman, Departmental Committee on Graduate Students

Essays on Uncertainty in Economics

by

Peter Lewis Klibanoff

Submitted to the Department of Economics
on May 12, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Chapter 1 offers a new equilibrium concept for finite normal form games motivated by the idea that players may have preferences which display uncertainty aversion. Uncertainty aversion occurs when individuals have a dislike of uncertainty (not knowing the relevant probabilities in a given decision environment) and is distinct from their attitude towards risk (not knowing which outcome will occur, but knowing the probability of each outcome). More specifically, this chapter examines and interprets the representation of uncertainty averse preferences presented in Gilboa and Schmeidler (1989). Then an equilibrium with uncertainty aversion is defined and applied to a number of simple games. This equilibrium concept generalizes both Nash equilibrium and maxmin play. One interesting feature of the equilibrium is that it provides a new justification for some mixed strategy equilibria based on hedging. It also admits a natural channel through which some unmodelled aspects of the game can influence the analyst's choice of equilibrium. A refinement of equilibrium with uncertainty aversion incorporating the notion of common knowledge of rationality is introduced. The notion of weak admissibility is discussed and incorporated into the solution concept.

Chapter 2 develops a dynamically consistent theory of decision making that incorporates the notion of uncertainty aversion. A large body of experimental work has demonstrated the existence of uncertainty averse behavior. The theory developed here is needed if uncertainty aversion is to be used in modelling many interesting economic problems, as the existing static theory of uncertainty aversion cannot be extended in a dynamically consistent manner simply by updating beliefs. Additionally, this paper proves an extension to the case of an uncertainty averse searcher of a reservation-price rule result of Rothschild(1974) and Bikhchandani and Sharma (1989) in a model of price search without recall from an unknown distribution.

Chapter 3 investigates the interaction between externalities and uncertainty arising through private information in a decentralized setting. Uncertainty is treated in the traditional manner, assuming uncertainty neutrality.

In the competitive model, externalities lead to inefficiencies, and inefficiencies increase with the size of externalities. However, as argued by Coase, these problems may be mitigated in a decentralized system through voluntary coordination. Chapter 3 shows how coordination is limited by the combination of two factors: respect for individual autonomy and the existence of private information. Together they imply that efficient outcomes can only be achieved through coordination when external effects are *relatively large.* Moreover, unlike many previous mechanism design models of bargaining, there are instances in which coordination cannot yield any improvement at all, despite common knowledge that social gains from agreement exist. This occurs when external effects are *relatively small*, and this may help to explain why coordination is so seldom observed in practice. When improvements are possible, we describe how simple taxes or subsidies can be used to implement second-best solutions and explain why standard solutions, such as Pigouvian taxes, cannot be used. Possible extensions to issues arising in the structure of research joint ventures, assumptions in the endogenous growth literature, and the location of environmental hazards are also described.

Thesis Supervisor: Drew Fudenberg
Title: Professor of Economics, Harvard University

Thesis Supervisor: Jean Tirole
Title: Professor of Economics

Acknowledgments

I would like to thank some of the many people who enriched my life during the last four years.

Learning from and working with my advisors, Drew Fudenberg and Jean Tirole, was truly a privilege. Their keen insight, kindness, and superb skill as both teachers and researchers is inspiring. I benefitted greatly from their time.

My debt to Jonathan Morduch goes far beyond that owed for his collaboration on Chapter 3 of this thesis. He has been a great friend and mentor with whom I've had the pleasure of many hours of conversation.

My fellow students have been a large part of my social and intellectual life. For their companionship and stimulating interaction I would especially like to thank Stacey Tevlin, David Laibson, Owen Lamont, Fiona Scott-Morton, Koleman Strumpf, Bob Majure, Donald Marron, Chris Snyder, David Frankel, Guy Debelle, Matt Slaughter, Toshi Baily, and Glenn and Sara Ellison.

I thank Emily Gallagher for making MIT a little friendlier.

My roommate and friend, Carrie Suzawa, made home a nice place to be. My long-time friends and fellow followers of the academic road, Thomas Malaby and Nina Kushner, helped make Boston and Cambridge a great place to live.

I am grateful for a National Science Foundation Graduate Fellowship and an Alfred P. Sloan Foundation Doctoral Dissertation Fellowship which supported my studies.

Finally, my greatest thanks go to my family. To my sister Kathryn and my father Allen, for their tremendous support and love. I dedicate this thesis to them and to the memory of my mother, Patricia, for whom no words are enough.

Contents

1 Uncertainty, Decision, and Normal Form Games 6

1.1 Introduction 6

1.2 Decision Theory 8

1.3 Game Theory with Uncertainty Aversion 15

1.4 Adding Common Knowledge of Rationality 25

1.5 Weak Admissibility 29

1.6 Conclusion 36

2 Dynamic Choice with Uncertainty Aversion 42

2.1 Introduction 42

2.2 The Model 47

2.3 Axioms and a Static Theory 49

2.4 Dynamic Consistency and a Dynamic Representation 58

2.5 Search From an Unknown Distribution 64

2.6 Conclusion 71

3 Decentralization, Externalities, and Efficiency (with Jonathan Morduch) 76

3.1 Introduction 76

3.2 The Model 82

3.3 What Can Coordination Achieve? 89

3.3.1 Expected Costs and Benefits of Coordination 90

3.3.2 Characterization of Optimal Transfers 93

3.3.3 Obtaining First-Best Outcomes 94

3.3.4 Improvements on the Autonomous Allocation 95

3.3.5 Summary of Results 99

3.4 Interpretations and Extensions 101

3.4.1 Public Expenditures and Public Goods 101

3.4.2 Information Structure 102

3.4.3 Budget Balance 103

3.4.4 Multiple Projects 104

3.4.5 Multiple Spillovers 104

3.5 Applications 106

3.5.1 Research and Development 106

3.5.2 Models of Long-Run Growth 106

3.5.3 Siting a Toxic Waste Dump 107

3.6 Conclusion 108

A Proof of Theorems 5 and 6 of Chapter 1 **113**

B Appendicies for Chapter 3 **117**

B.1 Proof of Theorem 1 117

B.2 Proof of Proposition 1 122

B.3 Obtaining Ex Post Efficient Outcomes 123

B.4 Obtaining the Second-Best Outcome 125

B.5 Relaxing Budget Balance 127

Chapter 1

Uncertainty, Decision, and Normal Form Games

1.1 Introduction

Traditional decision theory and game theory have treated uncertainty (situations in which probabilities are unknown or subjective) with the same formalism as they have treated risk (situations where probabilities are known or objective); indeed, the word "uncertainty" is often used to describe both. This continues despite the fact that there is strong evidence which suggests that thoughtful decision-makers react to uncertainty differently than they react to risk.[1] The classic reference is the Ellsberg Paradox (Ellsberg 1961) a version of which may be demonstrated by the following choice situation:

An urn contains ninety balls, identical except for their color. Thirty of these balls are black. The remaining sixty are either red or yellow in unknown proportion. One ball will be drawn at random from the urn. You are asked to consider the six bets shown in the figure (see next page), whose payoffs depend on the color of the drawn ball.

The preference ordering of many decision-makers when faced with these bets is

[1] Knight (1921) and Shackle (1949, 1949-50) were among earlier economists who argued for a distinction between uncertainty and risk.

		Bets	
	Black	Red	Yellow
1	$100	$0	$0
2	$0	$100	$0
3	$0	$0	$100
4	$100	$0	$100
5	$100	$100	$0
6	$0	$100	$100

$6 \succ 5 \sim 4$ and $1 \succ 2 \sim 3$. This ordering cannot be reconciled with any subjective probability assessment. Moreover, as Ellsberg (1961) recounts:

"The important finding is that, after rethinking all their ’offending’ decisions in the light of the axioms, a number of people who are not only sophisticated but reasonable decide that they wish to persist in their choices. This includes many people who previously felt a ‘first-order commitment’ to the axioms, many of them surprised and some dismayed to find that they wished, in these situations, to violate the Sure-thing Principle." [p. 656]

I may note in passing that I count myself among those who display the Ellsberg preferences and who maintain these choices after contemplation of the axiomatic arguments in Savage (1954) and elsewhere. (Whether I am sophisticated and reasonable, I leave for others to decide.) Further, the fact that many people do not change their behavior even when confronted with their violation of the standard axioms distinguishes this behavior from some other types of violations such as intransitivity in choice.[2] Although intransitivities are observed experimentally, when the violations of transitivity are pointed out subjects often wish to change their choices so as to make them transitive. I would argue that theories of reasoned or rational behavior as well as purely descriptive theories should try to incorporate those types of violations which persist. The fact that many thoughtful people are not convinced by the arguments for the standard axioms should cause us to at least question their predominance in economic analysis.

[2] For experimental evidence on this point see e.g. Slovic and Tversky (1974).

Fortunately, Gilboa and Schmeidler (1989) (See also Schmeidler 1989, 1986 and Gilboa 1987) have recently developed an axiomatic decision theory which allows for Ellsberg-type preferences.[3] A common explanation for the Ellsberg preferences is that decision makers dislike uncertainty or ambiguity. This is consistent with the fact that bet 1 (which has a known probability of one-third of paying $100) is preferred to bets 2 and 3 and bet 6 (which pays $100 with probability two-thirds) is preferred to the uncertain bets 4 and 5.[4] Thus Gilboa and Schmeidler view their theory as allowing for uncertainty aversion on the part of the decision maker. In the next section, I briefly present the Gilboa-Schmeidler theory, put forth an interpretation and show how the theory applies to the Ellsberg example. In the third section I present a new solution concept for normal form games in which players are Gilboa-Schmeidler decision makers. This section also contains examples to which the concept is applied. The fourth section presents a refinement of the solution concept and some more examples. The fifth section reconsiders the theory and proposes a modification which is then applied to games. The sixth section concludes. An Appendix contains some proofs.

1.2 Decision Theory

The workhorse of decision theory in economics since von Neumann and Morgenstern (1947) has been axiomatic representation theorems. The Gilboa-Schmeidler theory is in this spirit and adopts the "lottery-acts" framework of Anscombe and Aumann (1963). Let X be a set of "prizes" (e.g. cash rewards). Let Y be the set of distributions over X with finite support. We call elements of Y lotteries. Let S be a set and let Σ be an algebra on S. Elements of Σ are called events, while elements of S are states

[3] Some alternative theories and further experimental evidence are described in the survey paper by Camerer and Weber (1992).

[4] One important question which Ellsberg's example does not address is whether this uncertainty aversion is more than lexicographic. In other words, would a decision maker be willing to give up anything to avoid uncertainty? Ellsberg himself (1961, p.664) provides evidence for this when he reports that many subjects :naintain the above preferences even after one black ball is removed from the urn. Many subsequent studies (cited in Camerer and Weber (1992)) have found ambiguity premia which are strictly positive and are typically around $10-20\%$ in expected value terms.

of the world. Let L_0 be the set of Σ-measurable finite step functions from S to Y. Let L_c be the set of constant functions in L_0. Let L be a convex subset of Y_S which includes L_c. We call functions from S to Y acts. Preferences will be defined over acts. Consider the following axioms on the preference relation $\succeq$ on L:

A.1 (Weak Order) (a) $\forall f, g \in L, f \succeq g$ or $g \succeq f$ or both.

(b) $\forall f, g, h \in L, \{f \succeq g$ and $g \succeq h\} \Rightarrow f \succeq h$.

A.2 (Certainty Independence)

$\forall f, g \in L$ and $h \in L_c$ and $\alpha \in (0,1), f \succ g \Leftrightarrow \alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.

A.3 (Continuity)

$\forall f, g, h \in L$, if $f \succ g$ and $g \succ h$ then $\exists \alpha, \beta \in (0,1)$ such that

$\alpha f + (1-\alpha)h \succ g$ and $g \succ \beta f + (1-\beta)h$.

A.4 (Monotonicity)

$\forall f, g \in L$, if $f(s) \succeq g(s)$ on S then $f \succeq g$.

A.5 (Uncertainty Aversion)

$\forall f, g \in L$ and $\alpha \in (0,1)$, if $f \sim g$ then $\alpha f + (1-\alpha)g \succeq f$.

A.6 (Non-degeneracy)

Not for all $f, g \in L, f \succeq g$.

The only non-standard axioms are Certainty Independence and Uncertainty Aversion. Note that Certainty Independence is a strict weakening of the traditional Independence axiom when applied to the lottery-acts framework, as it requires that strict preference be preserved only under mixtures with constant acts. Gilboa and Schmeidler defend Certainty Independence by the argument that the decision maker can more easily visualize mixtures with a constant act than with an arbitrary one and the fact that this axiom allows for hedging whereas the standard Independence axiom in this context would not. How are we to understand what is meant by hedging here? I interpret hedging in this context to mean that, even if bets (or assets) paid off in utility terms (so that risk-aversion is controlled for) a bettor (or investor) might prefer to spread the payoffs over several states of the world than have them all lumped in one state. In this light, the axiom of Uncertainty Aversion can be interpreted as saying that the decision maker does not dislike hedging.

The main result of Gilboa and Schmeidler is the following representation theorem:

Theorem 1 *(Gilboa and Schmeidler 1989)*

Let $\succeq$ be a binary relation on L_0. Then the following are equivalent,

(1) $\succeq$ satisfies A.1 - A.5 for $L = L_0$

(2) $\exists$ an affine function $u : Y \rightarrow \mathcal{R}$ and a non-empty, closed, convex set C of finitely additive probability measures on Σ such that $\forall f, g \in L_0$, $f \succeq g$ if and only if $\min_{p \in C} \int u \circ f dp \geq \min_{p \in C} \int u \circ g dp$.

Furthermore, the function u is unique up to a positive affine transformation and, if and only if A.6 holds, the set C is unique.

The reader is referred to the paper for the proof.[5] It is also shown there that the theorem can be extended to preferences over all Σ-measurable bounded acts. Gilboa and Schmeidler do not interpret the set of measures, C, which appears in the representation. For reasons which will become clear later on, I would like to interpret C as the closure of the convex hull of the set of "possible" subjective probability distributions from the decision maker's point of view. Is there any justification for this in the way that C is constructed in the proof? I believe that there is. Imagine a fixed choice environment (i.e. a set S, and an algebra Σ). Now consider the space B of bounded, Σ-measurable functions from S to $\mathcal{R}$. These functions can be thought

[5] The first step in the proof is to observe that, as constant acts may be identified with the choice set in the von Neumann-Morgenstern setting, axioms A.1-A.3 applied to constant acts give the function u through the von Neumann-Morgenstern expected utility theorem. For this reason, I interpret this theory as implying that a decision maker behaves as an expected utility maximizer in situations where the probabilities are objective (i.e. where only risk but not uncertainty is present). This is not the only possible interpretation of the Gilboa-Schmeidler theory. Other researchers (see Quiggin 1982, Yaari 1987, Wakker 1990, among others; Fishburn 1988, chapt. 2 has a survey) have interpreted representations which are special cases of the related non-additive representation of Schmeidler (1989) as models of decision making under risk where the probabilities are distorted by the decision maker. As the representation considered above is isomorphic to the Schmeidler (1989) representation under certain conditions, such an interpretation could also be applied here. However, although these interpretations are attractive from a descriptive point of view (e.g. are consistent with the Allais Paradox (Allais 1953)), they do not seem normatively compelling as they operationally require a decision maker to take perfectly known, objective probabilities and distort them before using them to weight outcomes. This seems much more objectionable than allowing that, in situations where probabilities are not known, the decision maker may not act as if he subjectively "knows" the probabilities (i.e. has a unique subjective probability distribution in mind). More importantly for our purposes, perhaps, this interpretation does not allow for Ellsberg-type behavior.

of as a superset of the utility payoffs from acts (i.e. functions of the form $u \circ f$). Thus, if we can find a functional I on B such that, $\forall f, g \in L_0, f \succeq g$ if and only if $I(u \circ f) \geq I(u \circ g)$, where u is as described above, then this functional can be said to represent preferences. The key to the Gilboa-Schmeidler proof is showing that if preferences satisfy the axioms, there exists an I that has the nice property that, for each $b \in B$, there exists a finitely additive probability measure P_b such that $I(b) = \int b dP_b$ and $I(a) \leq \int a dP_b$ for all $a \in B$. The set C is then defined as the closure of the convex hull of these P_b.[6] Thus, C is the closure of the convex hull of exactly those probability measures which are used in place of an objective probability measure in valuing some subset of mappings from events to payoffs. In what sense then is C the set of "possible" subjective probability measures? In the standard theory of decision under uncertainty, a probability measure q is said to be the subjective probability measure of a decision maker when she behaves as if she were maximizing the expectation of her affine utility function on lotteries (elicited using standard techniques and objective probabilities) where the expectation is taken with respect to q. Here, a set C is the set of "possible" subjective probability measures when it is the closed, convex hull of those measures which are used to calculate expected utility for some acts and when the choice of which possible measure to use for which act is governed by which possible measure gives the minimum expected utility for that act. Notice that in expanding from a single measure to a set of measures I have had to specify a rule for assigning measures to acts. This is very important in obtaining a notion of the (i.e. unique) set of possible measures. If one were to allow both the set of measures and the assignment rule to vary, then there would be different sets of "possible" measures for different assignment rules even though the preferences being represented were not changing. Thus we must keep in mind that my interpretation of "possible" is contingent on the adoption of the G-S assignment rule.

Why consider the closed, convex hull? Considering the convex hull makes sense in the framework of possible beliefs or multiple priors since the crux of the individual's

[6] This is not quite correct, in that their construction of C uses only those measures that correspond to b such that $I(b) > 0$. However, properties of I can then be used to show that for any b the resulting set contains elements which satisfy the integral representation for that b.

problem is that he does not know how much to weight the different priors. If he knew this he could combine them into one prior as in the standard theory. Furthermore, even if C were not convex, the preferences under C would be identical to the preferences under the convex hull of C. Thus the two are not distinguishable in this setting. Closure seems like a more technical requirement although it does not seem unreasonable.

Now that we have adopted an interpretation of the Gilboa-Schmeidler result we can begin to explore the consequences of relaxing the standard theory in this way. Although the main focus of this paper is on the consequences for game theory, it is worthwhile to briefly explore some decision theoretic concerns. Does the G-S theory resolve the Ellsberg paradox and if so does it suffer from the traditional criticisms of Ellsberg's work? Recall the thought experiment described above. Suppose that the decision maker in the experiment is uncertainty averse and thus is unwilling to use a unique probability measure in situations where uncertainty is present. In the experiment as described, there is certainly substantial uncertainty in that the decision maker is given no information about the relative proportions of the three colors aside from the fact that one-third of the balls are black. A very natural candidate for the set C in this case is the set of all probability distributions over the three colors which assign probability one-third to black. Given this set, the reader may verify that a Gilboa-Schmeidler decision maker will display precisely the Ellsberg preferences. In fact, as long as the decision maker's set of possible probability measures includes at least one in which Prob(red) $>$ Prob(yellow) and another in which the reverse is true (assuming that all assign one-third to black), the decision maker will have these same preferences. Thus, in this very clear sense, it is the uncertainty about whether there are more red balls or yellow balls combined with the individual's dislike of uncertainty which results in the Ellsberg behavior.

Some observers have argued that the Ellsberg Paradox simply points out the need to teach people to obey the axioms of Savage-Anscombe-Aumann decision theory by presenting them with compelling examples that will persuade them that treating subjective probability differently than objective probability is a mistake. A leading proponent of this position is Howard Raiffa. In a comment (Raiffa 1961) published along with Ellsberg's original article, he uses two examples similar to ones offered by Ellsberg to make his argument.[7] In the first example, two questions are asked of a decision maker. First, the decision maker is asked to consider an urn containing fifty red balls and fifty black balls and to name the dollar amount that he would pay to be allowed to name a color and receive one hundred dollars if a ball drawn at random is of the named color. Raiffa reports that the amounts given clustered around thirty dollars (thus displaying risk aversion). These same decision makers were then asked to say how much they would pay for the same opportunity with an urn which contains red and black balls in unknown proportion. The answers in this case typically involved much smaller dollar amounts, thus violating the standard axioms. In subsequent discussion, Raiffa finds that the following argument convinces people to change their answer to the second question so that it matches their answer to the first question: Suppose that in the second setting you draw a ball at random and do not examine its color. Then flip a coin and say "red" if heads and "black" if tails. Notice that this results in an objective probability of winning of one-half independent of the true proportions of red and black balls. Certainly, it should not matter whether the ball is drawn before or after the coin is flipped since the processes are physically independent. Thus the second option (unknown proportions) should always be worth at least as much as the first option (known 50-50) since a coin flip can transform the second into the first.[8]

I find this argument compelling, and fully agree that an individual who values the second option less than the first is not acting rationally in the sense that once she thinks the problem through carefully (and either discovers or has pointed out to her the strategy of flipping a coin to decide) she will revise her decision. My complaint with this argument is that it does not contradict the results of Ellsberg in a similar experiment. In Ellsberg's version the set-up is the same but individuals are not given as much freedom in that they are asked to make specific pairwise comparisons be-

[7] See also chapter 5 in Raiffa 1968.

[8] Throughout this paper, as in decision theory and game theory generally, it is assumed that participants have costless access to independent, privately-observable randomizing devices.

tween bets. Thus many individuals say they prefer betting on red in the known urn to betting on red in the unknown one and prefer betting on black in the known urn to black in the unknown urn. Notice that these responses clearly violate the standard axioms but cannot be remedied by randomizing since subjects are asked to compare two fixed bets, whereas Raiffa is asking them to compare two betting environments. Specifically, the reader can check that a decision maker whose preferences are consistent with the Gilboa-Schmeidler axioms will *always* value Raiffa's unknown urn option at least as much as they value the known urn option, and may at the same time prefer any fixed bet on the known urn to the same bet on the unknown one.

Similarly, Raiffa argues that the Ellsberg preferences in the thought experiment in section 1.1 imply that the decision maker would prefer a 50-50 lottery between acts 1 and 6 to a 50-50 lottery between acts 3 and 5. He then points out, correctly, that these two lotteries result in objectively equal outcomes. Again, Raiffa's assumption is that each act in a lottery can be evaluated independently of the acts it is being mixed with. Thus, for this example to be convincing, the decision maker must have already accepted the independence axiom for acts, or no contradiction is implied. A Gilboa-Schmeidler decision maker is indifferent between the two lotteries even if she prefers 1 to 3 and 6 to 5.

The reason for this is simply that randomization between bets which pay off in different states of the world helps to reduce uncertainty by spreading the utility over more states, which is exactly what an uncertainty averse decision maker would like to do. Thus it is possible for such an individual to strictly prefer the randomization over two bets to either of the bets themselves. This feature of uncertainty aversion will play an important role in our discussion of game theory. We note that such a preference for randomization raises the issue of dynamic inconsistency, in the sense of wanting to randomize again once the outcome of the original randomization is known. However, Machina (1989) surveys many such dynamic inconsistency objections to non-expected utility theories and argues that this notion of consistency is inappropriate for such decision makers. The flavor of his argument can be expressed here by the notion that strictly preferring a randomization over two acts, A and B, includes the preference for act A over any mixture *conditional* on having borne a risk of B. Thus if the result of the randomization was A, the individual would indeed be willing to perform A. A similar argument is made for B. The reader who is unconvinced by Machina's arguments may want to think of the decision and game situations we will look at as situations in which the participants have available some means of committing to a mixed strategy. For example, they may be giving instructions to agents, or may be able to buy shares of more than one asset, or place a bet on more than one outcome.

Now that we have discussed and interpreted the preferences, some implications for game theory can be explored.

1.3 Game Theory with Uncertainty Aversion

Game theoretic situations are rife with uncertainty. Almost never can a player publicly commit to playing a given strategy. Thus, from the point of view of his opponent(s), there will often be great uncertainty about what this strategy will be. Much of game theory can be viewed as the search for concepts which narrow this uncertainty in convincing ways. Nash equilibrium, the leading solution concept for non-cooperative games, does this by combining two fundamental ideas. First, it borrows from decision theory the idea that rational players will choose a strategy which is the most preferred given their beliefs about what other players will do. Second, it imposes the consistency condition that all players' beliefs are, in fact, correct. One major criticism of Nash equilibrium has been the strength of the consistency condition. In many settings it is far from clear that players will have exactly correct beliefs about each other.[9] Moreover, even if it is common knowledge that all players in a game believe that Nash equilibrium is the proper concept to use in determining their beliefs about play, the problem of multiple Nash equilibria remains. In a game with multiple Nash equilibria, even if the players themselves accept (and are commonly known to accept) the Nash solution concept they may still face substantial

[9] For some work which has investigated conditions under which this will be true in particular repeated game settings see Fudenberg and Levine 1990 and Fudenberg and Kreps 1991.

uncertainty about the play of their opponents. Thus, there is wide scope for players' behavior under uncertainty to affect the conclusions of game theory. In this vein, I propose a solution concept for normal form games which generalizes the notion of Nash equilibrium and allows for players whose preferences can be represented as in theorem 1 above.

In related work, Dow and Werlang (1991) use Schmeidler's (1989) non-additive measure formulation to put forth a generalization of Nash equilibrium. Although the motivation for their work is similar, both the equilibrium concept and its implications differ from the ones proposed here. Additonally, they focus on the implications for backwards induction while I limit myself to static settings due to the difficulties with dynamic consistency inherent in Gilboa and Schmeidler's (1989) or Schmeidler's (1989) preferences (see Epstein and Le Breton (1993) and Klibanoff (1993a, b)).

Fix a finite normal form game G (i.e. a finite set of players $\{1, 2, \ldots, I\}$, a pure strategy space S_i for each player i such that $S = \times_i S_i$ is finite, and payoff functions $u_i : \times_i S_i \rightarrow \mathcal{R}$ which give player i's von Neumann-Morgenstern utility for each profile of pure strategies).

Definition: An *equilibrium with uncertainty aversion* of G is a $2 * I$-vector $(\sigma_1, \ldots, \sigma_I, B_1, B_2, \ldots, B_I)$ where $\sigma_i \in \Sigma_i$ (the set of mixed strategies for player i, i.e. the set of probability distributions over S_i) and the B_i are closed, convex subsets of P_{-i} (the set of probability distributions over $\times_{k \neq i} S_k$) such that, for all i,

(1) σ_i satisfies

$\min_{p \in B_i} \sum_s u_i(s_i, s_{-i}) \sigma_i(s_i) p(s_{-i}) \geq \min_{p \in Bi} \sum_s u_i(s_i, s_{-i}) \sigma'_i(s_i) p(s_{-i})$

for all $\sigma'_i \in \Sigma_i$, and

(2) $\prod_{k \neq i} \sigma_k(s_k) \in B_i$.

Condition (2) relaxes the consistency condition imposed by Nash equilibrium. It says that each player's beliefs must not be mistaken, in the sense that they contain the truth. More specifically, the truth must be contained in the closed, convex hull of the set of possible subjective probability distributions over the strategies of the other players. Condition (1) simply says that each player's strategy is optimal given her beliefs, assuming that her preferences can be represented as in theorem 1.

One important thing to note about the sets of beliefs is that eiements of these sets may allow for correlation between the strategies of the other players even though we require the true strategies to be independent. To see how this might arise, consider a three player game where each player may move either right or left. Player one might well believe that either two and three will both play right or two and three will both play left (because, for example, one knows that two and three grew up with the same social norm but one does not know what that norm is). Any convex combination of these two priors could only arise from correlation between two and three. In this way, one's uncertainty introduces subjective correlation into his beliefs even though he knows that only independent mixing is allowed.

Two polar special cases of this definition – when all the B_i's are singletons and when all the B_i's equal P_{-i} – yield familiar concepts as we observe in the following theorem.

Theorem 2 *(a) In any finite normal form game, a strategy profile σ is part of an equilibrium with uncertainty aversion where B_i is a singleton for all i if and only if σ is a Nash equilibrium profile.*

(b) In any finite normal form game, a strategy profile σ is part of an equilibrium with uncertainty aversion where $B_i = P_{-i}$ for all i if and only if, for all i, σ_i is a maximin strategy (i.e. a strategy which maximizes i's minimum payoff given that any opponents' play is possible).

Proof: (a) and (b) follow directly from the definition of equilibrium with uncertainty aversion. *QED*

Theorem 2 shows that equilibrium with uncertainty aversion spans the continuum between all players playing maximin strategies, a criterion often advocated in situations of complete ignorance, and Nash equilibrium where all players behave as if they had perfect knowledge of their opponents' strategies. Exactly when preferences in Theorem 1 coincide with subjective expected utility maximization, equilibrium with uncertainty aversion coincides with Nash equilibrium. Existence of an equilibrium with uncertainty aversion follows from the existence of a Nash equilibrium (Nash 1950).

The next observation shows that equilibria with uncertainty aversion are not often unique. This is to be expected as they are, by construction, very dependent on beliefs.

Observation 1 *A finite normal form game has a unique equilibrium with uncertainty aversion only if it has a unique Nash equilibrium and that Nash equilibrium consists of each player playing their unique maximin strategy.*

Proof: Any Nash equilibrium is an equilibrium with uncertainty aversion. From theorem 2, each player playing a maximin strategy is an equilibrium with uncertainty aversion. *QED*

Note that the converse is false, as is shown by the game in figure 1.

		Player 2		
		X	Y	Z
Player 1	A	1,2	4,3	1,4
	B	1,2	3,3	3,1
	C	2,2	4,1	2,1

figure 1

In this game, the unique Nash equilibrium is (C, X), which is also the unique maximin profile. However, (B, Y) is an equilibrium with uncertainty aversion if player 1 has a belief set consisting of all distributions over Y and Z, while player 2 has a belief set consisting of all distributions over A and B.

The best way to see the implications of this definition is through some examples. To keep things simple I will focus on 2 x 2 games. Consider the pure coordination game in figure 2.

This game has three Nash equilibria, (U, L), (D, R), and (1/3 U, 2/3 D; 1/3 L, 2/3 R). Let us focus on the mixed equilibrium. Notice that in the Nash setting, each player is indifferent between any pure or mixed strategy given their beliefs. Thus

	L	R
U	2,2	0,0
D	0,0	1,1

figure 2

there is no affirmative reason to mix with these proportions. This need not be true with uncertainty aversion. For example, if each player's belief set, B_i, consists of all mixtures over their opponents' pure strategies (as it would, for instance, if players are uncertainty averse and their sets of possible subjective probability measures include the Nash beliefs) then each player will strictly prefer to play the mixed strategy. This is true because by equalizing the payoff to, say, U and D under any distribution over L and R, the uncertainty is eliminated and the maximin payoff is achieved. In fact, as long as player 1 has some belief which assigns Prob(L) $<$ 1/3 and some belief which assigns Prob(L) $>$ 1/3, the mixed strategy is his strict best response. Similarly, if player 2 has some belief which assigns Prob(U) $<$ 1/3 and one which has Prob(U) $>$ 1/3, the mixed strategy is the strict best response. Thus equilibrium with uncertainty aversion can justify mixing as a response to strategic uncertainty. In contrast with Harsanyi's (1973) view of mixed equilibria as the limits of pure strategy equilibria in perturbed games, our setting allows common knowledge of payoffs to be taken seriously. Another advantage of this view of mixed strategies is that it can provide information about the likelihood or robustness of a mixed strategy outcome. Consider the game in figure 3, which has been commented on extensively in the literature.

	L	R
U	9,9	0,8
D	8,0	7,7

figure 3

In this game, unlike the game in figure 2, a mixed strategy will never be strictly preferred in equilibrium. This can be seen by noting that if any subjective distribution gives weight more than 1/8 to D (or R) then the best response is to play R (or D) and if all distributions give weight less than 1/8 to D (or R) then the best response is L (or U). The only beliefs for which mixing can occur in equilibrium are those which include giving weight 1/8 to D(or R) and possibly include some distributions which give weight less than 1/8 to D(or R). However mixing is not strictly preferred for these beliefs.

These two examples suggest that equilibrium with uncertainty aversion highlights mixed equilibria in some games but not in others. We would like to understand what it is about the game in figure 2 which leads to the possibility of a strict mixed equilibrium. Observe that, for player 1, U does better if 2 plays L while D does better if 2 plays R. Thus, as long as U does not weakly dominate D or vice-versa, a mixture over U and D will do better than D against L and will do better than U against R. Since an uncertainty averse individual cares about the minimum expected utility over her belief set, it is easy to see that mixing can raise this minimum as compared to either pure strategy for some beliefs.

In the game in figure 3, however, both U and D do better if 2 plays L. In this case, since both pure strategies are lower under R than under L, a mixed strategy will never raise the minimum expected utility compared to each of the pure strategies. More generally, if the expected payoffs to any two pure strategies are minimized (over B_i) by the same distribution $p \in B_i$, a mixture of the two will never be strictly preferred to each pure strategy by an uncertainty averse decision maker. This condition is only sufficient, however. This is easily seen by considering one strategy which strictly dominates another, but which is not minimized by the same distribution as the other. No mixing involving the dominated strategy will ever be preferred to the undominated strategy, yet these two strategies are not minimized by the same distribution. The following theorem gives necessary and sufficient conditions for not strictly preferring a mixture of two strategies to each strategy itself. In other words, these conditions characterize exactly when there is no gain to hedging between two strategies.

Theorem 3 *Fix a player i and two strategies σ_i and σ_i' such that $\sigma_i \succeq \sigma_i'$ (i.e. the* *minimum expected utility of σ_i is at least as big as the minimum expected utility of σ_i'). No mixture over σ_i and σ_i' will be strictly preferred to both σ_i and σ_i' if and only if there exists some $q \in B_i$ such that q minimizes the expected utility of σ_i and such that $\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) \geq \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)q(s_{-i})$.*

Proof: (sufficiency) Let there be such a q. Then the minimum expected utility of $\sigma_i = \sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) \geq \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)q(s_{-i}) \geq$ minimum expected utility of σ_i'. Therefore $\sum_s u_i(s_i, s_{-i})(\alpha\sigma_i(s_i) + (1-\alpha)\sigma_i'(s_i))q(s_{-i}) \leq$

$\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) =$ minimum expected utility of σ_i. This implies that $\sigma_i \succeq \alpha\sigma_i + (1-\alpha)\sigma_i'$ for all $\alpha \in (0,1)$.

(necessity) Assume that no mixture is strictly preferred and suppose, to the contrary, that for all $q \in B_i$ such that q minimizes the expected utility of σ_i it is true that $\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) < \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)q(s_{-i})$. Then for any such q, $\sum_s u_i(s_i, s_{-i})(\alpha\sigma_i(s_i) + (1-\alpha)\sigma_i'(s_i))q(s_{-i}) > \sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i})$ for all $\alpha \in (0,1)$. Now consider any $q^* \in B_i$ that does not minimize the expected utility of σ_i. By uniform continuity, there exists a $\delta > 0$ such that if $||q^* - q|| < \delta$ for a q which minimizes the expected utility of σ_i, then $\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q^*(s_{-i}) < \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)q^*(s_{-i})$. For any such q^* (i.e. one within δ of a minimizer),

$\sum_s u_i(s_i, s_{-i})(\alpha\sigma_i(s_i) + (1-\alpha)\sigma_i'(s_i))q^*(s_{-i}) > \sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i})$ for all $\alpha \in (0,1)$.

By definition of a minimizer, there exists an $\epsilon > 0$ such that for any $q^* \in B_i$ such that $||q^* - q|| \geq \delta$ for all q which minimize the expected utility of σ_i it is true that $\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q^*(s_{-i}) > \sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) + \epsilon$. Thus, for $\underline{\alpha}$ such that $\underline{\alpha}\epsilon + (1-\underline{\alpha})(\min_{p \in B_i} \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)p(s_{-i}) - \sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i})) = 0$, (which exists and is strictly less than one since the first term is positive and the second term is non-positive), it is true that for all $\alpha > \underline{\alpha}$, $\alpha\sigma_i + (1-\alpha)\sigma_i' \succ \sigma_i \succeq \sigma_i'$. This contradicts the assumption that no mixture of σ_i and σ_i' is strictly preferred to both strategies. This proves necessity. *QED*

In applying Theorem 3, it is often easier to check the sufficient condition mentioned above and given in the following corollary.

Corollary 3.1 *Fix a player i and two strategies σ_i and σ_i'. No mixture over σ_i and*

σ_i' will be strictly preferred to both σ_i and σ_i' if there exists some $q \in B_i$ such that q minimizes the expected utility of both σ_i and σ_i'.

Proof: Assume without loss of generality that $\sigma_i \succeq \sigma_i'$. Such a q then satisfies the conditions of Theorem 3. *QED*

This sufficient condition becomes even easier to check in 2 x 2 games, as reference to particular beliefs B_i can be omitted.

Corollary 3.2 *Fix a player i in a 2 x 2 game. If there is a pure strategy of i's opponent which minimizes the payoff to both of i's pure strategies, then i will never strictly prefer a mixed strategy to both of i's pure strategies.*

Proof: Call the pure strategies of *i*'s opponent a and b. Suppose that a minimizes the payoff to both of *i*'s pure strategies. No matter what *i*'s set of beliefs is, each of *i*'s pure strategies will have its expected utility minimized by the distribution in the belief set which puts the most weight on a. Therefore the existence of a q satisfying the conditions in Corollary 3.1 is guaranteed for any belief set. *QED*

In the special case of $\sigma_i \sim \sigma_i'$, the sufficient condition of Corollary 3.1 is also necessary.

Corollary 3.3 *Fix a player i and two strategies σ_i and σ_i' such that $\sigma_i \sim \sigma_i'$. No mixture over σ_i and σ_i' will be strictly preferred to both σ_i and σ_i' if and only if there exists some $q \in B_i$ such that q minimizes the expected utility of both σ_i and σ_i'.*

Proof: If $\sigma_i \sim \sigma_i'$ then the minimum expected utilities of σ_i and σ_i' must be equal. The only way to satisfy $\sum_s u_i(s_i, s_{-i})\sigma_i(s_i)q(s_{-i}) \geq \sum_s u_i(s_i, s_{-i})\sigma_i'(s_i)q(s_{-i})$ for a q which minimizes the left-hand side is to have the same q also minimize the right-hand side. *QED*

We can illustrate Theorem 3 (and Corollary 3.2 in particular) by again considering the game in figure 2. Suppose we modify this game by increasing player 1's payoff by one util when 2 plays L and increasing player 2's payoff by one util when one plays U. The modified game is as in figure 4.

Noting that each players' pure strategies now have their payoffs minimized by the distribution placing the most weight on R (or D), Corollary 3.2 tells us that a mixed strategy will never be strictly preferred. This contrasts with the earlier analysis of the

	L	R
U	3,3	0,1
D	1,0	1,1

figure 4

game in figure 2, in which mixed strategies were strictly optimal for a wide range of beliefs. In comparing the two games, the reader can check that not only are the Nash equilibria unchanged, but each player's best response correspondence is unchanged as well.[10] However the equilibria with uncertainty aversion are affected.

What has happened, intuitively, is that the change in payoffs has turned a game in which mixing helped hedge against uncertainty into one where it cannot play that role. On a more formal level, these changes have no effect in the standard theory because the independence axiom requires that preference between two acts (strategies) be preserved when they are mixed with a common third act. In the setting of Theorem 1 however, the independence axiom need only hold for mixing with constant acts, whereas adding one to player 1's payoff if 2 plays L, for example, is mixing the existing acts with a *non-constant* act. Thus such a transformation may change behavior.

The notion that this generalization of the Nash concept may allow for a natural way of refining predictions about the outcome of a game is another advantage of this approach. I view this equilibrium notion as allowing sharper prediction in the sense that it allows the use of information about players' beliefs in a way that the Nash concept does not. For example, in figure 3, if the players were known to be uncertainty averse and there was no compelling unmodelled feature of their environment which would lead each to include the mixed Nash strategy in their beliefs but not include any distribution which puts more weight on R (or D) (than the mixed Nash strategy),

[10] I use best response correspondence in the standard sense of a player's optimal strategy as a function of the opponents' strategies. An alternative notion of best response correspondence, defined as a player's optimal strategy as a function of that player's *beliefs* about the opponents' strategies, would, of course, give different correspondences for the two games.

I would be very reluctant to predict the mixed Nash equilibrium as the outcome of the game. Furthermore, in situations where the players are likely experiencing substantial uncertainty about others' play (for example, if they have never previously met their opponent and have not played the game before), I would be tempted to predict (D,R) as the outcome of the game in figure 3. The reasoning behind this is that greater uncertainty will be reflected in a larger set of beliefs, and thus (D, R) becomes more likely in the sense that if any belief assigns Prob(D) (or R) > 1/8, the player's best response switches to R (or D).[11] Thus a compelling feature of equilibria with uncertainty aversion is that "comparative statics" in uncertainty becomes possible in a well-defined sense.[12]

The set of equilibria with uncertainty aversion has been contained in the set of rationalizable[13] outcomes in the examples we have seen so far. This is not necessarily the case. Consider the game in figure 5.

	L	R
U	3,0	1,2
D	0,4	0,-100

figure 5

[11] There are other reasons why (D, R) is an attractive prediction in this game. Both the risk-dominance criterion of Harsanyi and Selten (1988) and Aumann's (1990) argument that pre-play communication is not likely to assist in coordination on (U, L) also lead to a prediction of (D, R). Note that the notion of risk-dominance in 2x2 games shares some of the flavor of uncertainty aversion but differs in important ways. Risk-dominance always produces a unique prediction, while equilibria with uncertainty aversion depend on players beliefs and uncertainty aversion. Furthermore, although figure 3 and heuristic considerations might lead one to think that the risk-dominant equilibrium is always the same as the equilibrium with uncertainty aversion when there is maximal uncertainty (or ignorance), this is not true. In figure 2, (U, L) is risk-dominant while the mixed strategy pair is picked out under ignorance and uncertainty aversion.

[12] This aspect of the theory could conceivably be tested in an experimental setting. After assessing subjects' utility functions (using objective probabilities) and using examples like those of Ellsberg to detect aversion to uncertainty, the experimenter would have the subjects play simple games. The level of uncertainty in their beliefs about their opponent could be manipulated by, say, providing or not providing a record of past games the opponent played; allowing or not allowing pre-play discussion or face-to-face contact etc. Subjects might also be asked to explicitly describe (ex-ante or ex-post) their beliefs about their opponent's play.

[13] For a definition of rationalizability see Bernheim (1984) and Pearce (1984) who introduced the concept, or Fudenberg and Tirole (1991), chapter 2.

In this game the unique Nash equilibrium is (U, R). This outcome can also be found by iterated strict dominance and is thus the unique rationalizable outcome as well. However, if any of player 2's subjective probability measures assigns weight at least 1/53 to D, then (U, L) will be an equilibrium with uncertainty aversion. In fact, letting 2's payoff from (D, R) approach $-\infty$, 2 will have to put an arbitrarily high minimum probability on U to be willing to play R.

This example makes several important points: (1) the set of equilibrium outcomes with uncertainty aversion is not in general contained in the set of rationalizable outcomes; (2) as Fudenberg and Tirole (1991, chapter 1) discuss, predicting (U,R) in a game like figure 5 relies crucially on the assumption that it is common knowledge that dominated (or non-rationalizable) strategies will never be used; and (3) to the extent that this common knowledge assumption is appropriate, the concept of equilibrium with uncertainty aversion may be too weak. In the next section, I pursue this line of reasoning by proposing a refinement of equilibrium with uncertainty aversion.

1.4 Adding Common Knowledge of Rationality

Consider the following definition that is motivated by the concept of correlated rationalizability (Pearce 1984, Brandenburger and Dekel 1987, Fudenberg and Tirole 1991, chapter 2). It is a natural generalization to the context where players can be described as in Theorem 1. The idea is to start from the whole set of strategies and eliminate, in each round of iteration, those strategies which are never a best response in the sense of Theorem 1 when the set of beliefs B_i is restricted to those beliefs which are compatible with the knowledge that other players only play best responses to the restricted sets of beliefs derived in the previous round. Thus, in the first round of iteration, those strategies which are never best responses to any beliefs are eliminated. In the second round, any strategies from the remaining set that are never best responses to any beliefs concentrated on that remaining set are eliminated, and so on. The successive rounds of iteration capture successive layers of knowledge of the rationality (in the sense of Theorem 1) of the players. Assume that all payoffs are common knowledge. Then the first iteration corresponds to the assumption that each player is rational. The second iteration corresponds to the assumption that each player is rational and knows that the other players are rational. The nth iteration corresponds to the assumption that each player is rational and knows that the other players know that the players know ... that the players are rational, where n-1 levels of knowledge are assumed.

Definition: Set $\Sigma_i^0 = \Sigma_i$, $P_{-i}^0 =$ the set of probability measures on $\times_{k \neq i} S_k$ such that for each $k \neq i$ the marginal distribution over S_k is an element of Σ_i. Recursively define for each integer $m > 0$:

$\Sigma_i^m = \{\sigma_i \in \Sigma_i^{m-1}$ such that there exists a closed, convex subset, B_i, of P_{-i}^{m-1} such that σ_i satisfies condition (1) in the definition of equilibrium with uncertainty aversion with Σ_i^{m-1} replacing Σ_i.$\}$, and

$P_{-i}^m =$ the set of probability measures on $\times_{k \neq i} S_k$ such that for each $k \neq i$ the marginal distribution over S_k is an element of the convex hull of Σ_k^m.

The *uncertainty aversion rationalizable strategies* for player i are $R_i = \bigcap_{m=0}^{\infty} \Sigma_i^m$.

The *uncertainty aversion rationalizable belief set* for player i is $Q_i = \bigcap_{m=0}^{\infty} P_{-i}^m$.

An alternate and often more useful characterization can be given in terms of iterated deletion of dominated strategies.

Theorem 4 *In finite normal form games the uncertainty aversion rationalizable strategies for player i, R_i, are exactly those strategies for player i which survive iterated deletion of strictly dominated strategies, (denoted by I_i).*

Proof: The definition of uncertainty aversion rationalizable strategies is equivalent to that of correlated rationalizable strategies when the set B_i is restricted to be a singleton. Since the set of correlated rationalizable strategies for player i is identical to the set of strategies for player i which survive iterated strict dominance (see Fudenberg and Tirole's (1991, chapter 2) modification of a proof by Pearce (1984)), R_i is a superset of I_i. As no strictly dominated strategy is a best response in the sense of condition (1) of the definition of equilibrium with uncertainty aversion, R_i is a subset of I_i. *QED*

Interpreting Q_i as the beliefs which are not ruled out by common knowledge of procedural rationality (i.e. maximization given beliefs) when preferences are restricted to obey the axioms in Theorem 1, a refinement of equilibrium with uncertainty aversion is offered.

Definition: An equilibrium with uncertainty aversion is an *equilibrium with uncertainty aversion and rationalizable beliefs* if and only if B_i is a subset of Q_i for all players i.

Note that equilibrium with uncertainty aversion and rationalizable beliefs can be viewed as a refinement of correlated rationalizability (and thus, using a result of Brandenburger and Dekel (1987), of a posteriori equilibria) in that it takes the rationalizability restrictions and adds to them a consistency requirement (condition (2) in the definition of equilibrium with uncertainty aversion). Note that correlated rationalizability already requires a condition equivalent to (2) in the case $B_i = Q_i$. Imposing the consistency condition for beliefs which are subsets of Q_i allows for knowledge about the other players, besides knowledge of their rationality, to be reflected in beliefs, and thus in the equilibrium. Condition (2) is an appropriate consistency condition for equilibrium in the sense that it requires that players not rule out strategies incorrectly. The basic idea is that the Nash consistency condition makes sense if you are sure of the distribution over strategies (i.e. B_i is a singleton), but the idea of not being surprised (i.e. not ruling out the strategy profile that is played) is more general than this, in that knowledge that rules out some, but not all, other options can be incorporated. For example, suppose I am a baseball player and I know that the opposing pitcher does not know how to throw a split-fingered fastball. Any outcome in which the pitcher does, in fact, throw this pitch is surely not much of an equilibrium. On the other hand, I may be unable or unwilling to summarize my beliefs in the form of a single distribution over the remaining pitches. Thus, a slider or a curveball or any randomization between the two might not surprise me, and could be part of what might be reasonably called an equilibrium.

To see that the set of equilibria with uncertainty aversion and rationalizable beliefs can be strictly smaller than the set of rationalizable outcomes, consider the "battle-of-the-sexes" game depicted in figure 6.

	U	D
U	2,1	0,0
D	0,0	1,2

figure 6

In this game, (U, D) is rationalizable but is not an equilibrium with uncertainty aversion and rationalizable beliefs.[14] To see this, observe that player 1 plays U only if he has no subjective beliefs which assign weight less than 1/3 to 2 playing U. Similarly, 2 plays D only if she has no subjective beliefs which assign weight less than 1/3 to 1 playing D. These beliefs fail the consistency condition (2). Thus this condition shares some of the flavor of Rabin's (1989) point that we might not want to assign an outcome a higher probability then either of the players could given that they are playing best responses. Another example where the set of equilibria with uncertainty aversion and rationalizable beliefs is strictly larger than the set of Nash equilibria is given in figure 7.

	U	D
U	2,1	0,0
D	1,1	1,2

figure 7

This game is a modification of the "battle-of-the-sexes" game which makes D more attractive to 1 and U more attractive to 2 than before. (D, U) is an equilibrium with uncertainty aversion and rationalizable beliefs. Any sets of beliefs that include any measures which put weight greater than 1/2 on 2 playing D will lead player 1 to play

[14] In fact, (U, D) is not even an equilibrium with uncertainty aversion. This, together with the example in figure 5, makes it clear that there is no general containment relation between the set of rationalizable (or correlated rationalizable) outcomes and the set of outcomes of equilibria with uncertainty aversion.

D. Similarly, if player 2 has any measures which assign probability greater than 1/2 to 1 playing U then 2 will play U. However, (D, U) is not a Nash equilibrium.[15] In fact, if we replace the payoff of (1, 1) with a payoff of (k, k) where $0 < k < 2$, (D, U) fails to be Nash but is an equilibrium with uncertainty aversion and rationalizable beliefs for an ever wider class of beliefs as k approaches 2. Of course the mixed Nash equilibrium does approach (D, U) as k approaches 2, but it seems that allowing for a wider range of beliefs is much more helpful in assessing which outcomes would be expected in which environments. From the point of view of equilibrium with uncertainty aversion, the mixed strategy Nash outcome for $0 < k < 2$ will never be strictly preferred.

1.5 Weak Admissibility

Consider the game in figure 8.

	U	D
U	1,1	0,1
D	0,0	0,2

figure 8

In this game no strategies are eliminated by iterated strict dominance, thus the restriction to rationalizable beliefs makes no difference. There are lots of Nash equilibria (a continuum in fact). Thus there are also many equilibria with uncertainty aversion. Notice, however, that as long as player 1 thinks that U is possible, player 1 should play U in response. Similarly, as long as player 2 thinks that D is possible, player 2 should play D in response. This reasoning leads one to think that in any equilibrium with uncertainty aversion where 1 plays D (or 2 plays U) all beliefs in the belief set must assign probability 0 to 2 playing U (1 playing D). That this is not

[15] Note that (U, D), as in figure 6, is rationalizable but is not an equilibrium with uncertainty aversion and rationalizable beliefs. Thus the game in figure 7 demonstrates that the set of equilibria with uncertainty aversion and rationalizable beliefs can lie strictly between the set of rationalizable profiles and the set of Nash equilibria.

true is easily seen by considering the case where both players have belief sets which contain all possible distributions. In this case, each player is indifferent between any two strategies since all strategies give a minimum expected utility of 0. This is one aspect in which I feel that the Gilboa-Schmeidler axioms are too weak.

A similar point can be made by reconsidering the Ellsberg example. Consider again the thought experiment of section 1.1, specifically options 1 and 4. It would seem irrefutable that unless a decision maker is certain that yellow will not be drawn she should prefer 4 to 1. However if the set of measures C simply includes a measure which assigns zero weight to yellow, even if other measures in C do not, then a Gilboa-Schmeidler decision maker will be indifferent between 1 and 4 (assuming that all measures in C assign one-third to black). Under our interpretation of C, such a decision maker considers it possible that yellow may occur in the sense that she is willing to use a measure which implies that in evaluating some acts. The fact that such a decision maker is indifferent is therefore unreasonable. To remedy this I use an additional axiom that appears in Schmeidler (1989).

Definition: An event $E \in \Sigma$ is *null* if and only if $\forall f, g \in L$ such that $\forall s \in S/E, f(s) \sim g(s)$, it is true that $f \sim g$.

Definition: Denote the *set of non-null events* by NNE $= \{E \in \Sigma$ such that E not null$\}$.

B.1 (Weak Admissibility)

$\forall f, g \in L$, if for all $s \in S, f(s) \succeq g(s)$ then $f \succeq g$ and $[f \succ g$ if an only if for some $E \in$ NNE, $f(s) \succ g(s), \forall s \in E]$.

Intuitively, a null event is a set of states which is never decisive. To be null in the context of the Gilboa-Schmeidler theory, an event must never be assigned positive probability by any measure in C. This can be seen by considering two acts, one of which gives utility 100 if E occurs while the other gives utility 0 if E occurs and both of which give utility 200 if E does not occur. The distribution in C used to evaluate each of these acts will be the one(s) which puts the most weight on E. Thus E has probability zero according to all measures in C if and only if the decision maker is indifferent between the two acts. So we conclude that a null event must be assigned probability zero by all probability measures in C. Furthermore, any event which is assigned zero probability by all measures in C is a null event. Weak admissibility says that state-by-state weak dominance (and indifference) holds on the set of events which are given positive probability by some measure in C. We obtain the following representation theorem:

Theorem 5 *Let $\succeq$ be a binary relation on L_0. Then the following are equivalent,*

(1) $\succeq$ satisfies A.1 - A.3, A.5 and B.1 for $L = L_0$.

(2) There exists an affine function $u : Y \to \mathcal{R}$ and a non-empty, closed, convex set C of finitely additive probability measures on Σ satisfying $[p(E) = 0$ if and only if $\forall p \in C, p(E) = 0]$ such that $\forall f, g \in L_0, f \succeq g$ if and only if $\min_{p \in C} \int u \circ f dp \geq \min_{p \in C} \int u \circ g dp$.

Furthermore, the function u is unique up to a positive affine transformation and, if and only if A.6 holds, the set C is unique.

Proof: See Appendix.

The new representation is identical to that in Theorem 1 except for the additional condition that each event be given either zero probability by all measures in C or positive probability by all measures in C (i.e. the measures in C are mutually absolutely continuous). This condition serves to impose the weak admissibility axiom (B.1). However, this requirement seems too strong. It does not allow a decision maker to be uncertain about whether a given event will occur with positive probability. In a two player game, for instance, this representation would not allow a player to include both a pure strategy and any other strategy (mixed or pure) in her belief set. In order to permit this type of uncertainty while maintaining weak admissibility, A.3 (continuity) will be relaxed. The intuitive idea is that weak admissibility is a second-order criterion, in the sense that A.4 (monotonicity) ensures that weak admissibility is only used to break ties in the original representation, thus engendering a possibly discontinuous preference relation. Unfortunately, simply dropping continuity only when applying weak admissibility directly to break ties will not allow us to maintain A.1 (weak order), which is in many ways the most fundamental axiom. To get around this problem, we allow for a finite number of hierarchically ordered preference relations, while placing conditions on these relations.

Consider a finite set of preference relations on L: $\succeq_i, i = 1, \ldots, N$.

Definition: The $\succeq_i$ *agree on* L_c if and only if $\forall y, z \in L_c$, $[y \succeq_1 z$ if and only if $y \succeq_2 z \ldots$ if and only if $y \succeq_N z]$.

Definition: The $\succeq_i$ *display non-increasing valuation of certainty* if and only if $\forall f \in L, y \in L_c, [f \sim_i y$ implies $y \preceq_{i+1} f]$ holds for $i = 1, \ldots, N-1$.

Now consider the following axiom on the preference relation $\succeq$ on L:

B.2 (N-Hierarchy)

There exist $N \geq 1$ preference relations on L: $\succeq_1, \succeq_2, \ldots, \succeq_N$ such that, $\forall f, g \in L, f \succeq g \Leftrightarrow [g \succ_i f \Rightarrow \exists k < i$, such that $f \succ_k g]$. Furthermore, each $\succeq_i$ satisfies A.1-A.5, and the $\succeq_i$ agree on L_c and display non-increasing valuation of certainty.

Observe that any preference relation $\succeq$ which satisfies A.1-A.5 will satisfy B.2 for $N = 1$. Thus imposing A.1, A.2, A.4, A.5, B.1, and B.2 is certainly no stronger than imposing A.1-A.5, and B.1. B.2 limits the way in which continuity can be relaxed. It says that there are a finite number of preference relations which are combined lexicographically to represent $\succeq$.[16] Furthermore, each of these N relations must satisfy the original axioms A.1-A.5, must order constant acts the same way. and reward constant acts versus uncertain ones (weakly) less and less. Thus the decision maker has first-order G-S preferences, second-order G-S preferences, etc., and aversion to uncertainty is not as important in breaking ties as it is in the ordering where the ties occur. One can think of the decision maker "accounting for" uncertainty in the manner of Theorem 1 with her first-order preferences, and, given that prospects are equal by this measure, being willing to venture a tie-breaking decision on the basis of preferences which do not give as much weight to uncertainty, since this weight has, in some sense, already been given. This type of refinement could continue through several levels.

An alternate scenario would be to think that instead of reducing uncertainty

[16] The only restriction beyond weak order in this requirement is that N be finite. See Fishburn (1974) and Chipman (1971) for more details.

aversion at each stage, the decision maker actually became more uncertainty averse in the case of ties. An important drawback to this case, however, is that weak admissibility would end up imposing precisely the conditions which we wanted to avoid in Theorem 5. For this reason, I work with the former case.

We obtain the following representation theorem:

Theorem 6 *Let $\succeq$ be a binary relation on L_0. Then the following are equivalent,*

(1) $\succeq$ satisfies B.1 and B.2 for $L = L_0$.

(2) $\exists$ an affine function $u : Y \rightarrow \mathcal{R}$ and $N \geq 1$ non-empty, closed, convex sets $C_i, i = 1, \ldots, N$, of finitely additive probability measures on Σ such that $\forall f, g \in L_0, f \succeq g$ if and only if $(\min_{p \in C_i} \int u \circ f dp)_{i=1}^N \geq_L (\min_{p \in C_i} \int u \circ g dp)_{i=1}^N$, where if $p(E) > 0$ for some $E \in \Sigma$, $p \in C_1$ then there exists an i such that $p(E) > 0$, for all $p \in C_i$, and where $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_N$.[17]

Furthermore, the function u is unique up to a positive affine transformation, and, if and only if A.6 holds, the set C_1 is unique.[18]

Proof: See Appendix.

The following corollary makes it clear that A.3 (continuity) is the only one of the G-S axioms which is being relaxed:

Corollary 6.1

$\succeq$ satisfies B.1 and B.2 implies $\succeq$ satisfies A.1, A.2, A.4, and A.5.

Proof: It is straightforward to verify that the representation in Theorem 6 satisfies A.1, A.2, A.4, and A.5. *QED*

This representation satisfies weak admissibility, while also allowing the set of possible probability measures, C_1, to include both measures that assign zero probability to an event and ones that give the event positive weight. An interpretation of the

[17] For $a, b \in \mathcal{R}^N$, $a \geq_L b \Leftrightarrow [b_i > a_i \Rightarrow \exists k < i$ such that $a_k > b_k]$.

[18] In a context where the independence axiom is assumed to hold for all acts (and thus uncertainty aversion is ruled out), Blume, Brandenburger, and Dekel (1991) obtain a similar lexicographic representation where the belief sets are singletons, $N \leq \#S$, and the superset relations are not required to hold. The added structure provided by independence allows a more attractive axiomatization than the one here, obviating the need to refer to a hierarchy of preference relations in the axioms. Unfortunately the properties of an ordered vector space which they use do not seem applicable here.

subsets C_2 through C_N is that the measures in C_k are considered infinitesimally more likely (or more important in terms of the decision) than the measures in C_{k-1}/C_k in the sense that if two acts are equally ranked using C_{k-1} then the decision maker will use the ranking under C_k to attempt to further discriminate, but if two acts are strictly ranked under C_{k-1} then the ranking under C_k is irrelevant. Viewed in this way, weak admissibility requires only that any event which is given positive weight by some measure in C_1 be considered at least infinitesimally more likely to occur with positive probability than to occur with zero probability.

The definition of equilibrium with uncertainty aversion can be extended to the preferences described in Theorem 6.

Definition: An *equilibrium with uncertainty aversion* of G is a $(N+1) * I$-vector $(\sigma_1, \ldots, \sigma_I, B_{11}, \ldots, B_{1N}, B_{21}, \ldots, B_{2N}, \ldots, B_{I1}, \ldots, B_{IN})$ where $\sigma_i \in \Sigma_i$ (the set of mixed strategies for player i, i.e. the set of probability distributions over S_i) and the B_{in} are closed, convex subsets of P_{-i} (the set of probability distributions over $\times_{k \neq i} S_k$) satisfying $B_{i1} \supseteq B_{i2} \supseteq \ldots \supseteq B_{iN}$ and [$p(s_{-i}) > 0$ for some $p \in B_{i1} \Rightarrow p(s_{-i}) > 0$ for all $p \in B_{in}$ for some n] such that, for all i,

(1) σ_i satisfies $(\min_{p \in B_i} \sum_s u_i(s_i, s_{-i}) \sigma_i(s_i) p(s_{-i}))_{n=1}^N \geq_L$
$(\min_{p \in Bi} \sum_s u_i(s_i, s_{-i}) \sigma_i'(s_i) p(s_{-i}))_{n=1}^N$ for all $\sigma_i' \in \Sigma_i$, and

(2) $\prod_{k \neq i} \sigma_k(s_k) \in B_{i1}$.

Using this definition, analogues of all of the theorems in sections 1.3 and 1.4 can be derived, although the results are not as clean as with the simpler definition. To apply the new definition, we return to the game in figure 8.

Recall that without weak admissibility, there was a great multiplicity in the equilibria with uncertainty aversion (with or without rationalizable beliefs) even when no strategy of the opponent was ruled out by all measures in the belief set. However, with this restriction, unless all of player 1's beliefs (the set B_{i1}) assign probability zero to U, 1 should play U. Similarly, unless all of player 2's beliefs assign probability zero to D, 2 should play D. Thus, if players are uncertainty averse and any degree of uncertainty exists in each of their minds, weak admissibility argues that (U, D) will be the outcome.

Note that (U, D) is also the outcome picked out by deletion of weakly dominated strategies. In general, however, weak admissibility is a much weaker condition than weak dominance. Weak admissibility allows the play of weakly dominated strategies when a player always assigns probability zero to the state(s) where the dominance is strict. The power of weak dominance in the Nash framework is precisely (and only) that it rules out the play of weakly dominated strategies even when the relevant states are assigned probability zero. I believe that weak admissibility is a more accurate formalization of the ideas which are often used to motivate weak dominance. If the reason that weakly dominated strategies should not be played is that players will almost never be in a situation where they can be sure that their opponent(s) will not play a particular strategy or strategies then that idea should be expressed directly, in terms of beliefs, rather than in a rule which is to be universally applied regardless of the beliefs in any particular situation. Weak admissibility makes clear this dependence on beliefs. For example, if only rationalizable beliefs are allowed, then strategies which would have been eliminated by weak dominance are allowed if they were strictly dominated only by those actions which rationalizable beliefs must assign probability zero. For example, consider the game in figure 9.

	X	Y	Z
A	1,1	0,1	1,2
B	0,0	0,2	1,1

figure 9

In this game, weak dominance eliminates B for player 1, whereas, since X is eliminated by iterated strict dominance, B is not eliminated by weak admissibility under the restriction to rationalizable beliefs.

1.6 Conclusion

The goal of this paper has been to explore some of the consequences for game theory of an attractive broadening in the decision theory used to describe the players. The concept of equilibrium with uncertainty aversion with or without a restriction to rationalizable beliefs turns out to have some nice features in normal form games. First, it provides a new justification based on hedging for some equilibria involving mixed strategies. Second, the flexibility of belief structure points out certain equilibria which I argue would not make very good predictions unless very definite information about beliefs were available. Third, this framework allows for oft-mentioned unmodelled features of the game environment, such as social norms, past experience of the players, and knowledge of equilibrium concepts to be incorporated in a natural way through their effects on the uncertainty which uncertainty averse players experience. When rationalizable beliefs are imposed, this solution concept can be viewed as a refinement of correlated rationalizability which is not as restrictive as Nash equilibrium. Finally, the flexibility of beliefs helps make weak admissibility a relevant condition.

There is obviously much that needs to be done if these ideas are to form the basis of a complete theory. The biggest missing piece is an extension of these concepts to extensive form, and thus dynamic, games. One route to follow here is to develop a satisfactory notion of updating the sets of probability measures. Gilboa and Schmeidler (1993) have done some preliminary work on this front. One procedure which they suggest which seems potentially appealing is, after an event occurs, to rule out some of the measures and update the rest by applying Bayes' rule to each one. However, it is known (see Epstein and Le Breton (1993) and Klibanoff (1993a)) that no update rule for sets of measures in the G-S framework guarantees dynamically consistent preferences. Klibanoff (1993b) responds to this by axiomatizing an alternative, explicitly dynamically consistent, representation of uncertainty aversion. Using such a theory to analyze dynamic games is a topic of future research.

Another thing missing from the present work is a discussion of games of incomplete information. However, it should not be difficult to apply a slightly adapted version of the present theory to such games. The basic change would involve an enlargement of the state space of player i, S_{-i}, to $S_{-i} \times \Theta$ where Θ is the space of unknown parameters. On a more applied level, it would be good to develop a full-blown application using these equilibrium concepts.

Bibliography

[1] Allais, M. [1953], "Le comportement de l'homme rationnel devant de risque: Critique des postulats et axiomes de l'ecole Americaine", *Econometrica* 21, pp. 503-546.

[2] Anscombe, F. J. and R. J. Aumann [1963], "A definition of subjective probability", *The Annals of Mathematical Statistics* 34, pp. 199-205.

[3] Aumann, R. [1990], "Communication need not lead to Nash equilibrium", Mimeo, Hebrew University of Jerusalem.

[4] Bernheim, D. [1984], "Rationalizable strategic behavior", *Econometrica* 52, pp. 1007-1028.

[5] Blume, L., A. Brandenburger and E. Dekel [1991], "Lexicographic probabilities and choice under uncertainty", *Econometrica* 59, pp. 61-79.

[6] Brandenburger, A. and E. Dekel [1987], "Rationalizability and correlated equilibria", *Econometrica* 55, pp. 1391-1402.

[7] Camerer, C. and M. Weber [1992], "Recent developments in modeling preferences: uncertainty and ambiguity", *Journal of Risk and Uncertainty* 5, pp. 325-370.

[8] Chateauneuf, A. [1991], "On the use of capacities in modeling uncertainty aversion and risk aversion", *Journal of Mathematical Economics* 20, pp. 343-369.

[9] Chipman, J. [1971], "On the lexicographic representation of preference orderings", in *Preferences, Utility, and Demand*, J.S. Chipman, L. Hurwicz, M.K. Richter and H. F. Sonnenschein (eds.), Harcourt Brace Jovanovich.

[10] Dow, J. and S. Werlang [1991], "Nash Equilibrium under Knightian uncertainty: Breaking down backward inductio", Mimeo.

[11] Dunford, N. and J.T. Schwartz [1957], *Linear Operators, Part I*, Interscience.

[12] Ellsberg, D. [1961], "Risk, ambiguity, and the Savage axioms", *Quarterly Journal of Economics* 75, pp. 643-669.

[13] Epstein, L. G. and M. Le Breton [1993]," Dynamically Consistent Beliefs must be Bayesian", *Journal of Economic Theory* 61, pp. 1-22.

[14] Fishburn, P. [1974], "Lexicographic orders, utilities and decision rules: a survey", *Management Science* 20, pp. 1442-1471.

[15] Fishburn, P. [1988], *Nonlinear Preference and Utility Theory*, Johns Hopkins Univ. Press.

[16] Fudenberg, D. and D. Levine [1990], "Steady-state learning and self-confirming equilibrium", Mimeo, MIT.

[17] Fudenberg, D. and D. M. Kreps [1991], "A Theory of Learning, Experimentation, and Equilibrium in Games", Mimeo, MIT.

[18] Fudenberg, D. and J. Tirole [1991], *Game Theory*, MIT Press.

[19] Gilboa, I. [1987], "Expected utility theory with purely subjective non-additive probabilities", *Journal of Mathematical Economics* 16, pp. 65-88.

[20] Gilboa, I. and D. Schmeidler [1989], "Maxmin expected utility with non-unique prior", *Journal of Mathematical Economics* 18, pp. 141-153.

[21] Gilboa, I. and D. Schmeidler [1993], "Updating Ambiguous Beliefs", *Journal of Economic Theory* 59, pp. 33-49.

[22] Harsanyi, J. [1973], "Games with randomly disturbed payoffs: A new rationale for mixed strategy equilibrium points", *International Journal of Game Theory* 1, pp. 1-23.

[23] Harsanyi, J. and R. Selten [1988], *A General Theory of Equilibrium Selection in Games*, MIT Press.

[24] Klibanoff, P. [1993a], "On Updating Sets of Measures", Mimeo, MIT.

[25] Klibanoff, P. [1993b], "Dynamic Choice with Uncertainty Aversion", Mimeo, MIT.

[26] Knight, F. [1921], *Risk, Uncertainty, and Profit*, Houghton-Mifflin.

[27] Machina, M. [1989], "Dynamic consistency and non-expected utility models of choice under uncertainty", *Journal of Economic Literature* 27, pp. 1622-1668.

[28] Nash, J. [1950], "Equilibrium points in n-person games", *Proceedings of the National Academy of Sciences* 36, pp. 48-49.

[29] von Neumann, J. and O. Morgenstern [1947], *Theory of Games and Economic Behavior*, second edition, Princeton Univ. Press.

[30] Pearce, D. [1984], "Rationalizable strategic behavior and the problem of perfection", *Econometrica* 52, pp. 1029-1050.

[31] Quiggin, J. [1982], "A theory of anticipated utility", *Journal of Economic Behavior and Organization* 3, pp. 323-343.

[32] Rabin, M. [1989], *Predictions and solution concepts in noncooperative games.*, Ph.D. dissertation, Department of Economics, MIT.

[33] Raiffa, H. [1961], "Risk, ambiguity, and the Savage axioms: comment", *Quarterly Journal of Economics* 75, pp. 690-694.

[34] Raiffa, H. [1968], *Decision Analysis Introductory Lectures on Choices under Uncertainty*, Addison-Wesley.

[35] Savage, L. J. [1954], *The Foundations of Statistics*, Wiley. Revised and enlarged edition, Dover, 1972.

[36] Schmeidler, D. [1986], “Integral representation without additivity”, *Proceedings of the American Mathematical Society* 97, no. 2, pp. 255-261.

[37] Schmeidler, D. [1989], “Subjective probability and expected utility without additivity”, *Econometrica* 57, pp. 571-587.

[38] Shackle, G. L. S. [1949], *Expectation in Economics*, Cambridge Univ. Press.

[39] Shackle, G. L. S. [1949-50], “A non-additive measure of uncertainty”, *Review of Economic Studies* 17, pp. 70-74.

[40] Slovic, P. and A. Tversky [1974], “Who Accepts Savage’s Axiom?”, *Behavioral Science* 19, pp. 368-73.

[41] Wakker, P. [1990], “Under stochastic dominance Choquet-expected utility and anticipated utility are identica”,*Theory and Decision* 29, pp. 119-132.

[42] Yaari, M. E. [1987], “Dual theory of choice under uncertainty”, *Econometrica* 55, pp. 95-115.

Chapter 2

Dynamic Choice with Uncertainty Aversion

2.1 Introduction

In this paper I develop a dynamically consistent theory of decision making that incorporates the notion of *uncertainty aversion.* Uncertainty aversion occurs when individuals have a dislike of uncertainty (not knowing the relevant probabilities in a given decision environment) and is distinct from their attitude towards risk (not knowing which outcome will occur, but knowing the probability of each outcome). While attitudes towards risk can be captured by varying the shape of the utility function in expected-utility theory (with either objective probabilities as in the theory of von Neumann and Morgenstern (1947), with subjective probabilities as in the theory of Savage (1954) or with both types of probabilities as in Anscombe and Aumann (1963)), there is no way to incorporate attitudes towards uncertainty in this expected-utility framework.

A classic example of behavior which is not compatible with standard theory and is naturally explained by uncertainty aversion is one by Ellsberg (1961). Here an individual faces an urn which contains 90 balls identical except for color. Thirty of the balls are black. The other possible colors are red and yellow, and there is no information about the proportions of these two colors. One ball will be drawn at random from the urn. In these circumstances, many decision-makers prefer to bet on black (i.e. they win 100 dollars if the ball drawn is black, 0 dollars otherwise) than to bet on red. At the same time, they prefer to bet on [red or yellow] (i.e. win 100 dollars if the ball is red or yellow) than to bet on [black or yellow]. This is incompatible with rational choice using any fixed probabilities of the different colors to weight outcomes. A natural explanation of this behavior is that the individual assigns a premium to bets for which the odds are known (betting on black or betting on [red or yellow]) over those for which the odds are not known (betting on red or betting on [black or yellow]). A large body of experimental work has found support for this type of behavior (see, e.g., the discussion in Camerer and Weber (1992)).

A static representation theory that allows for uncertainty averse behavior has been developed by Gilboa and Schmeidler (1989). In their theory, beliefs are represented by sets of probability measures and individuals choose actions which maximize the minimum expected utility where the minimum is taken over the measures in the belief set. While some exploration of the implications of such behavior in economic settings has been undertaken (e.g. Dow and Werlang (1991, 1992) on portfolio choice and game theory and Klibanoff (1992) on game theory), progress on this front has been hampered by the lack of a satisfactory dynamic theory to complement the static theory. For most interesting economic problems (for example, any problem modelled naturally as an extensive form game) we need a dynamic, and not simply a static theory.

The usual way of extending a static choice theory under uncertainty to a dynamic setting is to make assumptions about the way in which beliefs (as separated from utilities in the representation) are updated as new information becomes available. In the standard subjective expected utility theory (SEU) of Savage or Anscombe and Aumann, the usual assumption is that beliefs are updated by Bayes' rule whenever possible. This has great appeal for at least two reasons, one economic and one aesthetic. The economic appeal is that Bayes' rule is the only updating rule for a SEU maximizer that will never lead to dynamic inconsistency (see e.g. Brown (1976)). The aesthetic appeal arises from the agreement with Bayes' rule for conditional probabilities which follows from the definition of conditional probability and the axioms of probability theory.

Unfortunately, there exists no rule for updating beliefs which guarantees dynamic consistency in general when the decision maker's static preferences display Gilboa and Schmeidler's (1989) formulation of uncertainty aversion. A simple proof of this fact by way of two examples is contained in Klibanoff (1993).

In fact, a much stronger result is true as has been shown in Epstein and Le Breton (1993), namely that dynamic consistency and the relatively uncontroversial axioms of Savage (1954) (in particular, excluding the Sure-Thing Principle (Axiom P2)) imply that beliefs must be represented by a qualitative probability relation.[1] In other words, there must be a weak order on events expressing the relation "at least as likely as" where likelihood is operationalized by willingness to bet in the Savage sense.

In the uncertainty aversion framework of Gilboa and Schmeidler, this result implies that, for any events A and B, if the belief set contains one measure which assigns $p(A) > p(B)$ then all measures in the set must assign $p(A) > p(B)$ if dynamic consistency is to be maintained through updating beliefs.[2] In particular, requiring dynamic consistency in this setting rules out the behavior displayed in the Ellsberg Paradox, which is a major motivation and justification for the whole body of literature on uncertainty aversion.

Why not then just accept dynamic inconsistency as a price to pay when modelling uncertainty averse behavior in a dynamic setting? Why not adopt, say, an assumption of sophisticated behavior in the presence of dynamic inconsistency, as in Pollack (1968), Laibson (1992), Karni and Safra (1990), and many others? A major reason is that except in relatively few cases, it is very hard to analyze dynamic problems using these types of preferences. For example, the standard tools of dynamic and stochastic dynamic programming are not applicable because of the lack of consistency. This reason may not be very deep, but it is nonetheless quite important if these preferences are to be widely used in modelling. Another reason is that dynamic

[1] Their result relies mainly on a theorem of Machina and Schmeidler (1992).

[2] This is only a necessary condition for dynamic consistency. It need not be sufficient.

consistency has appeal as a normative principle. Finally, it is important to know if dynamic inconsistency is a necessary price to pay for modelling uncertainty aversion, or whether there are ways to get around this.

The approach I will take is based on the idea of Kreps and Porteus (1978) of modelling the consequences at each time (or stage) t as composed of an immediate payoff and an opportunity set from which the action at time $t+1$ will be chosen. However, this approach substantially departs from Kreps and Porteus along several dimensions.

First, my setting is a "states of the world" Anscombe-Aumann subjective probability framework as opposed to a world of purely known or objective probabilities. To my knowledge the only other work which axiomatizes dynamically consistent intertemporal utility in a subjective probability framework is Skiadas (1991), which axiomatizes recursive utility in a generalized Savage style framework.

Second, at each time t, preferences are not required to conform to the standard axioms which give rise to an expected utility representation. Chew and Epstein (1989) is an example of a paper in this vein, as they consider preferences which may violate the independence axiom but satisfy betweeness in an objective probability framework. (Betweeness requires a mixture $\alpha a+(1-\alpha)b$ to be "between" a and b in the preference ordering. (i.e. $a \succeq b \Rightarrow a \succeq \alpha a+(1-\alpha)b \succeq b$) Since I allow for uncertainty aversion, the preferences here do not even satisfy the analogue of betweeness for uncertain acts (although they do satisfy it when no uncertainty is present).)

Finally, since I want to represent uncertainty aversion and require dynamic consistency, utilities over payoff/opportunity set pairs must be allowed to be state dependent. This is needed for the same reason that value functions in dynamic programming are state dependent: the value of an opportunity set will vary depending on how uncertainty resolves tomorrow. I will discuss this analogy further in an example at the end of section 2.2. This work is thus also related to the literature on state dependent preferences (e.g. Karni (1985, 1993)). For the reasons discussed by these authors and others, allowing state dependence may be important in capturing many decision problems in a coherent way. Thus, a derivation of static, state dependent, uncertainty averse preferences, which this paper provides as a step in building a dynamic theory, may be important in its own right. However, viewed purely as a theory of state dependent preferences with uncertainty aversion, the results presented here are not as general as might be desired. The reason is that I require that the preference overlap between states be complete. In other words, for any outcome in state a, there is an outcome in state b which is just as good, and vice-versa. As discussed later on, this assumption is probably a reasonable one in the context of the dynamic theory that I develop and for the applications that I have in mind. However, it seriously limits the allowed nature of the state dependence, and thus may not be useful in some settings that the above-mentioned theories are, such as the analysis of life and health insurance.

As well as developing a dynamic choice theory, this paper also contains a brief application of the theory to the question of the existence of a reservation price rule when searching without recall from an unknown distribution of prices. I extend the results of Rothschild (1974) and Bikhchandani and Sharma (1989) to the case of an uncertainty averse searcher. Although the result is of independent interest, the main point is to demonstrate the fact that the theory developed here is a tractable tool for economic analysis. Dynamic consistency is a crucial feature in making this true. A full-blown application of preferences that look like those I characterize is presented in Epstein and Wang (1992) who examine an intertemporal model of asset demand with uncertainty averse individuals. This paper provides an axiomatic foundation for the preferences they assume.

The rest of the paper is laid out as follows. In section 2.2, I set out the notation, formally describe the decision environment, and present an example. In section 2.3, axioms for time t preferences are introduced and a static, state dependent, uncertainty averse representation theorem is presented. In section 2.4, the notion of dynamic consistency is formalized, and is used to knit the static preferences together. In section 2.5, a result on searching from an uncertain distribution is presented. Section 2.6 concludes.

2.2 The Model

Let J be a set of prizes or consequences. These are the ultimate outcomes — for example amounts of money or consumption goods. Consider a discrete-time, finite horizon setting with horizon $T > 0$. I will now recursively define several constructs that are important in describing the decision problem. G_t is the set of lotteries over "outcomes" in period t. In this setting, outcomes are pairs in which the first element is an immediate payoff and the second element is an opportunity set from which the next period's action will be chosen. S_t is the set of states of the world at time t. Note that in this paper I will assume that S_t is finite for all t. N_t is a set of time t acts. Acts are the choice variables in the model and each act maps states to lotteries over outcomes. X_t is the set of all non-empty, closed, convex sets of time t acts which can be generated by taking the convex hull of a finite number of time t acts; thus X_t is the set of opportunity sets of time t acts. I formally define these constructs iteratively, starting from the end. Let G_T be the set of all distributions with finite support over J. Let S_{T+1} be a singleton set of states of the world. Let N_T be the set of all bounded functions from S_{T+1} to G_T. For all $0 \leq t \leq T$, let X_t be the set of all non-empty, closed, convex subsets of N_t which can be generated as the convex hull of a finite number of elements of N_t. For all $1 \leq t \leq T$, let G_{t-1} be the set of all distributions with finite support over $J \times X_t$. For all $1 \leq t \leq T$, let N_{t-1} be the set of all bounded functions from S_t to G_{t-1} such that each element of the set induces a marginal over J which is constant in S_t. This last requirement guarantees that tomorrow's state realization does not affect the marginal distribution over prizes received today. This describes the fact that these prizes are "immediate" and thus unaffected by future uncertainty.

The decision problem is a sequential one starting at $t = 0$. At each time $0 \leq t \leq T$, preferences over N_t are assumed to exist and these preferences govern the individual's choice among the acts in N_t that are available. By choosing over acts in N_t the individual makes only that part of the overall dynamic decision which must be made at time t, namely picking a lottery (possibly degenerate) over immediate payoffs and a function which will determine the choice set the individual will face at $t+1$ as a function of the uncertainty that resolves between t and $t+1$ (captured by S_{t+1}).

To give these constructions a context and help the reader fix ideas, I now set up an example of a dynamic choice problem. The example is a problem of sequential price search without recall from an unknown distribution of prices. I will return to this example in section 2.5.

Consider the following problem. There is an item which you value at w dollars. The price of this item is set every day by an independent random draw from a fixed distribution of prices. You desire only one unit of the item and get no additional utility from having more than one unit. The process of going on a given day to check the price gives you disutility $c > 0$. Each day, you must decide whether to buy at that day's price and stop shopping or wait to see what tomorrow's price will be. Here, the S_t correspond to the possible prices at time t. The prizes include the item, the search cost, and monetary values associated with paying for the item. An example of a time t act is "stop and buy at the current price p_t". This act gives an immediate payoff $w - p_t$ (assuming risk neutrality). It gives an opportunity set which contains only what I will call the null act, since the decision problem ends once you have bought. Another act is "search once more". This act gives an immediate payoff of $-c$ since there is a cost to search, and gives an opportunity set which varies across S_{t+1}. If $s_{t+1} = p_{t+1}$ then the opportunity set consists of the time $t+1$ acts "stop and buy at price p_{t+1}" and "search once more" and any convex combination of the two. Note that a half-half mixture of these two acts yields an act that gives a half-half lottery over the outcomes of the two acts in each state.

To see why I must allow utility functions to be state dependent, think about how dynamic programming is used to solve a sequential problem. First, through backwards induction type arguments I derive a value function which gives the value of being at time t' with price $p_{t'}$ and beliefs $B_{t'}$ given optimal behavior from then on. Now consider the utility of continuing to search at time $t' - 1$. Assuming utility is time separable and discounted, the value is $-c$ (the immediate payoff) + $\delta V_{t'}(p_{t'}, B_{t'})$ if the state is $s_{t'} = p_{t'}$. Thus the utility of the opportunity set consisting of "search once more" clearly will depend on the state at time t'.

Observe that we don't usually think of dynamic programming situations as involving "state dependent utility" because utility is usually defined over payoffs only. The reason I use the Kreps-Porteus (1978) approach of considering choice over pairs of immediate payoffs and opportunity sets is that, as discussed in the introduction, assuming uncertainty aversion directly over payoffs won't yield dynamic consistency. Just as Kreps and Porteus relate their preferences to non-indifference towards the timing of the resolution of risk, the preferences derived here differ from assuming uncertainty aversion directly over payoff streams in that the timing (real or perceived) of the resolution of uncertainty matters. At each point in time (or at each decision node) the decision-maker cares directly about only that uncertainty which will resolve before the next decision. The remaining uncertainty (and aversion to it) is incorporated only indirectly, through its effect on future utility.

2.3 Axioms and a Static Theory

I will proceed by proposing axioms for the decision-maker's preferences to obey at each time t and for each history of prize and state realizations up to t, y_t. These axioms imply all of the axioms in Gilboa and Schmeidler (1989) except for state independence (which is implicit in their monotonicity axiom).

A1 (Weak Order)

For each t and $y_t \in Y_t$, there exists a complete and transitive binary relation $\succeq_{y_t}$ that represents the decision-maker's choices among elements of N_t.

A2 (Continuity)

For each t and $y_t \in Y_t$, $\forall n_t, n_t', n_t'' \in N_t$ if $n_t \succ_{y_t} n_t'$ and $n_t' \succ_{y_t} n_t''$ then there exists $\alpha, \beta \in (0,1)$ s.t. $\alpha n_t + (1-\alpha)n_t'' \succ_{y_t} n_t'$ and $n_t' \succ_{y_t} \beta n_t + (1-\beta)n_t''$.

Axioms A1 and A2 are completely standard and serve to guarantee the existence of a real-valued representation.

A3 (Uncertainty Aversion)

For each t and $y_t \in Y_t$, $\forall n_t, n'_t \in N_t$ and all $\alpha \in (0,1)$, $n_t \sim_{y_t} n'_t$ implies $\alpha n_t + (1-\alpha)n'_t \succeq_{y_t} n_t$.

This is the uncertainty aversion axiom of Gilboa and Schmeidler (1989). The interpretation is that mixing two acts can help hedge against uncertainty. As an example, imagine that are two states, s_{t+1} and s'_{t+1}. Imagine further that one act gives a zero immediate payoff and gives an opportunity set which contains only an act giving a prize with utility one if the state is s_{t+1}, and an opportunity set that contains only an act giving a prize with utility zero if the state is s'_{t+1}. There is also another act which again gives immediate payoff zero and the same two opportunity sets, however this act gives the utility one opportunity set in state s'_{t+1} and the utility zero opportunity set in state s_{t+1}. If there is uncertainty about the probabilities of s_{t+1} and s'_{t+1}, then these acts each give uncertain payoffs. However, by mixing these acts with probabilities half-half, the expected utility in both states becomes one-half. Thus two uncertain acts have been hedged against each other to make a certain act. Axiom A3 simply says that hedging between indifferent acts is not disliked. Note that in standard subjective expected utility (SEU) theory, this axiom holds with indifference replacing weak preference. For a discussion of hedging in the context of normal form games see Klibanoff (1992).

I now introduce the concept of a state lottery. State lotteries can be thought of as acts in a world where the probabilities of the states are known or objective. Thus state lotteries may involve risk but cannot involve uncertainty. The idea of using preferences over state lotteries in a state dependent framework to separate utilities from beliefs is due to Karni, Schmeidler, and Vind (1983).

Definition: A state lottery, $\hat{n}_t$, is a measurable map from $S_{t+1} \times G_t \rightarrow [0,1]$ s.t. $\sum_{S_{t+1}} \int_{G_t} \hat{n}_t(s_{t+1}, g_t) = 1$, and such that $\hat{n}_t(s_{t+1}, g_t) > 0$ implies $\hat{n}_t(s_{t+1}, g'_t) = 0$ for all $g'_t \neq g_t$. The set of all state lotteries is $\hat{N}_t$.

Definition: $\hat{n}_t \in \hat{N}_t$ is full-support if, for all states $s_{t+1} \in S_{t+1}$, there exists a $g_t \in G_t$ for which $\hat{n}_t(s_{t+1}, g_t) > 0$.

The next three axioms concern preferences over state lotteries, represented by the binary relation $\hat{\succsim}_{y_t}$ on $\hat{N}_t$. A4-A6 are the standard von Neumann-Morgenstern axioms, thus they impose standard (although state dependent) expected utility behavior in choosing over state lotteries. This emphasizes that it is only the effect of uncertainty which will distinguish the theory developed here from the more standard theory developed by Kreps and Porteus (1978) under risk.

A4 (Weak Order on state lotteries)

A1 applied to $\hat{\succsim}_{y_t}$ over $\hat{N}_t$.

A5 (Continuity on state lotteries)

A2 applied to $\hat{\succsim}_{y_t}$ over $\hat{N}_t$.

A6 (Independence on state lotteries)

For each t and $y_t \in Y_t$, $\forall \hat{n}_t, \hat{n}'_t, \hat{n}''_t \in \hat{N}_t$ and all $\alpha \in (0,1)$, $\hat{n}_t \hat{\succ}_{y_t} \hat{n}'_t$ if and only if $\alpha\hat{n}_t + (1-\alpha)\hat{n}''_t \hat{\succ}_{y_t} \alpha\hat{n}'_t + (1-\alpha)\hat{n}''_t$.

A7 (Non-triviality of prize preference in each state)

For all t, $y_t \in Y_t$, for any full-support $\hat{n}_t \in \hat{N}_t$, there exists, for each state $s^*_{t+1} \in S_{t+1}$, a full-support $\hat{n}'_t(s^*_{t+1}) \in \hat{N}_t$ such that $\hat{n}'_t(s^*_{t+1})$ equals $\hat{n}_t$ outside of s^*_{t+1} and not $\hat{n}_t \hat{\sim}_{y_t} \hat{n}'_t(s^*_{t+1})$.

Axiom A7 requires that in each state there be strict preference between at least one pair of outcomes. This will help insure the uniqueness of beliefs in the representation since it will allow me to distinguish between a state assigned zero probability and a state in which all consequences are indifferent by ruling out the latter.

Definition: Define $H : \hat{N}_t \rightarrow N_t$ such that, for all full-support $\hat{n}_t$, $H(\hat{n}_t)$ gives g_t conditional on state s_{t+1} if and only if $\hat{n}_t(s_{t+1}, g_t) > 0$.

Notice that H is the natural transformation from state lotteries to acts. For a state lottery which places positive probability on each state there is a well-defined lottery over outcomes conditional on the occurrence of each state. H simply returns the act that yields these lotteries in the appropriate states. It is as if the probability distribution over states were "stripped off", turning a situation of risk into one of uncertainty. The next two axioms, which give conditions on the relationship between preferences over acts, $\succeq_{y_t}$, and preferences over state lotteries, $\hat{\succeq}_{y_t}$, will use H in establishing this correspondence.

A8 (Monotonicity)

(a)For any full-support $\hat{n}_t, \hat{n}'_t \in \hat{N}_t$ such that $\hat{n}_t$ and $\hat{n}'_t$ are equal outside of some state s^*_{t+1}, $H(\hat{n}_t) \succ_{y_t} H(\hat{n}'_t)$ implies $\hat{n}_t \hat{\succ}_{y_t} \hat{n}'_t$, $\hat{n}_t \hat{\succ}_{y_t} \hat{n}'_t$ implies $H(\hat{n}_t) \succeq_{y_t} H(\hat{n}'_t)$, and $\hat{n}_t \hat{\sim}_{y_t} \hat{n}'_t$ implies $H(\hat{n}_t) \sim_{y_t} H(\hat{n}'_t)$.

(b)For any full-support $\hat{n}_t, \hat{n}'_t \in \hat{N}_t$ such that $\hat{n}_t$ and $\hat{n}'_t$ have the same distribution over S_{t+1}, define $\hat{n}''_t(s^*_{t+1})$ to be the state lottery which equals $\hat{n}_t$ on s^*_{t+1} and $\hat{n}'_t$ elsewhere. If $\hat{n}''_t(s^*_{t+1}) \hat{\succ}_{y_t} \hat{n}'_t$ for all $s^*_{t+1} \in S_{t+1}$, then $H(\hat{n}_t) \succ_{y_t} H(\hat{n}'_t)$.

Condition (a) of axiom A8 says (in light of the assumptions in axioms A4-A6) that preferences over acts satisfy a weak monotonicity property, namely that if the outcomes of an act are, in each state, at least as good as those of another act, then the first act must be at least as preferred as the second act. Condition (b) says that strict preference over outcomes in every state implies strict preference between the acts.

A9 (Cardinal Invariance)

For any $n_t, n_t{}', n_t{}'', n_t{}''' \in N_t$, consider full-support state lotteries $\hat{n}_t, \hat{n}'_t, \hat{n}''_t, \hat{n}'''_t \in \hat{N}_t$ defined so that $H(\hat{n}_t) = n_t, H(\hat{n}'_t) = n_t{}', H(\hat{n}''_t) = n_t{}''$, and $H(\hat{n}'''_t) = n_t{}'''$ and so that $\hat{n}_t$ and $\hat{n}''_t$ give the same distribution over S_{t+1} and $\hat{n}'_t$ and $\hat{n}'''_t$ give the same distribution over S_{t+1}. If all such $\hat{n}_t, \hat{n}'_t, \hat{n}''_t, \hat{n}'''_t$ satisfy $\hat{n}_t \hat{\succeq}_{y_t} \hat{n}'_t$ if and only if $\hat{n}''_t \hat{\succeq}_{y_t} \hat{n}'''_t$, then $n_t \succeq_{y_t} n_t{}'$ if and only if $n_t{}'' \succeq_{y_t} n_t{}'''$.

This axiom says that if the preference between the members of one pair of state lotteries is always the same as the preference between the members of another pair of state lotteries whenever the same pair of distributions over states are generated by each pair, then the preference between the associated acts in each pair must also be the same. In short, the effect of uncertainty on preference does not depend on the particular cardinal transformation of the utility function used to value the outcomes of acts, thus the name "Cardinal Invariance". Specifically, this axiom, together with axioms A4-A6, guarantee that if one pair of acts have a utility representation that is a positive affine transformation of the representation of another pair, then preference between the members of the first pair should correspond to preference between the members of the second pair. If this axiom were violated and A4-A6 held, then some other aspect of an act would have to matter apart from the preferences over its consequences.

A10 (Compensability across states)

Consider any degenerate state lottery $\hat{n}_t \in \hat{N}_t$ which has a distribution over S_{t+1} which puts all the mass on a single state, s^*_{t+1}. For any such $\hat{n}_t$ and for each $s_{t+1} \in S_{t+1}$, there exists a degenerate state lottery $\hat{n}'_t \in \hat{N}_t$ which has a distribution over S_{t+1} which puts all the mass on s_{t+1}, such that $\hat{n}_t \hat{\sim}_{y_t} \hat{n}'_t$.

Axiom A10 says that for any pair of states, and any outcome in the first state, one can find an outcome in the second state which is just as good. There are circumstances where this assumption is unappealing – for example, if one state were death and another were perfect health. However, in the context of a dynamic decision problem or an extensive form game, where states are typically moves by nature or the opponents, it is reasonable to assume that one could always be compensated by a large enough change in immediate payoff to make up for any differential value of an opportunity set due to the action of the opponents.

The following definition is used in the proof.

Definition: A constant-utility act $n_t \in N_t$ is an act s.t. $\forall$ full-support $\hat{n}_t, \hat{n}'_t \in \hat{N}_t$ for which $H(\hat{n}_t) = n_t$ and $H(\hat{n}'_t) = n_t$, it is true that $\hat{n}_t \hat{\sim}_{y_t} \hat{n}'_t$.

With these axioms we can prove a representation theorem for preferences over acts and state lotteries at each time t and history y_t. My proof draws heavily on techniques used in Gilboa and Schmeidler (1989) and Chateauneuf (1991).

Theorem 1 *For every t, y_t, the following are equivalent:*

(1) $\succeq_{y_t}$ and $\hat{\succeq}_{y_t}$ satisfy A1-A10

(2) There exists a function $w_{y_t} : J \times X_{t+1} \times S_{t+1} \rightarrow \mathcal{R}$ and a non-empty, closed and convex set C_{y_t} of additive probability measures on S_{t+1} such that:

(a) $n_t \succeq_{y_t} n_t'$ if and only if

$\min_{p \in C_{y_t}} \sum_{S_{t+1}} p(s_{t+1}) \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1})(n_t(s_{t+1})(j, x_{t+1})) \geq$
$\min_{p \in C_{y_t}} \sum_{S_{t+1}} p(s_{t+1}) \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1})(n_t'(s_{t+1})(j, x_{t+1}))$

(b) $\hat{n}_t \hat{\succeq}_{y_t} \hat{n}_t'$ if and only if $\sum_{S_{t+1}} \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1}) \hat{n}_t((j, x_{t+1}), s_{t+1}) \geq$
$\sum_{S_{t+1}} \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1}) \hat{n}_t((j, x_{t+1}), s_{t+1})$.

(c) The w_{y_t} above is unique up to a positive affine transformation.

(d) The set C_{y_t} above is unique.

(e) The $w_{y_t}(\cdot, s_{t+1})$ have the same range for all $s_{t+1} \in S_{t+1}$.

Proof:

I start by proving that (1) implies (2). Using axioms A4-A6 apply the von Neumann-Morgenstern theorem to give a utility function $w_{y_t} : J \times X_{t+1} \times S_{t+1} \rightarrow \mathcal{R}$, unique up to a positive affine transformation, such that for $\hat{n}_t, \hat{n}_t' \in \hat{N}_t$, 2(b) holds. Note that summations can be used since S_{t+1} is assumed finite and elements of G_t have finite support, so that elements of $\hat{N}_t$ give lotteries with finite support.

For each act $n_t \in N_t$, if $\sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1})(n_t(s_{t+1})(j, x_{t+1}))$ is constant over S_{t+1}, define $K(n_t) =$ that constant value $= \sum w_{y_t} n_t$. By compensability across states, monotonicity (A8) and the other axioms, for any $n_t \in N_t$ there exist constant utility acts $\overline{n}_t$ and $\underline{n}_t \in N_t$ such that $\overline{n}_t \succeq_{y_t} n_t \succeq_{y_t} \underline{n}_t$. We can find such acts by first writing n_t as an $|S_{t+1}|$-vector of utilities, finding the highest utility level, using compensability across states (A10) to find equivalent outcomes in the other states, and defining $\overline{n}_t$ as the act with those outcomes. We then do a similar procedure for the lowest utility to define $\underline{n}_t$. The preference ordering relative to n_t follows from the weak monotonicity of preference over acts in terms of utility vectors which is derived from part (a) of monotonicity (A8) and the representation in 2(b). Next, by continuity of $\succeq_{y_t}$ (A2), cardinal invariance (A9), and the other axioms, there exists a unique $\alpha \in [0,1]$ such that $n_t \sim_{y_t} \alpha\overline{n}_t + (1-\alpha)\underline{n}_t$. Now define for each act $n_t \in N_t$, $K(n_t) = K(\alpha\overline{n}_t + (1-\alpha)\underline{n}_t) = \alpha\sum w_{y_t}\overline{n}_t + (1-\alpha)w_{y_t}\underline{n}_t$. As constructed, since $(\sum w_{y_t}n_t) > (\sum w_{y_t}n_t')$ (i.e. the first vector is strictly greater in every element than the second vector) implies $n_t \succ_{y_t} n_t'$ by part (b) of monotonicity (A8), $K(\cdot)$ represents preferences on N_t in the sense that $n_t \succeq_{y_t} n_t'$ if and only if $K(n_t) \geq K(n_t')$.

Fix w_{y_t} so that $w_{y_t} \geq 0$ for all $((j, x_{t+1}), s_{t+1})$. This is possible since w_{y_t} is unique only up to a positive affine transformation. Let V be the space of all bounded functions from S_{t+1} to $\mathcal{R}_+$. Elements of V are thus $|S_{t+1}|$-dimensional vectors of non-negative numbers.

Lemma 1 *There exists an $I : V \to \mathcal{R}$ such that*

(i) For all $n_t \in N_t$, $I((\sum w_{y_t}n_t)) = K(n_t)$

(ii) I is monotonic (i.e. $a, b \in V, a \geq b \Rightarrow I(a) \geq I(b)$)

(iii) I is homogeneous of degree 1 (i.e. $a \in V, \alpha \geq 0 \Rightarrow I(\alpha a) = \alpha I(a)$)

(iv) I is C-Independent (i.e. $a \in V, \beta \geq 0 \Rightarrow I(a + (\beta, \ldots, \beta)) = I(a) + I(\beta, \ldots, \beta)$)

(v) I is superadditive (i.e. $a, b \in V \Rightarrow I(a+b) \geq I(a) + I(b)$).

Proof of Lemma:

Define I on the subset of V which is mapped out by vectors of the form $(\sum w_{y_t}n_t)$ for $n_t \in N_t$ by (i). Thus I represents preferences in the sense that $n_t \succeq_{y_t} n_t'$ if and only if $I((\sum w_{y_t}n_t)) \geq I((\sum w_{y_t}n_t'))$. For a constant utility act with constant utility $\beta \geq 0$, $I((\beta, \ldots, \beta)) = \beta$. I will now show that I is homogeneous of degree 1 on this subset of V. Let $n_t, n_t' \in N_t$ be such that $(\sum w_{y_t}n_t) = \alpha(\sum w_{y_t}n_t')$, where $\alpha \in (0,1]$. Now let $\tilde{n}_t$ be an act with constant utility $\sum w_{y_t}\tilde{n}_t = K(n_t')$. We know that such a $\tilde{n}_t$ exists and is indifferent to $n_t{}'$ by the construction of $K(\cdot)$. By cardinal invariance (A9), $n_t{}' \sim_{y_t} \tilde{n}_t$ implies $n_t \sim_{y_t} \alpha\tilde{n}_t$ where $\alpha\tilde{n}_t$ is a constant utility act with utility representation $(\alpha \sum w_{y_t}\tilde{n}_t)$. (Being careful, we must show that such an act exists. We know that there exists some constant utility act indifferent to n_t. If this act has constant utility less than or equal to $\alpha \sum w_{y_t}\tilde{n}_t$ then by convexifying between this and $\tilde{n}_t$ a constant act with the desired utility will exist and A9 will show that the utilities must have been equal. If, on the other hand, this act has constant utility greater than $\alpha \sum w_{y_t}\tilde{n}_t$ then we can use A9 and convexity of the set of acts to show that this constant act is indifferent to an act which has a utility vector which strictly dominates that of n_t. But then, since this strict dominance translates into strict preference, and preference is transitive, we would have contradicted the assumption that the constant utility act we started with was indifferent to n_t.) Now, by (i), $I(\alpha(\sum w_{y_t}\tilde{n}_t)) = \alpha \sum w_{y_t}\tilde{n}_t = \alpha I((\sum w_{y_t}\tilde{n}_t))$. Thus, $I((\sum w_{y_t}n_t)) = I(\alpha(\sum w_{y_t}\tilde{n}_t)) = \alpha I((\sum w_{y_t}\tilde{n}_t)) = \alpha I((\sum w_{y_t}n_t{}'))$. The case $\alpha > 1$ follows by dividing by α. So, I is homogeneous of degree 1 on the subset of V containing $(\sum w_{y_t}n_t)$. Now extend I to the rest of V by homogeneity. Note that this is possible since the construction of $K(\cdot)$ guarantees that the subset is spanned by constant utility acts, so extension by homogeneity covers any remaining parts of V. By construction I is homogeneous of degree 1 on V. As preferences are monotonic in $(\sum w_{y_t}n_t)$ (as shown earlier in the proof of the theorem) and homogeneity respects monotonicity, I is also monotonic on V.

Now I show that I is C-independent. Consider $a \in V$ and $\beta \geq 0$. In general, $I(a) = rI((\beta, \ldots, \beta))$ for some $r \geq 0$. Assume $r > 0$. By homogeneity, I can assume without loss of generality that $\frac{1+r}{r}a$ and $(1+r)((\beta, \ldots, \beta))$ are in the subspace of V which correspond to utility vectors generated by acts. Let n_t be an act such that $(\sum w_{y_t}n_t) = \frac{1+r}{r}a$, and let $n_t{}'$ be an act such that $(\sum w_{y_t}n_t{}') = (1+r)(\beta, \ldots, \beta)$. Now consider the act $n_t{}''$ which is a convex combination of n_t and $n_t{}'$ with weights $\frac{r}{1+r}$ and $\frac{1}{r+1}$ respectively. Since this is a positive affine transformation of the vector $\frac{1+r}{r}a$, cardinal invariance (A9) implies, taking the same positive affine transformation of $(1+r)((\beta, \ldots, \beta))$ (which does have an act associated with it, since this gives the same vector back again), that $n_t \sim_{y_t} n_t'$ implies $n_t'' \sim_{y_t} n_t'$. Thus, by preference representation, $I(a+(\beta,\ldots,\beta)) = I((1+r)(\beta,\ldots,\beta))$. Using homogeneity and the fact that $I(a) = rI((\beta,\ldots,\beta))$, I conclude that $I(a+(\beta,\ldots,\beta)) = I(a)+I((\beta,\ldots,\beta))$. The case $r=0$ follows from the case above. Thus I is C-independent.

The next step is to show that I is superadditive. Consider $a, b \in V$. Suppose $I(a) = I(b)$. Since I is always non-negative, if $I(a) = I(b) = 0$ then it must be that $I(a+b) \geq 0 = I(a)+I(b)$. Consider, then the strictly positive case. For some $\alpha > 0$ there must exist acts $n_t, n_t' \in N_t$ such that $\alpha(\sum w_{y_t} n_t) = a$ and $\alpha(\sum w_{y_t} n_t') = b$. Preference representation implies $n_t \sim_{y_t} n_t'$. Uncertainty aversion (A3) then implies that $\frac{1}{2}n_t + \frac{1}{2}n_t' \succeq_{y_t} n_t$. Thus, $I(\frac{1}{2\alpha}a + \frac{1}{2\alpha}b) \geq I(\frac{1}{\alpha}a)$. By homogeneity, $I(a+b) \geq 2I(a) = I(a)+I(b)$. Now suppose $I(a) > I(b)$ (the case $I(b) > I(a)$ follows from this one by exchanging a and b). Let $\gamma = I(a) - I(b)$, and define $c = b + (\gamma,\ldots,\gamma)$.

$$\begin{aligned} I(a+b)+\gamma &= I(a+b+(\gamma,\ldots,\gamma))\text{by C-independence} \\ &= I(a+c) \text{ by definition of } c \\ &\geq I(a)+I(c) \text{ by the previous case as } I(a)=I(c) \\ &= I(a)+I(b)+\gamma \text{ by C-independence.} \end{aligned}$$

Thus $I(a+b) \geq I(a)+I(b)$. *QED*

Now apply the following lemma stated and proved in Chateauneuf (1991), the key to which is a theorem of Fan (1956).

Lemma 2 *Let $\mathcal{A}$ be a σ-algebra of subsets of a set S, and let I be a functional on the set V of bounded $\mathcal{A}$-measurable functions from S to $\mathcal{R}_+$. Then the following are equivalent:*

(1) I satisfies:

(i) For all $\alpha \geq 0, \beta \geq 0, X \in V : I(\alpha X + \beta^*) = \alpha I(X) + \beta$, where β^* denotes the function with value β on all of S.

(ii) $X, Y \in V$ implies $I(X+Y) \geq I(X) + I(Y)$.

(iii) $X, Y \in V, X \geq Y$ implies $I(X) \geq I(Y)$.

(2) There exists a non-empty, closed, convex set C of additive probabilities on $\mathcal{A}$

such that $I(X) = \min_{p \in C} \int X dp$, for all $X \in V$. Furthermore, there exists a unique such closed (in the weak star topology), convex set C.

I have shown that I satisfies all of the properties in (1), thus applying the lemma to this setting yields the representation in 2(a) of the theorem. The uniqueness of C_{y_t} similarly follows. The condition 2(e) on the range of the w_{y_t} follows from axiom A10. This completes the proof that (1) implies (2). It is straightforward to check that the representation implies the axioms. *QED*

The above theorem gives us a characterization of preferences at each point in time and each history of state realizations and payoff realizations. These preferences allow for both uncertainty aversion and state dependence. With this as the building block, the next section will formally define dynamic consistency and impose it in the form of an axiom which knits together the various representations at different times and histories. This knitting together leads to a representation theorem for dynamic choice with uncertainty aversion.

2.4 Dynamic Consistency and a Dynamic Representation

I now present a dynamic consistency axiom. The fundamental idea of dynamic consistency in a choice setting is that preferences at time t over objects at time $t+1$ *contingent on any new information that will be available at time* $t+1$ should agree with preferences when time $t+1$ arrives and the information is realized.

Definition: (Preference over opportunity sets)

$x_{t+1} \succeq_{y_t, j_t, s^*_{t+1}} x'_{t+1}$ if and only if for each $n'_{t+1} \in x'_{t+1}$ there exists a $n_{t+1} \in x_{t+1}$ such that $n_{t+1} \succeq_{y_t, j_t, s^*_{t+1}} n'_{t+1}$. In other words, for any choice in the set x'_{t+1} there is a choice in x_{t+1} which is at least as good.

A11 (Dynamic Consistency) For all $t, y_t, s^*_{t+1} \in S_{T+1}, x_{t+1} \in X_{t+1}$, and $j_t \in J$,

(1) If $x_{t+1} \succeq_{y_t, j_t, s^*_{t+1}} x'_{t+1}$ then $n_t \succeq_{y_t} n_t{}', \forall n_t, n_t{}' \in N_t$ such that

(a) n_t and n_t' give degenerate lotteries (i.e. a single pair $j \times x_{t+1}$) in each $s_{t+1} \in S_{t+1}$,

(b) the same element $j_t \in J$ is given by n_t and n_t', and

(c) $n_t(s^*_{t+1}) = j_t \times x_{t+1}, n_t'(s^*_{t+1}) = j_t \times x'_{t+1}$, and $n_t(s_{t+1}) = n_t'(s_{t+1}), \forall s_{t+1} \neq s^*_{t+1}$.

(2) If $x_{t+1} \sim_{y_t, j_t, s^*_{t+1}} x'_{t+1}$ then $\forall n_t, n_t' \in N_t$ which satisfy (a), (b), and (c), $\hat{n}_t \hat{\sim}_{y_t} \hat{n}'_t$ for all full-support $\hat{n}_t, \hat{n}'_t \in \hat{N}_t$ such that $H(\hat{n}_t) = n_t$ and $H(\hat{n}'_t) = n_t'$ and such that $\hat{n}_t$ and $\hat{n}'_t$ have the same distribution over S_{t+1}.

(3) If $x_{t+1} \succ_{y_t, j_t, s^*_{t+1}} x'_{t+1}$ and $n_t \sim_{y_t} n_t'$ for some $n_t, n_t' \in N_t$ which satisfy (a), (b), and (c), then for those n_t, n_t', it must be that $\hat{n}_t \hat{\succ}_{y_t} \hat{n}'_t$ for all full-support $\hat{n}_t, \hat{n}'_t \in \hat{N}_t$ such that $H(\hat{n}_t) = n_t$ and $H(\hat{n}'_t) = n_t'$ and such that $\hat{n}_t$ and $\hat{n}'_t$ have the same distribution over S_{t+1}.

This axiom says that, given equal immediate payoffs, choice today over opportunity sets is governed by tomorrow's preferences over the elements of those sets, except possibly in cases where the preference is strict tomorrow. In this case, indifference is allowed today (thus allowing for the possibility of assigning a state zero probability) if and only if any corresponding state lottery with full support would agree with tomorrow's strict preference. This last requirement makes sure that assigning zero weight to a state is the only reason for such a divergence in preference. Finally, note that condition (2) clarifies situations of "accidental" agreement, where the acts are indifferent due to a zero probability and the opportunity sets are ex-post indifferent as well, and makes sure that preferences would have agreed even under positive probability. I now show that this axiom allows us to tie together the utility functions for each time and history that were derived in section 2.3. As might be expected, the tying together comes in a recursive way: today's utility of an opportunity set in a given state is a strictly increasing function of the value (in minimum expected utility terms) of the best element of that opportunity set tomorrow.

Theorem 2 *Adding axiom A11 to the axioms used in theorem 1, in addition to the conclusions of that theorem, A1-A11 are equivalent to the existence of functions* U_{y_t} : $\{(j, s_{t+1}, r) \in J \times S_{t+1} \times \mathcal{R} | r =$ $\max_{n_{t+1} \in x_{t+1}} \min_{p \in C_{y_t,j,s_{t+1}}} \sum_{S_{t+2}} p(s_{t+2}) \sum_{J \times X_{t+2}} w_{y_t,j,s_{t+1}}((j, x_{t+2}), s_{t+2})(n_{t+1}(s_{t+2})(j \times x_{t+2}))$ *for some* $x_{t+1} \in X_{t+1}.\} \rightarrow \mathcal{R}$*, which are strictly increasing in their third argument and which satisfy,*

$$
\begin{aligned}
w_{y_t}((j, x_{t+1}), s_{t+1}) \quad = & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.1) \\
& U_{y_t}(j, s_{t+1}, \max_{n_{t+1} \in x_{t+1}} \min_{p \in C_{y_t,j,s_{t+1}}} \sum_{S_{t+2}} p(s_{t+2}) \\
& \sum_{J \times X_{t+2}} w_{y_t,j,s_{t+1}}((j, x_{t+2}), s_{t+2})(n_{t+1}(s_{t+2})(j \times x_{t+2}))),
\end{aligned}
$$

where the w's and C's are the ones derived in theorem 1. For fixed w_{y_t}*'s the* U_{y_t}*'s are unique.*

Proof:

Fix w_{y_t}'s from theorem 1. Assume that such U_{y_t}'s exist and are unique. Let $V_{y_t}(x_{t+1}, s_{t+1}, j)$ be the value of the third argument of U_{y_t}. Note that, using the preference representations from theorem 1, $x_{t+1} \succeq_{y_t,j,s^*_{t+1}} x'_{t+1}$ if and only if $V_{y_t}(x_{t+1}, s^*_{t+1}, j) \geq V_{y_t}(x'_{t+1}, s^*_{t+1}, j)$. Also, $n_t \succeq_{y_t} n_t'$ if and only if $\min_{p \in C_{y_t}} \sum_{S_{t+1}} p(s_{t+1}) \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1})(n_t(s_{t+1})(j, x_{t+1})) \geq$ $\min_{p \in C_{y_t}} \sum_{S_{t+1}} p(s_{t+1}) \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1})(n_t'(s_{t+1})(j, x_{t+1}))$. Thus, since the U_{y_t} are assumed strictly increasing in the third argument, A11 (1) is verified. Examining the representation for state lotteries, $\hat{n}_t \hat{\succeq}_{y_t} \hat{n}'_t$ if and only if $\sum_{S_{t+1}} \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1}) \hat{n}_t((j, x_{t+1}), s_{t+1}) \geq$ $\sum_{S_{t+1}} \sum_{J \times X_{t+1}} w_{y_t}((j, x_{t+1}), s_{t+1}) \hat{n}'_t((j, x_{t+1}), s_{t+1})$, we see that A11 (2) and (3) are satisfied as well.

Now I must prove the other direction, and derive the U_{y_t} from the axioms. Again fix the w_{y_t}. Equation 2.1 uniquely defines the U_{y_t} if $V_{y_t}(x_{t+1}, s_{t+1}, j) = V_{y_t}(x'_{t+1}, s_{t+1}, j)$ implies $w_{y_t}((j, x_{t+1}), s_{t+1}) = w_{y_t}((j, x'_{t+1}), s_{t+1})$. This is implied by A11 (2) and the preference representation properties of V_{y_t} and w_{y_t} used in the other direction of this proof. To show U_{y_t} strictly increasing in the third argument it must be shown that $V_{y_t}(x_{t+1}, s_{t+1}, j) > V_{y_t}(x'_{t+1}, s_{t+1}, j)$ implies $w_{y_t}((j, x_{t+1}), s_{t+1}) > w_{y_t}((j, x'_{t+1}), s_{t+1})$. Since $V_{y_t}(x_{t+1}, s_{t+1}, j) > V_{y_t}(x'_{t+1}, s_{t+1}, j)$ implies $x_{t+1} \succ_{y_t, j, s_{t+1}} x'_{t+1}$ by the representation in theorem 1; A11 (1) and (3) and the representation in theorem 1 imply that $w_{y_t}((j, x_{t+1}), s_{t+1}) > w_{y_t}((j, x'_{t+1}), s_{t+1})$. *QED*

Notice that, by specifying utilities over J in the final period, beliefs in the initial period, a rule for the evolution of beliefs, and the functions U_{y_t} this theorem allows us to derive the rest of the utility functions in a recursive manner. This is very convenient from an analytical and computational point of view, especially when dependence on history is assumed to occur only through the evolution of beliefs. In such a case the U_{y_t} satisfy $U_{y_t}(j, s_{t+1}, r) = U_t(j, r)$. Further assuming that immediate and future utility are traded off in the same way at each time yields $U_t(j, r) = U(j, r)$. In the next section I will use such preferences to examine the problem of sequential price search without recall when the distribution of prices is unknown and the searcher is risk neutral, uncertainty averse, and has a constant rate of substitution between immediate and future utility.

However, before proceeding to the price search application, I want to briefly mention some circumstances in which the generality of the representation theorem is useful (but certainly not *required* for representing uncertainty aversion), in that dependence on history does not naturally occur only through updating beliefs. A classic example of "direct" (i.e. not simply through beliefs) state dependence occurs when preferences are defined in US dollar equivalents and part of the state is the current, say, French franc/dollar exchange rate. Presumably, the underlying source of my preferences on dollars is my preferences over what those dollars can buy. If the franc/dollar rate falls, all else equal, I can buy, for example, less *foie gras* and thus the utility of a given number of dollars is lower than it was under the previous exchange rate. Thus, preferences are state dependent in that the "prizes" (dollars) fluctuate in value with the state. In terms of the U_{y_t} functions derived in the representation, adding this prize state dependence to the dependence through updating gives $U_{y_t}(j, s_{t+1}, r) = U_t(Z_t(j, s_{t+1}), r)$.

Another example of "direct" state dependence might occur if part of the state gave the probability of living until the next time. Such a state would likely affect the weight assigned to utility generated in the future. This type of state dependence could be exhibited by writing $U_{y_t}(j, s_{t+1}, r) = U_t(j, D_t(s_{t+1}, r))$. Note that allowing time itself to also affect this probability is natural here.

One example of dependence of immediate payoffs on past realizations is the case of preferences leading to habit formation. In this example, past consumption of a good might increase the utility of consuming similar levels of the good today. Alternately, past consumption of gourmet meals may reduce the utility of consuming a gourmet meal tonight. This could be construed as a preference for intertemporal variety. Similarly, past and present state realizations could influence today's utility directly if, for example, states included information about how addictive past and present consumption goods were. Adding effects like these leads back to the most general form: $U_{y_t}(j, s_{t+1}, r)$.

An interesting question raised by these examples is whether there are easily stated restrictions on preferences which would restrict the U_{y_t} to one form or another. This turns out to be harder to answer than it might appear. A very partial answer is given in the following results.

First, it is useful to divide a history, y_t, into a history of past payoffs, z_t, and a history of past state realizations, q_t.

A12 (Payoff history independence)

For all $t, y_t(= (z_t, q_t)), n_t, n_t' \in N_t$, and $\hat{n}_t, \hat{n}_t' \in \hat{N}_t$, $n_t \succeq_{(z_t,q_t)} n_t'$ if and only if $n_t \succeq_{(z_t',q_t)} n_t'$, and $\hat{n}_t \hat{\succeq}_{(z_t,q_t)} \hat{n}_t'$ if and only if $\hat{n}_t \hat{\succeq}_{(z_t',q_t)} \hat{n}_t'$, $\forall z_t' \neq z_t$.

Theorem 3 *Adding axiom A12 to axioms A1-A11 yields the conclusions of theorem 1 and theorem 2 with y_t replaced everywhere by q_t.*

Proof: The conditions on preferences combined with the representations proved previously guarantee that the w_{y_t}, U_{y_t}, and C_{y_t} can be chosen so as not to vary with z_t. *QED*

Suppose that the valuation of prizes, J, is independent of time, state history, opportunity set, and state realization. The following axiom expresses this in terms of preferences.

A13 (Prize valuation independence)

Fix q_t, x_{t+1}, and s_{t+1}. Consider all state lotteries such that probability 1 is attached to x_{t+1}, s_{t+1}. (i.e. $\sum_J \hat{n}_t((j, x_{t+1}), s_{t+1}) = 1$.) Let $\hat{n}_t$ and $\hat{n}'_t$ be any two such state lotteries. Pick $q'_{t'}, x'_{t'+1}$, and $s'_{t'+1}$ arbitrarily and let $\hat{n}''_{t'}$ and $\hat{n}'''_{t'}$ be state lotteries which assign probability 1 to $q'_{t'}, x'_{t'+1}$ and give the same marginal distribution over J as $\hat{n}_t$ and $\hat{n}'_t$ respectively. It is true that $\hat{n}_t \hat{\succsim}_{q_t} \hat{n}'_t$ if and only if $\hat{n}''_{t'} \hat{\succsim}_{q'_{t'}} \hat{n}'''_{t'}$.

Theorem 4 *Adding axiom A13 to axioms A1-A12 is equivalent to the existence of a function $W : J \rightarrow \mathcal{R}$ and fuctions $\alpha_{q_t} : X_{t+1} \times S_{t+1} \rightarrow \mathcal{R}^+$ and $\beta_{q_t} : X_{t+1} \times S_{t+1} \rightarrow \mathcal{R}$ such that*

$$w_{q_t}((j, x_{t+1}), s_{t+1}) = \alpha_{q_t}(x_{t+1}, s_{t+1}) W(j) + \beta_{q_t}(x_{t+1}, s_{t+1}), \tag{2.2}$$

for all $t, q_t, j, x_{t+1}, s_{t+1}$.

Proof:

Fix t, q_t, x_{t+1}, and s_{t+1}. Define $W(j) = w_{q_t}((j, x_{t+1}), s_{t+1})$. Axiom A13 and the representation for state lotteries proved earlier imply that, for any $q'_{t'}, x'_{t'+1}, s'_{t'+1}$, $w_{q'_{t'}}((\cdot, x'_{t'+1}), s'_{t'+1})$ represents the same preferences over lotteries over J as $W(\cdot)$ does. Thus the $\alpha_{q'_{t'}}(x'_{t'+1}, s'_{t'+1})$ and $\beta_{q'_{t'}}(x'_{t'+1}, s'_{t'+1})$ are simply the slope and intercept of the positive affine transformation that relates the two functions. *QED*

The theorem shows that adding the assumption of prize valuation independence limits the role of state history, state, and opportunity set to (a) determining the weight given to immediate payoff utility relative to the future (through α), and (b) determining a utility from future opportunities (through β).

2.5 Search From an Unknown Distribution

Consider again the example of section 2.2. There is an item which you value at w dollars. The price of this item is set every day by an independent random draw from a fixed distribution of prices. You desire only one unit of the item and get no additional utility from having more than one unit. The process of going on a given day to check the price gives you disutility $c > 0$. Each day, you must decide whether to buy at that day's price and stop shopping or wait to see what tomorrow's price will be. You are risk neutral, have a discount rate δ, and have preferences that conform to the axioms in the previous sections. The uncertainty here concerns the true distribution of prices. According to the representation, beliefs can be represented by a closed, convex set of distributions over distributions (which of course translates into distributions over prices). The question I examine concerns the existence of a reservation price rule in this setting. In other words, at each time, given the prices realized previously, is there a price above which you will keep searching and below which you will stop and buy? Among the reasons for interest in reservation price rules is that they imply that sellers face well-defined downward sloping demand curves. In a well known paper, Rothschild (1974) shows that if the true distribution is multinomial, the searcher's prior over distributions is Dirichlet, the prior is updated in a Bayesian fashion, and the searcher is a risk neutral expected utility maximizer, then search follows a reservation price rule. More recently, Bikhchandani and Sharma (1989) have generalized this result to any combination of underlying distribution and learning process which results in the posterior given n observations equalling a convex combination of the prior and the empirical distribution determined by the n observations, and where the weights in this combination depend only on n, not on the specific realizations. Although I will stick to the case where prices have finite support, as in Rothschild, I will draw on the insight of Bikhchandani and Sharma that it is the taking of convex combinations of priors and empiricals which is important. I stick to a finite support of prices to match the finite state assumption in my representation. The logic of the proof does not depend in any essential way on this finiteness.

In particular, I am now in the position to ask whether the reservation price result extends to the case of an uncertainty averse decision maker who has a set of priors and who updates each one by taking a convex combination of the prior and the empirical distribution of the observations. Such a rule fits the mechanics of the preferences I described earlier since it is easy to show that a convex set of priors updated in this way gives rise to a convex set of posteriors. A special case would be a decision-maker with a closed, convex set of Dirichlet priors who updates each prior according to Bayes' rule.

Formally, let there be k possible prices. Let B be a non-empty, closed, convex set of probability measures on the space of distributions over these k prices. Let the utility of the item be w and the cost of each search be c. Fix a finite horizon $T > 0$. Let the utility at time T for time T outcomes be monetary value (thus risk neutrality is assumed). In terms of the representation in theorem 2, assume that the U_{y_t} functions all take the simple form $U(j, s, r) = j + \delta r$. In other words, utility for immediate payoffs is risk neutral at each time, and utility is discounted at rate δ between periods. All that remains to complete the specification of preferences is to describe the evolution of beliefs. To this end, it is more convenient to refer to beliefs by their cumulative distribution functions and to think of B as a set of cumulative distributions. Note that each distribution on distributions gives rise to a distribution on prices through the expectations operator, $E[\cdot]$. Define the empirical distribution on prices after observing prices $(p_1, \ldots, p_t)$ as,

$$H_{p_1,\ldots,p_t}(r) = \frac{1}{t}\sum_{i=1}^{t} 1_{[p_i,\infty)}(r) \tag{2.3}$$

where $1_{[p_i,\infty)}(r)$ equals 1 if $r \geq p_i$ and is 0 otherwise.

I assume, following Bikhchandani and Sharma (1989), that for each prior cumulative $F \in B$,

Assumption 1 *For all t, and observations $p_1, \ldots, p_t$, $E[F|p_1, \ldots, p_t](r) = (1 - a_t^F)E[F](r) + a_t^F H p_1, \ldots, p_t(r)$, for $a_t^F \in [0, 1]$.*

Thus the posterior distribution is a convex combination of the prior distribution and the empirical distribution. I am now ready to state and prove the reservation price rule result.

Theorem 5 *If an individual has preferences which are as described in the preceding paragraph, and the updating of beliefs satisfies assumption 1, then for the problem of search without recall when the set of possible prices is finite and the distribution of prices is unknown, the optimal stopping rule has the reservation price property.*

Proof:

I start by writing down $V_t(B)$, the utility of searching for one more period given that there are t periods remaining, that beliefs are B, and given the optimal continuation from that point on. $V_1(B) = w - [\max_{F\in B}\int_0^\infty r dE[F](r)] - c$, $V_2(B) = \min_{F\in B}\int_0^\infty \max\langle w-r, \delta V_1(B|r)\rangle dE[F](r) - c$. Thus, in general we have, $V_t(B) = \min_{F\in B}\int_0^\infty \max\langle w-r, \delta V_{t-1}(B|r)\rangle dE[F](r) - c$.

I now prove a useful lemma about the V_t.

Lemma 3 *Under the assumptions of the theorem, if $y_1 \geq y_2$ then $V_t(B|x^n, y_1) \leq V_t(B|x^n, y_2), \forall$ vectors of n observations $x^n, t \geq 1$.*

Proof of lemma:

For $t = 1$,

$$\begin{aligned}
& V_1(B|x^n,y_1) - V_1(B|x^n,y_2) \\
= & -\max_{F\in B|x^n,y_1}\int_0^\infty rdE[F](r) \\
& +\max_{F\in B|x^n,y_2}\int_0^\infty rdE[F](r) \\
= & -\max_{F\in B}[(1-a_{n+1}^F)\int_0^\infty rdE[F](r) \\
& +a_{n+1}^F\int_0^\infty rdH_{x^n,y_1}(r)] \\
& +\max_{F\in B}[(1-a_{n+1}^F)\int_0^\infty rdE[F](r) \\
& +a_{n+1}^F\int_0^\infty rdH_{x^n,y_2}(r)] \\
\leq & \max_{F\in B}[a_{n+1}^F[\int_0^\infty rdH_{x^n,y_2}(r) - \int_0^\infty rdH_{x^n,y_1}(r)]] \\
= & a_{n+1}^{F^*}[\frac{1}{n+1}(y_2-y_1)]
\end{aligned}$$

$$\leq 0, \text{ where } F^* \text{ is a maximizing distribution.}$$

Thus $V_1(B|x^n, y_1) \leq V_1(B|x^n, y_2)$, for all vectors of n observations x^n and observations $y_1 \geq y_2$.

Now I proceed by induction. Assume that for $t-1$, $V_{t-1}(B|x^n, y_1) \leq V_{t-1}(B|x^n, y_2)$, for all vectors of n observations x^n and observations $y_1 \geq y_2$.

I know that

$$\begin{aligned}
& V_t(B|x^n, y_1) - V_t(B|x^n, y_2) \\
= & \min_{F \in B|x^n, y_1} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1, r)\rangle dE[F](r) \\
& - \min_{F \in B|x^n, y_2} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2, r)\rangle dE[F](r).
\end{aligned}$$

By the induction hypothesis each integrand is non-increasing in r. Furthermore, the updating assumption in the statement of the theorem implies that $E[F|x^n, y_1]$ first-order stochastically dominates $E[F|x^n, y_2]$, $\forall F, x^n, y_1 \geq y_2$.

Thus the expression above is less than or equal to the same expression with the set $B|x^n, y_2$ substituted for $B|x^n, y_1$ in the first term:

$$\begin{aligned}
& \min_{F \in B|x^n, y_2} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1, r)\rangle dE[F](r) \\
& - \min_{F \in B|x^n, y_2} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2, r)\rangle dE[F](r) \\
= & \min_{F \in B|x^n, y_2} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, r, y_1)\rangle dE[F](r) \\
& - \min_{F \in B|x^n, y_2} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, r, y_2)\rangle dE[F](r) \\
\leq & \ 0.
\end{aligned}$$

The equality follows since Assumption 1 implies that changing the order of the observations does not change the set of beliefs. The inequality follows from the induction hypothesis. This proves $V_t(B|x^n, y_1) \leq V_t(B|x^n, y_2), \forall$ vectors of n observations $x^n, t \geq 1$, and $y_1 \geq y_2$. *QED*

Using the lemma, I proceed by showing that for all collections of n observations, x^n, and for all $t \geq 1$, $y_1 \geq y_2$ implies

$$V_t(B|x^n, y_1) - V_t(B|x^n, y_2) \geq y_2 - y_1. \tag{2.4}$$

To see that this suffices, note that it is optimal to stop after observing (x^n, y_1) with t periods remaining if and only if

$$w - y_1 \geq \delta V_{t-1}(B|x^n, y_1)$$

which is equivalent to

$$w \geq \delta V_{t-1}(B|x^n, y_1) + y_1.$$

But equation 2.4 implies that

$$V_t(B|x^n, y_1) + y_1 \geq V_t(B|x^n, y_2) + y_2,$$

and, using the lemma,

$$\delta V_t(B|x^n, y_1) + y_1 \geq \delta V_t(B|x^n, y_2) + y_2 \text{ for any } \delta \in [0, 1].$$

Therefore, whenever it is optimal to stop after drawing y_1 it is also optimal to stop if any lower price, y_2, is observed. This is precisely the reservation price property.

In fact, I will prove something stronger than equation 2.4, I will prove that

$$V_t(B|x^n, y_1^m) - V_t(B|x^n, y_2^m) \geq y_2 - y_1, \ \forall x^n, t, y_2 \leq y_1, \text{ and positive integers } m \tag{2.5}$$

where y_i^m means m observations of y_i.

The proof is again by induction on t. For $t = 1$,

$$\begin{aligned} & V_1(B|x^n, y_1^m) - V_1(B|x^n, y_2^m) \\ = & \max_{F \in B|x^n, y_2^m} \int_0^\infty r dE[F](r) \end{aligned}$$

$$
\begin{aligned}
& - \max_{F \in B|x^n, y_1^m} \int_0^\infty r dE[F](r) \\
= \; & \max_{F \in B}[(1 - a^F_{n+m}) \int_0^\infty r dE[F](r) + a^F_{n+m} \int_0^\infty r dH_{x^n, y_2^m}(r)] \\
& - \max_{F \in B}[(1 - a^F_{n+m}) \int_0^\infty r dE[F](r) + a^F_{n+m} \int_0^\infty r dH_{x^n, y_1^m}(r)]
\end{aligned}
$$

Let $F^* \in B$ be a maximizer of the second term.

$$
\begin{aligned}
\geq \; & (1 - a^{F^*}_{n+m})(0) + a^{F^*}_{n+m}[\int_0^\infty r dH_{x^n, y_2^m}(r) - \int_0^\infty r dH_{x^n, y_1^m}(r)] \\
= \; & a^{F^*}_{n+m}[\frac{m}{n+m}(y_2 - y_1)] \\
\geq \; & y_2 - y_1, \text{since } y_2 \leq y_1.
\end{aligned}
$$

Now suppose that

$$
V_{t-1}(B|x^n, y_1^m) - V_{t-1}(B|x^n, y_2^m) \geq y_2 - y_1, \forall x^n, y_2 \leq y_1, \text{ and positive integers } m.
$$

$$
\begin{aligned}
& V_t(B|x^n, y_1^m) - V_t(B|x^n, y_2^m) \\
= \; & \min_{F \in B|x^n, y_1^m} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dE[F](r) \\
& - \min_{F \in B|x^n, y_2^m} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dE[F](r) \\
= \; & \min_{F \in B}[(1 - a^F_{n+m}) \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dE[F](r) \\
& + a^F_{n+m} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dH_{x^n, y_1^m}(r)] \\
& - \min_{F \in B}[(1 - a^F_{n+m}) \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dE[F](r) \\
& + a^F_{n+m} \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dH_{x^n, y_2^m}(r)] \\
\geq \; & \min_{F \in B}[(1 - a^F_{n+m})(\int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dE[F](r) \\
& - \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dE[F](r)) \\
& + a^F_{n+m}(\int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dH_{x^n, y_1^m}(r) \\
& - \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dH_{x^n, y_2^m}(r))].
\end{aligned}
$$

Let F^* be a minimizer. Observe that $V_{t-1}(B|x^n, y_i^m, r) = V_{t-1}(B|x^n, r, y_i^{m-1}, y_i)$. By the Lemma, $V_{t-1}(B|x^n, r, y_1) \leq V_{t-1}(B|x^n, r, y_2), \forall y_2 \leq y_1$, so that $V_{t-1}(B|x^n, r, y_1^m) \leq V_{t-1}(B|x^n, r, y_2^m)$. Thus, it is true that $\max\langle w - r, \delta V_{t-1}(B|x^n, r, y_1^m)\rangle - \max\langle w - r, \delta V_{t-1}(B|x^n, r, y_2^m)\rangle$ $\geq \delta(V_{t-1}(B|x^n, r, y_1^m) - V_{t-1}(B|x^n, r, y_2^m)) \geq \delta(y_2 - y_1)$ by the induction hypothesis. Also, since I know what $H_{x^n, y_i^m}(r)$ is, I can explicitly write out the integral over it.

Using these facts yields

$$
\begin{aligned}
& \min_{F \in B}[(1 - a^F_{n+m})(\int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dE[F](r) \\
& - \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dE[F](r)) \\
& + a^F_{n+m}(\int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dH_{x^n, y_1^m}(r) \\
& - \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dH_{x^n, y_2^m}(r))] \\
= \; & (1 - a^{F^*}_{n+m})(\int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_1^m, r)\rangle dE[F](r) \\
& - \int_0^\infty \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^m, r)\rangle dE[F](r)) \\
& + \frac{a^{F^*}_{n+m}}{n+m} \sum_{i=1}^{n} (\max\langle w - r, \delta V_{t-1}(B|x^n, x_i, y_1^m)\rangle \\
& - \max\langle w - r, \delta V_{t-1}(B|x^n, x_i, y_2^m)\rangle) \\
& + \frac{a^{F^*}_{n+m}}{n+m} m(\max\langle w - r, \delta V_{t-1}(B|x^n, y_1^{m+1})\rangle \\
& - \max\langle w - r, \delta V_{t-1}(B|x^n, y_2^{m+1})\rangle) \\
\geq \; & (1 - a^{F^*}_{n+m})(\delta(y_2 - y_1)) \\
& + \frac{a^{F^*}_{n+m}}{n+m}(n\delta(y_2 - y_1)) \\
& + \frac{a^{F^*}_{n+m}}{n+m}(m(y_2 - y_1)) \\
\geq \; & y_2 - y_1.
\end{aligned}
$$

This completes the proof of the induction step, thus the result holds for all t. *QED*

Thus, we see that the reservation price property for beliefs that satisfy Assumption 1 is robust to the presence of uncertainty aversion. At this point at least two remarks are in order. First, other properties of the optimal search in this situation may not be so robust. For example, if we are concerned with properties of the dynamic evolution of the value function V_t it is true that the dynamic evolution for an uncertainty averse searcher will in general follow a path which is not possible for an expected utility maximizer. To see this, observe that if we start with a EU maximizer whose prior coincides with the particular prior that the uncertainty averse searcher starts off using to evaluate the utility of continuing (i.e. the prior which initially gives the minimum expected utility for continuing out of the set of priors), then, in general, the uncertainty averse individual will stop sooner than the EU maximizer since the uncertainty averter can only switch to a worse distribution. Second, due to the fact that the decision at each time in the search problem is a simple "continue" or "stop", the search behavior of an uncertainty averse individual, as described above, could be generated by an individual who has state and state-history dependent preferences, but is a bayesian. However, the form of state dependence required would be rather strange as it would have to mimic the effects of an uncertainty averter switching between distributions in the belief set. Furthermore, if the setting were enriched, and more choices were available at some times, the uncertainty averter's behavior could no longer be represented by state (and state-history) dependence alone.

2.6 Conclusion

In this paper, I have derived a dynamically consistent theory of choice with uncertainty aversion. This theory is needed, as the existing static theory of uncertainty aversion cannot be extended in a consistent manner simply by updating beliefs. Additionally, this paper proves an extension of a reservation-price rule result of Rothschild(1974) and Bikhchandani and Sharma (1989) to the case of an uncertainty averse searcher. Similar preferences are used (without axiomatization) in Epstein and Wang (1992) to examine intertemporal asset pricing.

Further work will focus on applying these preferences to an analysis of extensive form games. Another area of potential research is in the testing of this theory through experiments that may distinguish the dynamically consistent uncertainty aversion developed here from other forms of uncertainty aversion.

Bibliography

[1] Anscombe, F.J. and R.J. Aumann [1963], "A Definition of Subjective Probability", *The Annals of Mathematical Statistics* 34, pp. 199-205.

[2] Bikhchandani, Sushil and Sunil Sharma [1989], "Optimal Search with Learning", mimeo, UCLA.

[3] Brown, P.M. [1976], "Conditionalization and Expected Utility", *Philosophy of Science* 43, pp. 415-419.

[4] Camerer, Colin and Martin Weber [1992], "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity", *Journal of Risk and Uncertainty* 5, pp. 325-370.

[5] Chateauneuf, A. [1991], "On the Use of Capacities in Modeling Uncertainty Aversion and Risk Aversion", *Journal of Mathematical Economics* 20, pp. 343-369.

[6] Chew, S.H. and Larry G. Epstein [1989], "The Structure of Preferences and Attitudes Towards the Timing of Uncertainty", *International Economic Review*, 30, pp. 103-117.

[7] Dow, J. and S. Werlang [1991], "Nash Equilibrium Under Knightian Uncertainty: Breaking Down Backward Induction", mimeo.

[8] Dow, J. and S. Werlang [1992], "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of a Portfolio", *Econometrica* 60, pp. 197-204.

[9] Ellsberg, D. [1961], "Risk, Ambiguity and the Savage Axioms", *Quarterly Journal of Economics* 75, pp. 643-669.

[10] Epstein, Larry G. and Michel Le Breton [1993], "Dynamically Consistent Beliefs must be Bayesian", *Journal of Economic Theory* 61, pp. 1-22.

[11] Epstein, Larry G. and Tan Wang [1992], "Intertemporal Asset Pricing Under Knightian Uncertainty", mimeo.

[12] Fan, K. [1956], "On Systems of Linear Inequalities", in *Linear Inequalities and Related Systems, Annals of Mathematical Studies* 38.

[13] Gilboa, Itzhak and David Schmeidler [1989], "Maxmin Expected Utility with Non-unique Prior", *Journal of Mathematical Economics* 18, pp.141-153.

[14] Karni, Edi [1985], *Decision Making under Uncertainty: The Case of State Dependent Preferences*, Cambridge, MA: Harvard University Press.

[15] Karni, Edi [1993], "Subjective Expected Utility Theory with State-Dependent Preferences", *Journal of Economic Theory* 60, pp. 428-438.

[16] Karni, Edi and Zvi Safra [1990], "Behaviorally Consistent Optimal Stopping Rules", *Journal of Economic Theory* 51, pp. 391-401.

[17] Karni, Edi, Schmeidler, David, and Karl Vind [1983], "On State Dependent Preferences and Subjective Probabilities", *Econometrica* 51, pp. 1021-32.

[18] Klibanoff, Peter [1992], "Uncertainty, Decision, and Normal-form Games", mimeo, MIT.

[19] Klibanoff, Peter [1993], "On Updating Sets of Measures", mimeo, MIT.

[20] Kreps, David M. and Evan L. Porteus [1978], "Temporal Resolution of Uncertainty and Dynamic Choice Theory", *Econometrica* 46, no.1, pp. 185-200.

[21] Laibson, David I. [1992], "Self-Control and Saving", mimeo, MIT.

[22] Machina, M. and D. Schmeidler, [1992], "A More Robust Definition of Subjective Probability", *Econometrica* 60, pp. 745-80..

[23] Pollack, R.A. [1968], "Consistent Planning", *Review of Economic Studies* 35, pp. 201-208.

[24] Savage, L.J. [1954], *The Foundations of Statistics.* New York: John Wiley.

[25] Skiadas, Costis [1991], "Conditioning and Aggregation of Preferences", Northwestern University, Center for Mathematical Studies in Economics and Management, Discussion Paper No. 1010.

[26] von Neumann, J. and O. Morgenstern [1947], *Theory of Games and Economic Behavior*, second edition. Princeton University Press.

Chapter 3

Decentralization, Externalities, and Efficiency (with Jonathan Morduch)

3.1 Introduction

Decentralization has many benefits; most importantly, it takes advantage of local information and gives individual firms, agents or localities control over their affairs. However, it also has costs; spillovers from one jurisdiction or firm to another can undermine efficiency in a decentralized system. For example, emissions from factories in the United States contribute to acid rain in Canada, and New Jersey's spending on public schools benefits employers in New York. In the absence of coordination, societies end up with too much smoke and too little education.[1] Lessons drawn from competitive analysis suggest that these inefficiencies become more severe as external effects increase in size.

We argue, drawing on insights from the literature on mechanism design and bargaining, that these lessons are misleading in more realistic settings in which there are attempts to coordinate local activities or activities among firms. Given the su-

[1] See, for example, Laffont [1988] and the survey by Rubinfeld [1987].

periority of local information and respect for the autonomy of individual localities or firms, outcomes with coordination will be efficient only when external effects are *relatively large.* In contrast, when external effects are relatively small, coordination cannot yield improvements at all. These results run counter to the classical logic that small externalities lead to small inefficiencies while large externalities give rise to large inefficiencies. The contrast arises because the larger is the externality, the greater is the potential gain from coordination; this makes it easier to design a program which is acceptable to all parties and within budget.

It is well understood that asymmetric information increases the expected costs of coordination and that this can make it difficult to obtain efficient outcomes through bargaining (e.g., Laffont and Maskin [1979], Myerson and Satterthwaite [1983], Cramton, Gibbons and Klemperer [1987], and Farrell [1987].) While the results in the previous literature are suggestive, there has been little work on the particular problems associated with introducing externalities into such contexts. [2]

The following example illustrates our intuition. Imagine that the state of New York will benefit from reduced acid rain if Ohio Electric builds a new, cleaner power plant to replace an existing facility. Since Ohio Electric has the right to decide whether to build the plant, New York's only way to affect the decision is to offer to compensate Ohio Electric in exchange for a promise that the new plant will be built. However, when the net benefits to Ohio Electric of building the plant are not publicly known, New York does not know how much it needs to pay in order to secure an agreement. For example, it might be in Ohio Electric's interest to build a new plant anyway, in which case New York need pay nothing.

[2] Farrell [1987] presents a simple example of bargaining in the presence of externalities, in which private information can lead to inefficiencies. The example highlights the role of the individual rationality constraint (i.e., autonomy) as in Myerson-Satterthwaite [1983], but he does not consider the range of issues taken up here. Greenwood-McAfee [1991] address externalities and asymmetric information in the context of education; their paper centers on a case in which monotonicity conditions are binding (e.g., the government wants to devote extra resources to slow and fast learners, but not to average learners). This yields the inefficiency in their model, rather than individual rationality — which they do not impose. Pratt and Zeckhauser [1987] also do not consider individual rationality constraints; as above, they show that efficiency can be obtained in a wide range of environments through taxes and subsidies based on "expected externalities". In the present context, it is natural to assume that monotonicity is not a binding constraint, and we highlight the ways in which autonomy and externalities interact with private information to limit efficient coordination.

This uncertainty interferes with the ability to reach a mutually acceptable agreement in two ways. First, it increases the expected costs of coordination and, second, it decreases the expected benefits. Imagine that New York would benefit by $\$w$ if Ohio Electric built the plant. Ohio Electric would surely agree to build in exchange for an offer of $\$w$ if it stood to lose no more than that from the plant. But Ohio Electric's net loss on the plant could be a good deal less than $\$w$ (it might even have a net gain), and New York can't tell. Thus, New York expects to "overpay" Ohio Electric relative to compensation needed in a world of perfect information. In particular, suppose that New York offers $\$x$ in return for Ohio Electric building the new plant. New York knows that Ohio Electric will accept this offer if its net valuation of the plant is above -$\$x$. If Ohio Electric's net valuation is strictly larger than -$\$x$, New York will have overpaid and Ohio Electric will get an "informational rent" from its private information.

The second complication arises from uncertainty about Ohio Electric's actions *without an agreement.* In making its choices, New York considers two scenarios. In the first scenario, an agreement is reached, payment is made, and Ohio Electric promises to build the plant. In the second scenario, there is no agreement, but Ohio Electric might choose to build the plant of its own accord (if Ohio Electric's net valuation is positive.) The expected net benefit of coordination for New York is the *difference* in expected outcomes with coordination and without. When there is a positive probability that Ohio Electric will take the desired action on its own, New York's expected benefit from coordination will always be less than the full value of eliminating the externality. [3]

Taken together, the increase in expected costs and the decrease in expected benefits limits the attractiveness of coordination from New York's perspective. We can express these costs and benefits in a straightforward and compact way. Assume that Ohio Electric's net valuation of building a new plant to replace the old one is in the interval $[\underline{v}, \overline{v}]$ where $\underline{v} < 0 < \overline{v}$, and that, from the point of view of an outsider, this

[3] So, as is further described in Section 3.3, New York would refuse to make a Pigouvian transfer to Ohio Electric — i.e. a subsidy equal to the full value of the externality ($\$w$).

valuation is distributed according to the cumulative distribution function $F(\cdot)$. The probability that Ohio Electric will build the plant if New York offers \$$x$ is $1 - F(-x)$ as depicted in Figure 1. Figure 1 also illustrates the probability $(1 - F(0))$ that Ohio Electric will build if no offer (an offer of zero) is made. So, if New York offers to pay Ohio Electric \$$x$ if the plant is built, the expected cost of this offer is \$$x$ times the probability that Ohio Electric's net valuation is above -\$$x$, or $x(1 - F(-x))$. The expected benefit to New York of making this offer is \$$w$ times the *increase* in the probability that Ohio Electric will build the plant compared to the case where New York does not offer any money. Formally, this is $w[(1 - F(-x)) - (1 - F(0))] = w(F(0) - F(-x))$. For an offer of \$$x$ to be beneficial to New York the expected benefits must outweigh the expected costs. This is true if and only if

$$w(F(0) - F(-x)) \geq x(1 - F(-x)). \tag{3.1}$$

This cost/benefit inequality is central to understanding when coordination will occur and how much it can accomplish.

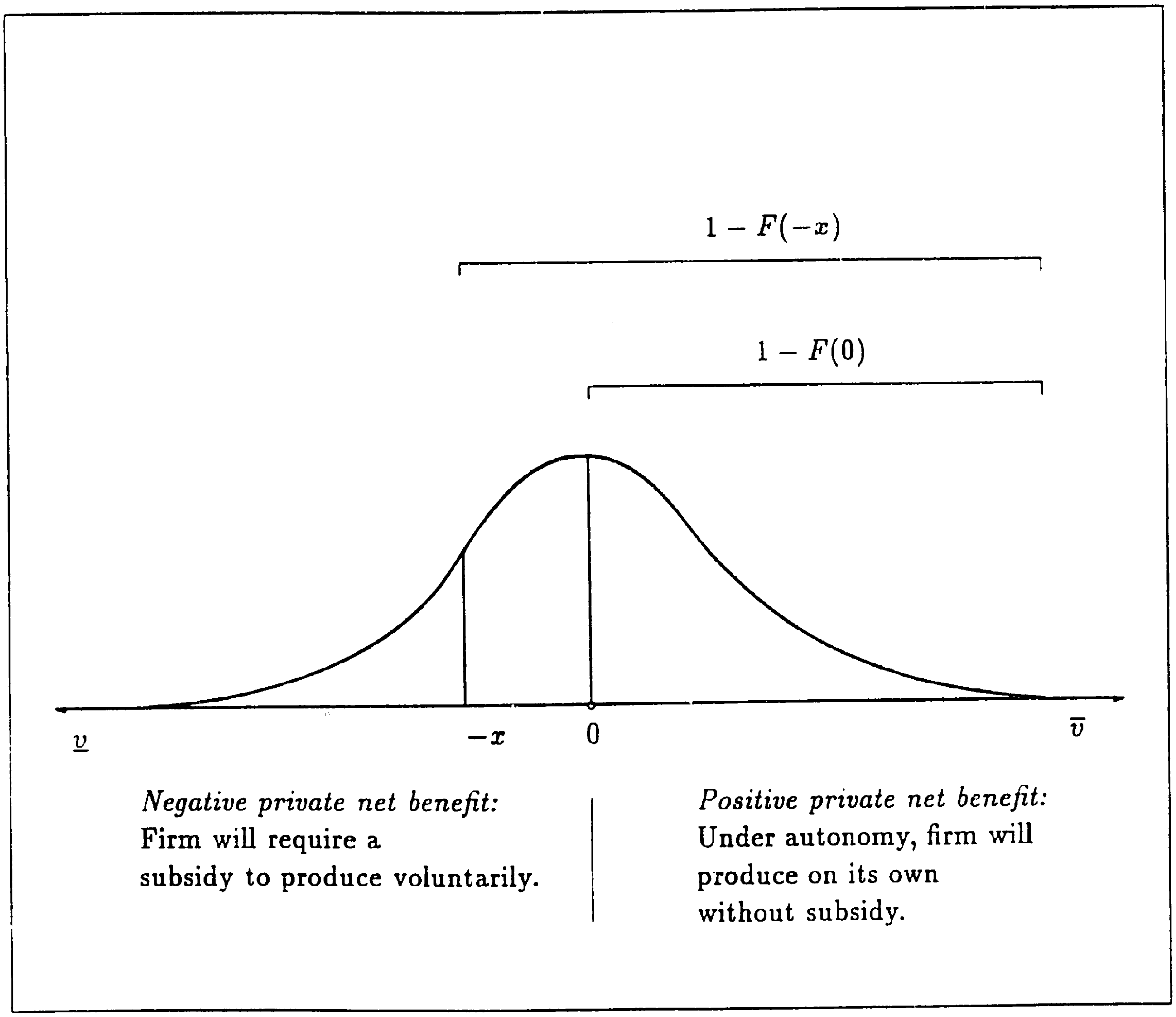
$1 - F(-x)$
$1 - F(0)$
$\underline{v}$
$-x$
0
$\overline{v}$
Negative private net benefit: Firm will require a subsidy to produce voluntarily.
Positive private net benefit: Under autonomy, firm will produce on its own without subsidy.

Figure 3-1: Distribution of Private Net Benefits, $f(v)$

We show that the combination of asymmetric information and externalities can increase expected costs and decrease expected benefits to such a degree that, for a wide range of cases, the outcome under a coordinated agreement cannot improve on the autonomous allocation — i.e., it is no better than doing nothing at all. For example, returning to equation (3.1), suppose that $F(\cdot)$ is the cdf for the uniform distribution on $[\underline{v}, \overline{v}]$ (i.e. $F(x) = \frac{x-\underline{v}}{\overline{v}-\underline{v}}$). Then the condition for expected benefits to exceed expected costs becomes $w \geq \overline{v} + x$. Thus if the externality, w, is smaller than the largest possible net benefit, $\overline{v}$, no positive offers will be made and no improvement over the autonomous allocation is possible. This finding may help to explain why coordination is so infrequently observed in practice. In a sense, our model shows how asymmetric information and autonomy generate "transactions costs" endogenously, in the form of informational rents. Commonly cited reasons such as high "exogenous" transactions costs or ill-defined property rights need not be operative (Coase [1960]).

This result contrasts with bargaining results in contexts without externalities in which (as long as there is a positive probability of gains from trade *ex ante*) there is always a second-best mechanism which offers at least some advantage over the autonomous allocation. In the example above, if $\underline{v} = -1/2, \overline{v} = 1$, and $w = 1$ then there is simultaneously common knowledge that social gains from agreement exist for all possible net valuations and no mutually acceptable improvement over the no agreement (autonomous) outcome.

We thus arrive at a fundamental paradox. We argue that the same forces which make decentralization appealing — respect for the autonomy and superior information of localities or firms — conspire to undermine the efficiency of the system in the presence of externalities.[4] Previous work has shown that, in the presence of externalities, neither attribute alone necessarily leads to inefficiencies, since efficient outcomes

[4] Our framework takes as given that the autonomy of localities or firms is inviolable. Presumably, in framing a constitution, a degree of autonomy is guaranteed in order to protect localities against the possibility that future governments will abuse their power (Madison, Hamilton and Jay [1787] *Federalist* X). As a recent example, enhancing local autonomy has been a key point of the political changes in China; the political reforms have strengthened economic reforms by ceding authority to provincial and county governments and thus making future reversals more difficult (Weingast [1993]). An argument of the present paper is to show that these guarantees can have costs in terms of forsaken efficiency.

can be achieved through Pigouvian taxation, where only asymmetric information is at issue, or through decentralized "Coasian" bargaining, where only autonomy is at issue. Here, however, we show that *in combination* these two fundamental elements of decentralized systems place limits on the ability to internalize externalities.

The next section describes and solves the problem of designing the cptimal coordination policy. Section 3.3 interprets the results in terms of the intuition developed above, characterizes optimal transfers, and describes when improvements can be made over the autonomous allocation and when the first-best allocation can be achieved. Section 3.4 considers extensions of the basic model, and Section 3.5 describes potential applications to choices by firms about research and development, assumptions underlying the "new growth theory", and the siting of environmental hazards like a waste dump or polluting factory. Section 3.6 concludes.

3.2 The Model

For ease of exposition, we describe the model in terms of two firms rather than a firm and a state, as in the example of the previous section. The model itself is general enough to encompass several interpretations. We discuss some of these in Section 3.4.

We begin by assuming that there are two firms, $i = 1, 2$, each solely concerned with its own welfare. Firm 1 has a project which it could undertake; this might be, for example, building a new plant or introducing a new worker training program.

The benefits of undertaking the project do not accrue just to the firm which undertakes it — there may also be spillovers to the other firm. The value of the spillovers is given by w^*. In the case of a worker training program, for example, there may be positive externalities ($w^* > 0$), as some of the trained workers may go to work for the other firm. The spillover parameter w^* is assumed to be public knowledge, whereas the private net benefit of the project is known within the firm that can undertake it only. This information structure arises because the firm has special knowledge about the cost or profitability of the project, while outsiders do not.

Formally, welfare is determined by the investment, X, in the project. This variable is binary (either 0 or 1).[5] Firm 1's objective function is given by:

$$u_1(X, v, t) = vX + t, \tag{3.2}$$

where $v \in [\underline{v}, \overline{v}]$ is a parameter which reflects the private net benefit of the project; it is drawn from distribution $F(\cdot)$ with strictly positive, continuous density $f(\cdot)$ on the domain $\underline{v} < 0 < \overline{v}$.[6] Net transfers from firm 2 to firm 1 are given by t.

Firm 2's objective function is given by:

$$u_2(X, w^*, t) = w^*X - t. \tag{3.3}$$

Given these objectives, we consider the ability to coordinate the activities of the firms. A government (or mediator) attempts to achieve efficient outcomes by offering an appropriately designed menu of options to the firms. A given option provides a subsidy/tax coupled with a production plan (specified as a probability of producing), based on the announced net benefits of production $\hat{v}$.

We model the problem as a three-stage game. In the first stage, the government proposes the menu of options to the firms. In the second stage, each firm accepts or rejects the menu. Then, in the third stage, if *both* accept, the programs are implemented with enforcement by the government. Otherwise, there is no agreement, and firm 1 is free to pursue its production decision independently.

Note that this is not the most general proposal that we could allow. Consider the case of a positive externality. Suppose that the government could sign a contract with firm 1 (*without* the approval of firm 2) which required firm 1 not to undertake the project unless firm 2 agreed to pay a transfer of w^*. Firm 1 would agree to sign such a contract, since if firm 2 believed that firm 1 would not do the project absent

[5] The assumption that the project is {0,1} is equivalent to assuming constant returns to scale in production and constant marginal benefits, along with an upper bound on project size — i.e., that there is a linear objective function with continuous project choice from an interval $[0, \overline{X}]$.

[6] Note that we have no inefficiency if v is *always* less than zero or if v is *always* greater than zero, since there is then no ambiguity as to whether firm 1 will produce or not under autonomy. In this case, either autonomy is efficient or Pigouvian taxes will work.

an agreement to pay w^*, firm 2 would indeed be willing to pay w^* contingent on production. Thus, the socially efficient outcome would result. [7]

The key assumption needed here is that the above side-contract is a credible one. However, the government and firm 1 have an incentive to secretly negotiate an escape clause which says that if firm 2 does not agree to pay w^*, then firm 1 is free to choose whether to produce or not. For this type of scheme to work, therefore, one of the parties must be able to credibly commit not to negotiate such a clause. Such a commitment might be plausible for a long-lived, patient government which knows that it will be involved in many such mechanisms and can develop a reputation for not secretly (re)negotiating. However, a reputation story could also work for firm 2 if it was to be involved in many similar circumstances and found it desirable to build a reputation for not giving in to such contracts. Furthermore, it may be difficult to verify government enforcement of punishments specified in the side-contract with firm 1.

In general, our inclination is that the level of commitment needed to make these "rent extraction" contracts credible is very high. Consequently, we focus on the no side-contracts case as an upper bound on what is achievable in most situations.[8]

Now we proceed to state and solve the problem. As in similar problems of mechanism design, the government's problem is simplified via the revelation principle, which states that, without loss of generality, the menu of programs can be limited to direct revelation mechanisms which induce truth-telling.[9] We thus consider direct revelation mechanisms of the form:

$$\langle p(v),\ t(v)\rangle, \tag{3.4}$$

[7] We thank Eric Maskin for suggesting this type of contract to us.

[8] The reader may be wondering why our setup requires any less commitment than the one we are ruling out. The mechanism design modelling makes it seem that, after firm 1 has revealed its private information, the two firms and the government might have an incentive to renegotiate the mechanism. In general, this is true; however, for our problem, we show in Section 3.3 that the optimal mechanism can be implemented by a tax/subsidy contingent on production. Thus, the only time that any information gets revealed is when firm 1 actually decides to undertake the project or not. When that decision has been made, the tax or subsidy is the only thing left to negotiate about, and, since it is simply a transfer between firms there is no scope for renegotiation.

[9] Fudenberg and Tirole [1991], Chapter 7, e.g., provides a good overview of issues in mechanism design and the revelation principle.

which gives the probability of producing the project and the net monetary transfer from firm 2 to firm 1 as a function of firm 1's type (the fact that p is a probability requires that $0 \leq p(v) \leq 1, \forall v$.) By allowing only for a transfer between firms, we are imposing budget balance. Budget balance is natural in considering a decentralized setting since it restricts attention to programs which do not require support from higher authorities.[10]

In evaluating a given menu, firms consider expectations of production plans and net transfers under an agreement. The expectations of firm 1 are conditional on the private net benefit v of producing, as this is known to it. Those of firm 2 however are not. Accordingly, define:

$$\begin{aligned} P &\equiv E(p(v)), \\ T &\equiv E(t(v)). \end{aligned}$$

If there is an agreement, firm 1's expected probability that it will produce is given by $p(v)$; P gives firm 2's expectation that firm 1 will produce under an agreement; and $t(v)$ gives firm 1's expected net transfers, while -T gives firm 2's expected net transfers. The firms' expected utility under an agreement as a function of type is then:

$$U_1(v) = vp(v) + t(v), \tag{3.5}$$

$$U_2 = w^* P - T. \tag{3.6}$$

The government maximizes the sum of expected utilities over all firms, weighting

[10] Note that introducing other firms which are also autonomous — but not affected by these projects — does not relax budget balance in a way relevant to the present problem. In Section 3.4.3 and Appendix B.5 we analyze a case in which budget balance is not required.

production according to the valuations of both firms affected :[11]

$$\max_{p(\cdot)} \int_{z=\underline{v}}^{\overline{v}} (U_1(z) + U_2) f(z) dz = \max_{p(\cdot)} \int_{z=\underline{v}}^{\overline{v}} (z + w^*) p(z) f(z) dz \tag{3.7}$$

subject to incentive compatibility (IC) and individual rationality (IR) constraints:

$$\text{(IC)} \qquad U_1(v) \geq vp(\hat{v}) + t(\hat{v}), \ \ \forall v, \hat{v},$$

$$\text{(IR1)} \qquad U_1(v) \geq \max(v, 0), \ \ \forall v,$$

$$\text{(IR2)} \qquad U_2 \geq w^*(1 - F(0)).$$

The incentive compatibility constraint ensures that firm 1 has no incentive to misrepresent its type, and the individual rationality constraints ensure that expected utility under an agreement for each firm is at least as great as that without.[12,13] Here, we see the role of respect for autonomy, which requires that participation must be strictly voluntary. Unlike common problems of bargaining over control of a single good without externalities (e.g., Myerson-Satterthwaite [1983]), the autonomous (i.e., "no trade") position can involve utility generated by the actions of the other party

[11] While we assume here that the government is utilitarian (in that it wishes to maximize the unweighted sum of utilities over firms), our model applies equally well to decentralized bargaining. For example, if the objective function puts zero weight on firm 1, this corresponds to a bargaining process in which firm 2 makes a take-it-or-leave-it offer to firm 1. The results in this case are the same as those presented in Appendix B.5 except that the $\frac{\lambda}{1+\lambda}$ is replaced by 1. Thus the qualitative features all carry over from the analysis of the evenly-weighted case.

[12] Under autonomy (i.e. in the absence of an agreement) firm 1 produces on its own if $v > 0$. Since each firm knows this, from the point of view of firm 2, firm 1 will produce with probability $(1 - F(0))$ under autonomy. Thus, without coordination, firm 1 will obtain $\max(v, 0)$ from its own production and firm 2 expects to get $w^*(1 - F(0))$ from firm 1's production.

[13] Note that under autonomy each firm has a dominant strategy; thus, we do not need to consider the possibility that the information revealed in the mechanism approval/disapproval stage will affect the autonomous outcome.

even if participation is rejected.[14,15]

The problem in equation (3.7) is not easy to solve since there is a continuum of constraints. The following theorem allows us to reduce the set of constraints to just two:

Theorem 1 *Suppose that $p(v)$ is non-decreasing in v. Then a direct revelation mechanism $\langle p(\cdot), t(\cdot)\rangle$ satisfies (IC), (IR1), and (IR2) if and only if:*

$$(\overline{A.1}) \quad \int_{z=\underline{v}}^{\overline{v}} p(z)\left(z + w^* + \frac{F(z)}{f(z)}\right) f(z)dz \geq \overline{v} + w^*(1 - F(0)),$$

and

$$(\underline{A.1}) \quad \int_{z=\underline{v}}^{\overline{v}} p(z)\left(z + w^* - \frac{1 - F(z)}{f(z)}\right) f(z)dz \geq w^*(1 - F(0)).$$

Proof of Theorem 1

See Appendix B.1.

Thus, we can write the central government's problem as

$$\max_{p(\cdot)} \int_{z=\underline{v}}^{\overline{v}} (z + w^*)p(z)f(z)dz.$$

subject to

$$(\overline{A.1})$$

$$(\underline{A.1})$$

and $p(v)$ non-decreasing.

[14] The framework can be naturally extended to coordination among a number of different firms, as long as the assumption is maintained that agreement must involve *all* firms or none. While we have not formally investigated the case where some firms coordinate their activities while others opt out, this can only make efficiency more difficult to achieve since the individual rationality constraint will be made more stringent if a firm expects others to agree even if it opts out. So, again, our results can be seen as placing an upper bound on what is achievable without extraordinary commitment. A more complete treatment of the multiple firm case would examine issues of coalition formation and how partial acceptance of mechanisms would affect the form of the second-best outcome.

[15] The presence of this type-contingent outside option for firm 1 can create countervailing incentives (see e.g. Lewis and Sappington [1989] and Maggi and Rodriguez [1993]) in our problem. Whether the incentive to overstate v or understate v is dominant will determine where (IR1) binds.

We make the additional assumptions that

$$\frac{d}{dv}\left(\frac{F(v)}{f(v)}\right) \geq 0 \tag{3.8}$$

and

$$\frac{d}{dv}\left(\frac{1-F(v)}{f(v)}\right) \leq 0. \tag{3.9}$$

These assumptions are satisfied for many common distributions and ensure that the monotonicity constraint is satisfied at the solution.[16] We can now solve using Kuhn-Tucker multipliers. Let $\overline{\lambda} \geq 0$ be the multiplier on $(\overline{A.1})$ and $\underline{\lambda} \geq 0$ be the multiplier on $(\underline{A.1})$. We can rewrite the maximization problem as:

$$\begin{aligned}\max_{p(\cdot)} \quad & \int_{z=\underline{v}}^{\overline{v}} \left((1+\overline{\lambda}+\underline{\lambda})(z+w^*) + \overline{\lambda}\frac{F(z)}{f(z)} + \underline{\lambda}\frac{(-1+F(z))}{f(z)}\right) p(z)f(z)dz \\ & - \overline{\lambda}(\overline{v}+w^*(1-F(0))) - \underline{\lambda}(w^*(1-F(0))).\end{aligned}$$

The first order conditions yield that production by firm 1 is determined by a simple "cut-off" rule:

$$p(v) = \begin{cases} 1 & \text{if } v + w^* + \frac{\overline{\lambda}}{1+\overline{\lambda}+\underline{\lambda}}\frac{F(v)}{f(v)} + \frac{\underline{\lambda}}{1+\overline{\lambda}+\underline{\lambda}}\frac{(F(v)-1)}{f(v)} \geq 0. \\ 0 & \text{otherwise} \end{cases}$$

While the setup allowed that the optimal agreement could incorporate an element of randomization, the result above is in fact straightforward to implement: firm 1 produces with probability equal to 1 if its announced valuation is above a cut-off value and does not produce otherwise. We will show that in the case of positive externalities, firm 1 then receives a simple matching grant for producing, and in the case of negative externalities, it faces a simple per unit tax.

The first two terms of the cut-off rule, $v + w^*$, give the social benefit of firm 1's production. If these were the only terms, we would have the rule: produce if and only

[16] Bagnoli and Bergstrom [1989]. The distributions include the uniform, normal, exponential, logistic, chi-squared, Laplace, and, with parameter restrictions, the Weibull, gamma, and beta distributions.

if social benefits of production are positive, which is the first-best outcome. However, the presence of asymmetric information leads to the addition of the final two terms (the first of these terms is positive and the second is negative, with weights given by which constraints are binding.) These two terms give deviations from first-best production levels, and in the next section we describe how they affect the limits to coordination.

First, note that the fact that the solution takes the form of a cut-off rule makes it easy to see which constraint, $(\overline{A.1})$ or $(\underline{A.1})$, will be binding. Let $\tilde{v}$ be the cut-off value defined by the first-order condition. Constraint $(\underline{A.1})$ requires that

$$\int_{\tilde{v}}^{\overline{v}} [-1 + F(z) + (z + w^*)f(z)]dz \quad \geq \quad w^*(1 - F(0))$$

or, simplifying,

$$w^*(F(0) - F(\tilde{v})) \geq -\tilde{v}(1 - F(\tilde{v})). \tag{3.10}$$

This condition is the relevant one when externalities are positive — i.e., when more production is desirable ($\tilde{v} < 0$). Notice that this is the same cost/benefit inequality as equation (3.1) of Section 3.1 except that x is chosen optimally to equal $-\tilde{v}$. Similarly, constraint $(\overline{A.1})$ requires that

$$w^*(F(0) - F(\tilde{v})) \geq \tilde{v}(F(\tilde{v})). \tag{3.11}$$

This condition is relevant for negative externalities, when less production is desirable ($\tilde{v} > 0$).[17] We provide intuition for these conditions below.

3.3 What Can Coordination Achieve?

The conditions above have a simple interpretation in terms of the expected costs and benefits from coordination and lead to an easily implemented system of optimal

[17] Note that at the autonomous allocation ($\tilde{v} = 0$), where production is neither encouraged nor discouraged by the program, both $(\overline{A.1})$ and $(\underline{A.1})$ are binding.

transfers. Following this interpretation, we show when coordination can make any improvements at all over the autonomous allocation and when the first-best outcome can be achieved.

3.3.1 Expected Costs and Benefits of Coordination

The constraints ($\underline{A.1}$) and ($\overline{A.1}$), simplified as equations (3.10) and (3.11), have a straightforward interpretation. Each says that the expected benefits to firm 2 of setting the cut-off at type $\tilde{v}$ must outweigh the expected costs associated with that cut-off.

Following the intuition in the introduction, in the case of positive externalities, the right hand side of equation (3.10), $-\tilde{v}(1 - F(\tilde{v}))$, gives exactly the expected cost of the subsidies necessary to implement a cut-off of $\tilde{v} < 0$. This is because if the cut-off type is paid $-\tilde{v}$ all other types that produce must receive the same amount since net benefits are not observed.

Similarly, with negative externalities, the expected cost of paying the subsidy is $\tilde{v}F(\tilde{v})$, reflecting the fact that all types $v \geq \tilde{v}$ must be paid *not* to produce. This cost is the right hand side of equation (3.11).

Figure 2 illustrates these expected costs for the case of positive externalities. The horizontal axis gives net costs faced by the cut-off type. The upper curve reflects expected costs under asymmetric information, $-\tilde{v}(1 - F(\tilde{v}))$. In contrast, the lower curve shows expected costs in a world with perfect information. In this "Coasian" setting, expected costs are just $-\int_{\tilde{v}}^{0} zf(z)dz$ for any cut-off type; here, each type, marginal or infra-marginal, is paid exactly the smallest amount required to induce them to produce. The space between the two curves gives the "information costs" which arise from the informational asymmetry.

The expected benefits of coordination (over and above the autonomous outcome) are given by the left hand sides of the inequalities in (3.10) and (3.11), $w^*(F(0) - F(\tilde{v}))$.

This is illustrated in Figure 3, again for the case of positive externalities. Both curves here are defined for a given positive externality, w^*. The upper curve gives the

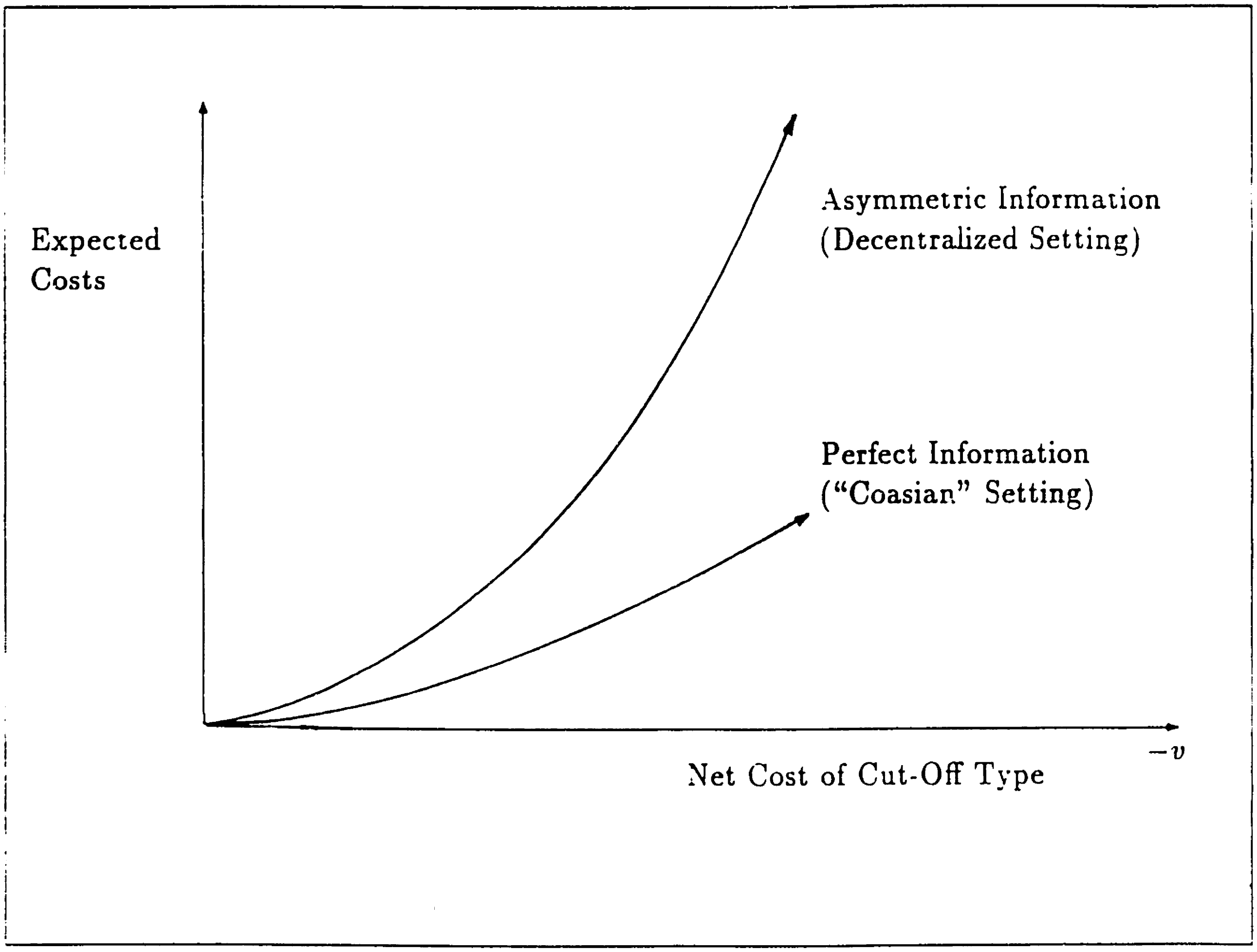

Figure 3-2: Expected Costs of Coordination with Positive Externalities, $v \sim U[-k, k]$

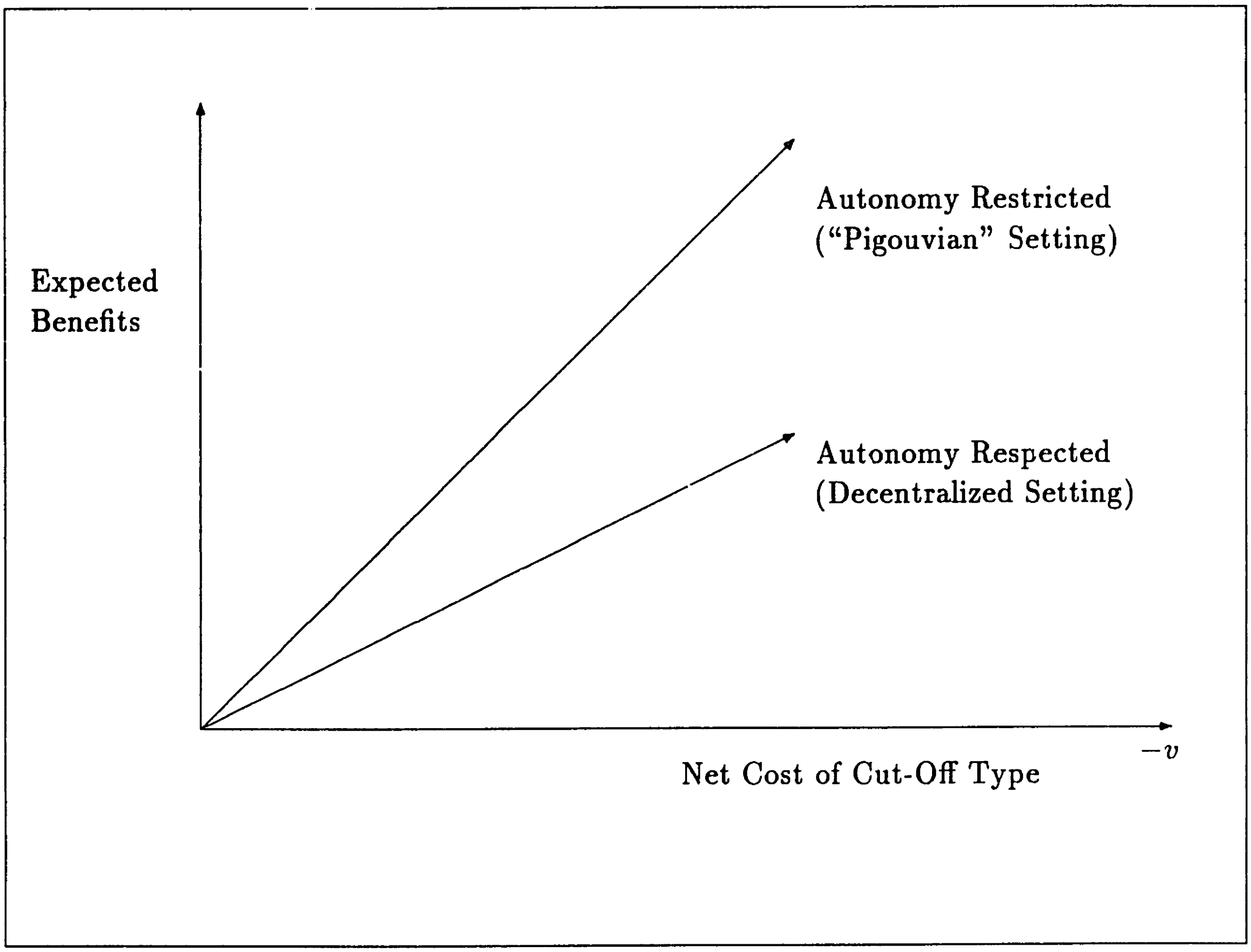

Figure 3-3: Expected Benefits of Coordination with Positive Externalities, $v \sim U[-k, k]$

expected benefits when firms are prohibited from producing outside of a coordinated agreement (the "Pigouvian" world), and the lower curve reflects expected benefits in the decentralized setting of our model. The difference in the curves is exactly $w^*(1 - F(0))$ in that the lower one accounts for the probability that a firm may make the desired production choice in the absence of an agreement, while in the higher one this probability is zero.

3.3.2 Characterization of Optimal Transfers

Once the cut-off type, $\tilde{v}$, has been determined, implementing the optimal program here is simple: we require firm 1 to pay $\tilde{v}$ to firm 2 if firm 1 produces. If $\tilde{v} > 0$, this is a per unit tax on production, and if $\tilde{v} < 0$, this is a per unit subsidy, or matching grant, on production. We can calculate $\tilde{v}$, and thus the size of an optimal tax/subsidy, by assuming the relevant inequality ((3.10) or (3.11)) holds with equality. Thus, for example, for a positive externality $\tilde{v}$ solves

$$w^*(F(0) - F(\hat{v})) + \hat{v}(1 - F(\hat{v})) = 0. \tag{3.12}$$

Observe that this equation will, in general, have multiple solutions (for instance $\hat{v} = 0$ (i.e. the autonomous outcome) is always a solution). The optimal $\hat{v}$, i.e. $\tilde{v}$, will be the minimum of these solutions (i.e. the smallest tax or largest subsidy), since this will encourage the most production and will still be acceptable to firm 2.

It is interesting to compare these transfers with standard Pigouvian taxes. In our framework, Pigouvian taxes correspond to the case where $-\hat{v} = w^*$. The needed assumption here is that, in the case of positive externalities, firm 2 is willing to pay in full for the external benefits, and thus firm 1 is subsidized in exactly the amount of the externality. In this case the socially optimal level of production is achieved.

In a truly decentralized setting, however, firms do not face involuntary restrictions on their actions. So, firm 2's expected benefit from participating in the scheme is reduced to the extent that these benefits would be forthcoming under autonomy as well. Firm 2 will thus not be willing to pay transfers as large as w^*, as required in the

Pigouvian case, and a form of second-best "Pigouvian" taxes/subsidies instead involve transfers equal to $\tilde{v}$. These second-best taxes/subsidies can be called "Pigouvian" in the sense that they are paid uniformly to all firms that produce, irrespective of actual benefits and costs. The fact that these transfers are strictly smaller than w^* is proved in Proposition 1.

Proposition 1 *If $\underline{v} < 0 < \overline{v}$ and $w^* \neq 0$, transfers will be lower than the "Pigouvian" level (i.e., $|\tilde{v}| < |w^*|$.)*

Proof of Proposition 1

See Appendix B.2.

3.3.3 Obtaining First-Best Outcomes

When can the first-best, ex post efficient outcome be reached? Consider the case of positive externalities ($w^* > 0$). Efficiency requires that firm 1 undertake the project if $v \geq -w^*$. In light of Proposition 1, therefore, we know that efficiency will not be possible if $w^* < -\underline{v}$, since in this case firm 1 produces if and only if $v \geq \tilde{v}$ which is gieater than $-w^*$. This yields too little production. Thus if we are to achieve efficiency at all, it can only occur when the external effect is greater than the largest possible cost ($w^* > -\underline{v}$) so that in the first-best *all* possible types of firm 1 are required to produce. Equation (3.10) tells us when the expected cost of compensating all possible types of firm 1 to produce (by paying a transfer equal to the greatest possible cost of producing, $-\underline{v}$) is less than the expected benefits:

$$\underline{v}(1 - F(\underline{v})) + w^*(F(0) - F(\underline{v})) \;=\; \underline{v} + w^*F(0) \geq 0.$$

That is, the net benefit of guaranteeing that all types produce, $w^*F(0)$, must be greater than the expected cost incurred by paying $-\underline{v}$ to firm 1 with probability equal to 1.[18]

So, when externalities are positive, coordination can lead to all localities producing only when

$$w^* \geq \frac{-\underline{v}}{F(0)} \tag{3.13}$$

To obtain a sense of relative magnitudes, assume that private net benefits are distributed uniformly on the interval $[-k, k]$. Then, the condition implies that

$$w^* \geq 2k$$

is necessary to obtain the first-best outcome.[19] That is, coordination will only achieve efficiency if external effects are at least *twice as large* as the *largest possible* private net benefit.

While the result suggests that external effects must be large relative to private net benefits for the first-best to be achieved, there may be common situations in which "large enough" externalities exist. For example, if a public service is not very "local", such as a waste disposal site which can serve many localities in a region, then the externalities, taken together, are likely to be very large relative to private net benefits. Similarly, even with two firms or localities, if production costs are a large fraction of benefits, then the externality could be large compared to private *net* benefits.

3.3.4 Improvements on the Autonomous Allocation

How large must the externality be in order to obtain an improvement over the autonomous allocation? Again, take the case of positive externalities ($w^* > 0$). Improving on the autonomous outcome is only possible when the externality, w^*, is large enough so that the weight on the marginal type induced to produce, $f(0)$, multiplied by the gain from production, w^*, is larger than the weight on transfer payments to all types at least as large, $(1 - F(0))$. These latter types would have produced any-

[18] When implementing efficiency it is sufficient to have transfers of size $-\underline{v} < -\tilde{v}$ instead of requiring larger transfers equal to $-\tilde{v}$ as in section 3.3.2.

[19] Appendix B.3 formally derives this result and the symmetric result for negative externalities.

way without compensation and thus enter only in the cost calculation and not in the benefits.[20]

Thus, when externalities are positive, the external effect must be at least as big as

$$w^* \geq \frac{1 - F(0)}{f(0)} \tag{3.14}$$

to improve on the autonomous allocat:on.[21] Again, to obtain a sense of relative magnitudes, consider the case in which private net benefits are distributed uniformly on the interval $[-k, k]$. Then, equation (3.14) implies that

$$w^* \geq k$$

must hold for any improvements to be implementable. That is, coordination will be worthwhile only if external effects are *at least as large* as the *largest possible* private net benefit.

Figure 4 illustrates the three regimes (no improvement, some improvement, and efficiency) in terms of the expected costs shown in Figure 2 and the expected benefits shown in Figure 3. In order to achieve a gain over autonomy, the expected benefit curve must be above the expected cost curve at some $-\tilde{v} > 0$. In other words, there must be some beneficial tax/subsidy which is acceptable to firm 2. For this to occur, the slope of the expected benefit curve must exceed the slope of the expected cost curve at the origin. This is precisely what equation (3.14) captures. The three net benefit curves in Figure 4 reflect different-sized externalities corresponding to each of the three regimes.

The lowest expected benefit curve (for w_1^*) is always below the expected cost curve,

[20] The sharpness of this result arises from considering whether or not firms undertake investments of a fixed size $\{0,1\}$, as in Myerson-Satterthwaite [1983], Cramton-Gibbons-Klemperer [1987], and much of the bargaining literature. The assumption is equivalent to assuming constant marginal net benefits of production up to a finite limit. If, instead, firms make continuous, unbounded choices about levels of production, increasing subsidies induces a firm to raise levels of production and this has social benefits. Here, however, increasing subsidies to a firm which would have made the investment anyway does not affect their actions.

[21] Appendix B.4 formally derives the result, as well as the symmetric result for negative externalities.

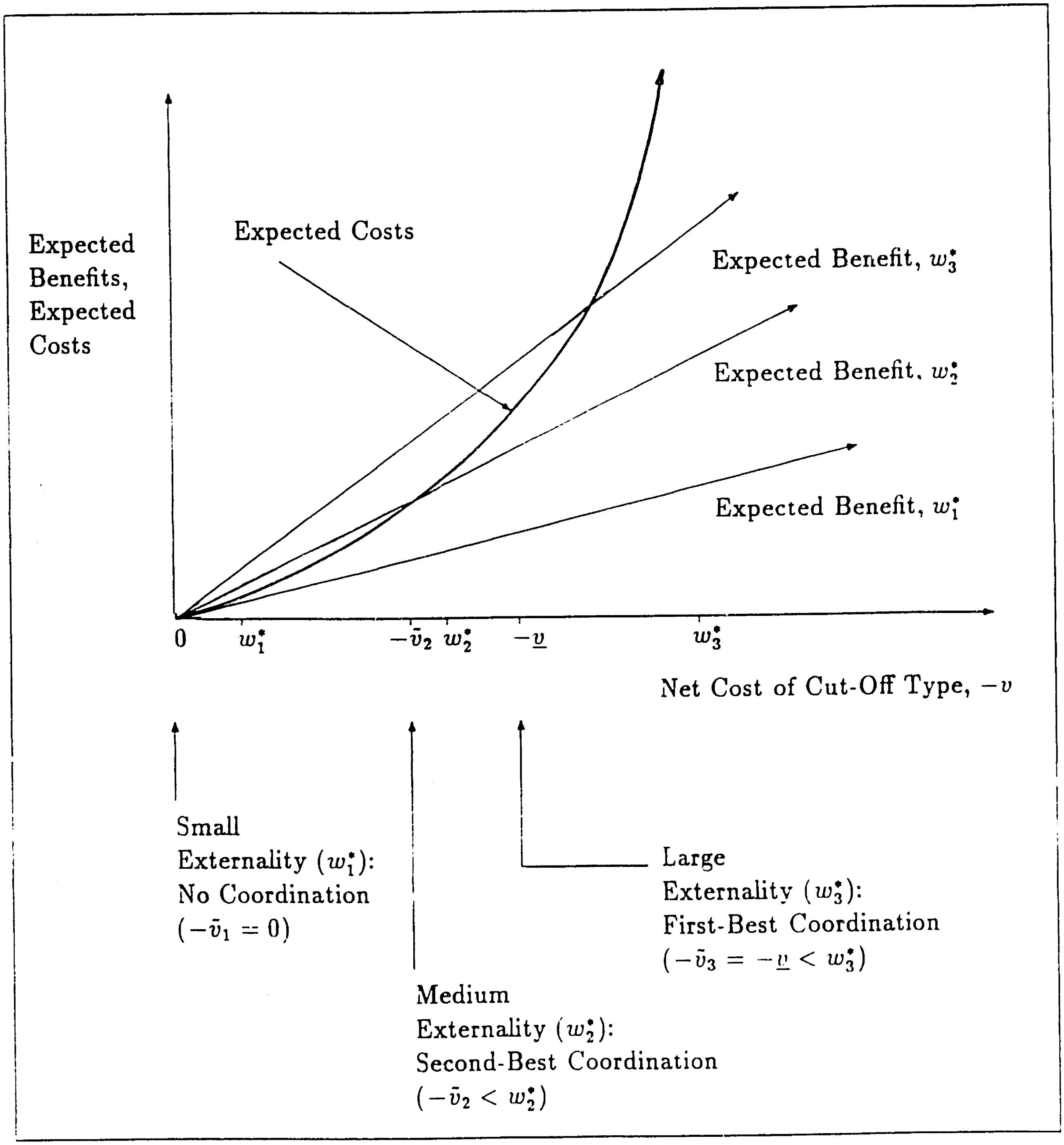

Figure 3-4: Expected Costs and Benefits of Coordination under Decentralization: Increasing Positive Externalities ($w_1^* < w_2^* < w_3^*$)

so that no agreement satisfying the constraints will improve on the autonomous allocation. The middle expected benefit curve (for w_2^*) lies partially above the expected cost curve but intersects it at a point below that where first-best efficiency (i.e., $-\tilde{v}_2 = w^*$) is obtained. The intersection point identifies the optimal cut-off type (and thus the optimal subsidy), since there are always gains from increasing the $-\tilde{v}$ as long as $-\tilde{v} \leq w^*$, and only cut-off types for which benefits exceed costs satisfy the constraint. The highest expected benefit curve (for w_3^*) reflects a level of externalities sufficient to achieve first-best efficiency. Here, although the point of intersection is below w_3^*, it is greater than $--\underline{v}$ and thus efficiency is obtained in that all types are induced to produce.

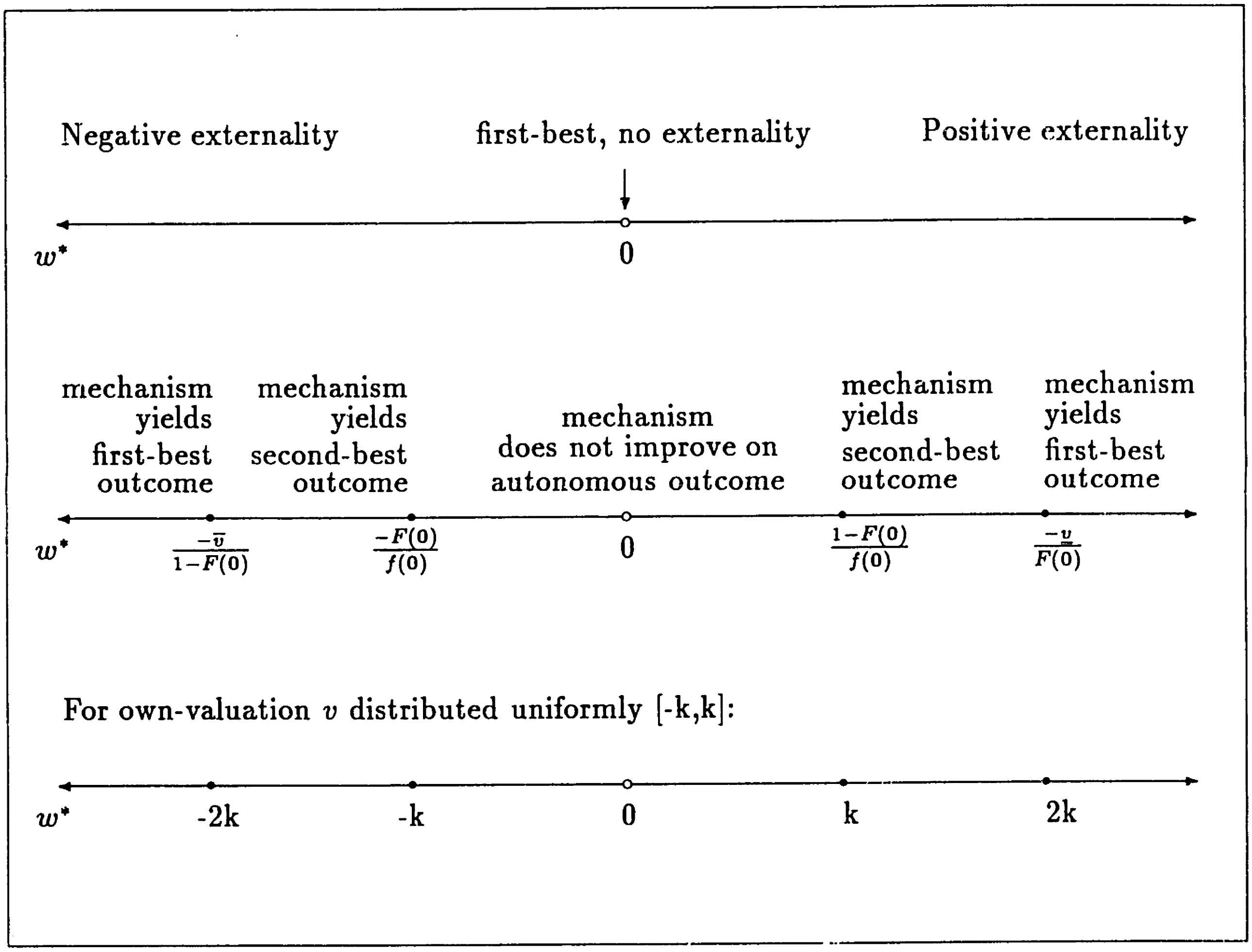
Negative externality
first-best, no externality
Positive externality
w^*
0
mechanism yields first-best outcome
mechanism yields second-best outcome
mechanism does not improve on autonomous outcome
mechanism yields second-best outcome
mechanism yields first-best outcome
w^*
$\frac{-\bar{v}}{1-F(0)}$
$\frac{-F(0)}{f(0)}$
0
$\frac{1-F(0)}{f(0)}$
$\frac{-\underline{v}}{F(0)}$
For own-valuation v distributed uniformly [-k,k]:
w^*
-2k
-k
0
k
2k

Figure 3-5: Summary of Results — Externalities (w^*) and Outcomes Implementable through Voluntary Agreement

3.3.5 Summary of Results

We have argued that larger externalities allow increased efficiency and that for small externalities no gains from an agreement are possible. Figure 5 shows this result for both positive and negative externalities when private net benefits, v, are distributed uniformly on the interval [-k,k].[22] When w^* is between -k and k, coordination will not improve on the autonomous outcome. Only if w^* is less than -2k or greater than 2k will the first-best outcome be attainable.

[22] It can be shown that all of the intervals in Figure 5 are well-defined.

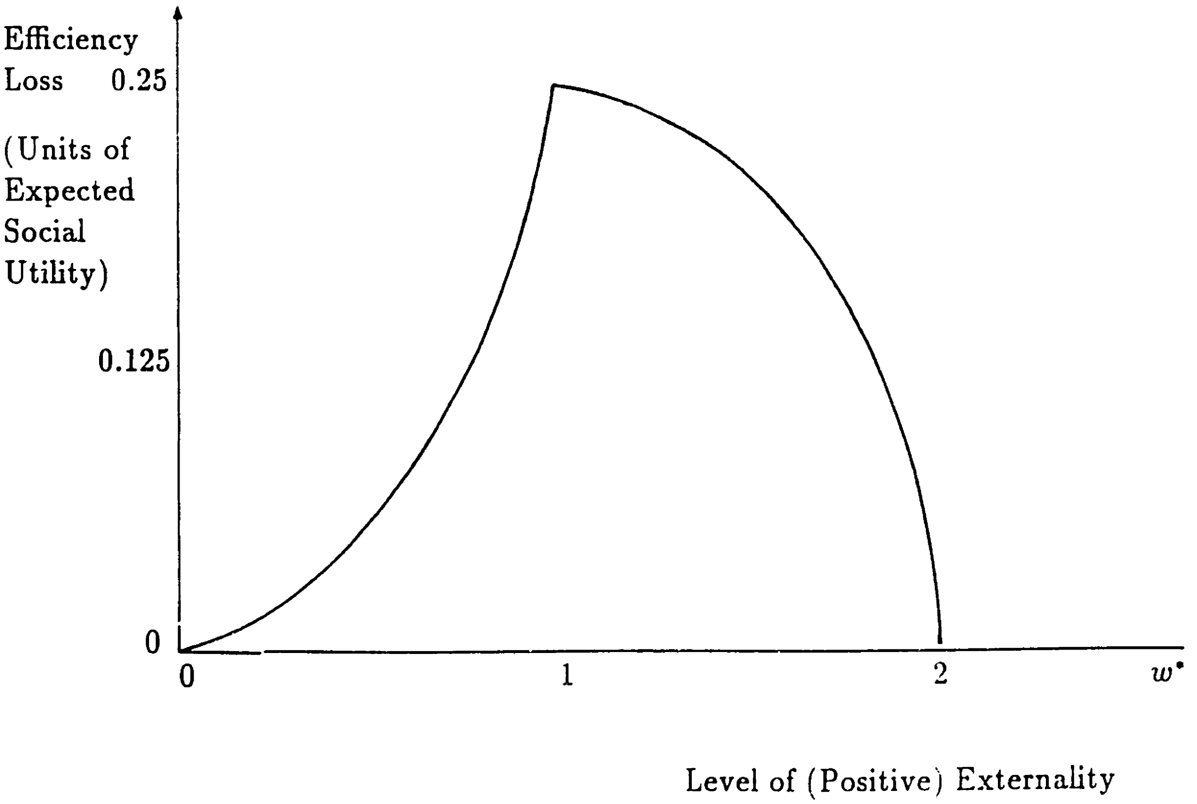

Figure 3-6: Deviations from Efficiency with Coordination, Private Net Benefit $v \sim U[-1,1]$

Figure 6 shows efficiency losses (relative to the first-best) associated with positive externalities in the case in which private net benefits, v, are distributed uniformly on the interval [-1,1]. Note here that when w^* is between 0 and 1, coordination does not improve on autonomy, so initially larger externalities are associated with larger inefficiencies. But, beyond this range, coordination does improve on the autonomous allocation. When w^* is between 1 and 2, coordination serves to reduce inefficiencies so that beyond $w^* = 2$, the first-best outcome can be achieved and no efficiency is lost. Thus, only when external effects are relatively small does inefficiency rise with externalities. At its height ($w^* = 1$), the efficiency loss equals one quarter of social welfare under the first-best allocation. As the size of externalities increases, inefficiency falls until is eventually eliminated.

If the private net benefit were instead distributed with an unbounded distribution, a similar graph would emerge. However, efficiency would only be reached at the limit as the size of the externality approaches ∞. In the case in which $v \sim N(0,1)$, the peak efficiency loss would correspond to an externality of size $\frac{1}{2}\sqrt{2\pi} \approx 1.25$.

3.4 Interpretations and Extensions

3.4.1 Public Expenditures and Public Goods

There is nothing about our model which is specific to firms. As in our introductory example, one of the parties might be a government, or, indeed, both parties might be. For example, consider two localities, one of which can invest in improvements in its public education system, generating a positive externality. We model each locality as maximizing the welfare of a representative resident. Our results then characterize optimal coordination between benevolent governments.

Alternatively, our model can be used to analyze certain public goods problems where unanimous approval is required to implement a mechanism which would determine a provision and funding scheme. A simple example is as follows. Suppose there are two consumers who may differ in their valuation of the public good. Normalize production costs to zero, and suppose that the valuations are independently distributed on $[\underline{v}, \overline{v}]$ according to $F(\cdot)$, where $\underline{v} < 0 < \overline{v}$. Unanimous approval is required for the public good to be provided. In this situation, if one consumer has a positive valuation v, they are only willing to subsidize the other consumer for voting "yes" if v is large enough. This is precisely because the other consumer might vote "yes" even without a subsidy. Thus the size of the valuations here play the role of the size of externalities in our model.

3.4.2 Information Structure

We have assumed that the net benefits of investing in projects are only known privately. The critical aspect of this assumption is that other firms or localities and any higher level of government or mediator are uncertain about whether investment will take place under autonomy (i.e., if no agreement is reached).[23] To see this, observe that if externalities are positive and it is publicly known whether v is larger or smaller than zero, a policy of offering a transfer of w^* to any locality which has $v < 0$, in return for an agreement to produce, will yield the first-best outcome. This policy satisfies the incentive, individual rationality and budget constraints.

Similarly, if v is publicly known, but the size of the externality w^* is private information of the firm or locality affected, then the first-best outcome can always be obtained. For example, if an external benefit $w^* = \frac{1}{2}$ is received by firm 2, it will be willing to offer firm 1 exactly the smallest subsidy required to guarantee production, as long as firm 1's net costs are no greater than the external benefit, $-v \leq \frac{1}{2}$. This means that all types such that $v + w^* \geq 0$ will produce. Thus, we see that asymmetric information concerning the externalities is not sufficient to create inefficiencies in our problem.

If asymmetric information about both direct and indirect effects (i.e., both v and w^*) were considered, the formal analysis becomes more difficult and lies beyond the scope of this paper. However, we conjecture that the presence of this extra asymmetry,

[23] The side-contracts we discussed earlier were precisely attempts to remove the uncertainty about what would happen under autonomy by committing to a particular action.

beyond the one necessary for our results, can only make it more difficult to improve on the decentralized outcome.

3.4.3 Budget Balance

A possible argument against imposing budget balance in the model (interpreted as one of government coordination rather than bargaining) might be that "actual governments do not balance their budgets." There are several responses to this. First, at the level of state and local governments, budget balance is often mandated through legislation or constitutional provision. Second, in the long run, all governments must balance their budgets; i.e., current deficits necessitate future surpluses. Since our model is purely static, this sort of intertemporal shifting of resources cannot occur. Presumably, in a more complex model, there would be a trade-off between running a deficit today and alleviating a current incentive problem versus running a surplus tomorrow and exacerbating (or mitigating to a smaller degree) another incentive problem then. However, the fundamental point remains unchanged: a lack of outside subsidies limits the ability to reach socially desirable outcomes.

A related criticism of the budget balance assumption is that, although governments may balance budgets overall, they often have discretionary revenues which can be shifted among different items in the budget. Thus, in a more complex model where the government is concerned with many projects beyond the one at hand, it might subsidize one project by taking funds from another, rendering the assumption of budget balance too stringent for our framework.

We show in Appendix B.5, however, that the qualitative conclusions of our model are robust to this type of story as long as diverting funds to the project has some positive social opportunity cost. Such a social opportunity cost may arise naturally from distortions introduced by taxation, for example.

3.4.4 Multiple Projects

Rather than just considering a single producer, we can consider cases in which both firms have an investment opportunity which generates an externality on the other. If the effect of each investment is independent, and the private information about the net benefits of each investment is independent, the problems are completely separable and our analysis holds for each individually.[24] However, if there are complementarities or substitution effects across projects, the analysis becomes more complex. One example where such effects might be present is the case of two neighboring states, each considering whether or not to build a road in the direction of the other. A state will benefit more from the other state's road if it has built its own connecting road. Thus the size of the externality is affected by the state's own production decision. This complementarity lends an aspect of coordination to the problem that is absent in our setting. Now, the distortion that externalities generate in a state's investment decision will vary with the actions of the other state. Therefore the actual autonomous outcome will depend very much on the beliefs that the states hold about each other. Thus, the relevant individual rationality constraints will also depend on these beliefs, and the problem becomes difficult to set up since beliefs can change with actions taken within the mechanism.

3.4.5 Multiple Spillovers

In some situations more than one firm is affected by the spillovers; several difficulties may then need to be considered. For example, the assumption that all firms participate or not is a much stronger one when there are more than two firms. If we allow for participation by subsets of the firms while others opt out, the effect will

[24] Recent work by McAfee and Reny [1992] and Crémer and McLean [1985, 1988], among others, has shown that when private information is correlated in mechanism design problems, typically the first-best can be achieved, even with ϵ correlations. These results are striking, but the types of mechanisms which they require are unrealistic, necessitating very large bets. Auriol and Laffont [1991] have shown that the results of the independent case go through while allowing for some correlation if preferences are additive in a common component and an idiosyncratic component and both components are known by the firm. It is then only the idiosyncratic component which matters in the mechanism design problem.

be to increase the welfare of firms when they refuse an agreement. This will make coordination and efficiency harder to achieve than in the model in Section 3.2. This is closely related to the way that we think about free-rider problems in that assuming that the other firms will form an agreement if one firm opts out is like assuming that it is possible for a firm to free-ride.

One issue that arises in a multiple firm setting which can be easily dealt with in our framework is the interpretation of externalities. There are two polar cases to consider. First, externalities can, themselves, have the quality of a public good in that the total external effect increases proportionally with the number of firms affected. Second, gross external effects may be fixed. Then considering more firms reduces the per firm externality.

In the model above, we have defined the external effect w^* in per firm terms. Because the external effect is non-rival in the first case, adding firms does not affect externalities elsewhere, so, here, the government would want to consider the sum of external effects across the n firms. If, for example, all firms are equally affected and equal-sized, a per firm externality equal to $\frac{1}{n-1}\overline{w^*}$ is required to obtain efficient outcomes, where $\overline{w^*}$ is the level required for efficiency in the case with two firms. Thus, when the external effect has the non-rival attributes of a pure public good, the more firms that are affected, the more likely it will be that an efficient outcome can be obtained.

In the second case, in which adding firms reduces the per firm externality proportionally, the basic results carry over unchanged from Section 3.3. Doubling the number of firms affected reduces per firm external effects by half. Thus, the sum of external effects is invariant to the number of firms involved. Here adding firms does not change the likelihood of reaching the first-best outcome.

3.5 Applications

3.5.1 Research and Development

The creation of new products often arises through cross-fertilization of different endeavors. Firms often have a range of research and development projects underway, each overlapping and depending in some way on the other, often through the accumulation of skills or new insights. Without coordination, however, firms will under-invest in projects with positive spillovers. Our framework extends naturally to help explain the optimal behavior of firms in this situation.

In particular, the sort of coordination we describe with respect to local governments has a natural analogue in research joint ventures like Sematech, where microchip manufacturers joined forces to create a new generation of semi-conductors. The problem of designing the initial agreement involves the sort of individual rationality and budget constraints considered here, although budget balance is often violated due to heavy government subsidization. Our analysis in Appendix B.5, where budget balance is not required, suggests that externalities must be relatively large to make joint ventures efficient in the presence of private information (about, say, the net costs of R & D).

One aspect of the joint venture problem which is not present in our model is that the externalities are often partly endogenous as a result of output or profit-sharing agreements. A fully worked-out application would need to incorporate this feature.

3.5.2 Models of Long-Run Growth

The renewed interest in models of long-run economic growth has been been spurred by the explanatory power of new models which feature positive spillovers in production (e.g., Lucas [1988]). The spillovers provide a justification for the assumption that firms face decreasing returns individually while there are constant or increasing returns to aggregate production. Thus, equilibria are competitive, keeping the models simple, but, unlike the standard neo-classical model, growth rates need not converge across economies and capital will not necessarily flow from rich to poor economies.

The fundamental assumption of these models is that there is no coordination. If it were possible to fully internalize externalities, individual firms would face increasing returns, and the non-convexity in the production function would diminish the likelihood of reaching a competitive equilibrium (Laffont [1988]).[25]

The lack of coordination in these models is invoked, rather than explained. If there are an infinite number of producers and no central authority, then presumably coordination would be difficult indeed. Still, even with many producers, there is no reason that governments cannot create tax and subsidy schemes to address externalities; see, for example, Barro [1991].

The present paper suggests that if there is asymmetric information about the direct costs and benefits of production, then even a central authority may not be able to fully internalize externalities through voluntary programs. While our model is static, the intuition carries over to a dynamic framework. However, we have shown that if the gains from coordination, reflected by w^*, are large enough, then we would expect efficient coordination, counter to the assumptions of the new growth literature. In considering deviations from efficient growth paths, we expect that those gains *would* be large, since the benefits to coordination accrue for all time.

3.5.3 Siting a Toxic Waste Dump

Where should toxic waste dumps be situated? Can they be situated efficiently? While localities understand the need for waste dumps, no one wants one in their own "back yard". Clearly, if no one had to have a dump, under autonomy no one would. But, given that a site must be chosen, the problem involves determining which locality can bear the burden at least cost. This choice framework can be captured by adding to our model the constraint that the sum of probabilities of building a dump must equal one: the dump must go somewhere, but only one is needed.

Consider the case in which the gross external effect is constant no matter where

[25] Strictly speaking, a competitive equilibrium could be maintained as long as coordination is not so effective that it introduces non-convexities into firms' problems.

the dump is sited. If no single locality accepts the communal dump, then all localities build private dumps which yield a level of utility equal to zero.[26] A simple auction can be created (e.g., sealed bid, second price) to allocate the dump, such that the n localities bid to receive a given transfer conditional on building the dump.

If the transfer is at least as large as $-\underline{v}$, the largest possible cost of building the dump, localities will always participate in the auction. Thus the central government taxes each locality $-\underline{v}/(n-1)$; localities will voluntarily pay these taxes as long as $w^* \geq -\underline{v}/(n-1)$. Thus, the auction is consistent with individual rationality, and it does not require running a budget deficit — indeed, the program runs at a surplus, with the central government keeping the money paid by the highest bidder. But, while the auction will lead to the efficient location of the dump, it will not lead to the first-best social outcome. This is because the program runs at a surplus, and there is no return to money in the hands of the government in the present model.

The proportional welfare loss falls as the external effect w^* increases beyond $-\underline{v}/(n-1)$, since the gains from having the dump increase with w^* while the surplus from the auction stays constant. Although this is not a formal analysis, it gives some intuition as to how large externalities can help improve this type of allocation problem.

3.6 Conclusion

The fact that information is known only locally or by individual firms provides a strong reason for favoring decentralized arrangements. Decentralization is also appealing for both political and philosophical reasons, in that it limits the coercive powers of central authorities. However, we have shown that autonomy and private information together can make it very difficult to internalize externalities. This can lead to substantial losses in social efficiency.

Respect for autonomy (the essence of decentralization in this paper) is critical for

[26] More realistically, the level of utility flowing from a private dump would equal v. We do not address this scenario here.

this result, since even with private information and externalities, a central government with coercive powers can implement efficient outcomes. In this case, a system of Pigouvian taxes and subsidies can be used to achieve efficiency.[27]

When autonomy is respected, and when firms or localities decide whether to make investments of a given size, voluntary agreements may not improve the outcome in a large range of cases. This occurs when the size of the external effect is relatively small compared with net benefits to producers (where "relatively small" may nevertheless be large in an absolute sense.) This may help to explain why coordination is so rarely observed relative to the number of activities associated with externalities.

While these principles are derived in a fairly general framework, the optimal plan is easily implementable. When improvements are possible, they can be achieved through simple unit taxes and subsidies. The results suggest that asymmetric information can lead to large costs in terms of efficiency. However, we have also shown that when externalities are relatively large, the first-best outcome can always be obtained.

[27] That is, a system of taxes/subsidies based on w^* and contingent on production.

Bibliography

[1] Auriol, E. and J.-J. Laffont [1991], "Regulation by Duopoly", Mimeo, IDEI, Toulouse.

[2] Bagnoli, M. and T. Bergstrom [1989], "Log-Concave Utility and its Applications", Discussion Paper 89-23, University of Michigan.

[3] Barro, Robert [1991], "Government Spending in a Simple Model of Long-Run Growth", *Journal of Political Economy* 98(5), Part 2:S103-S125.

[4] Coase, Ronald H. [1960], "The Problem of Social Cost", *Journal of Law and Economics* 3: 1 – 44.

[5] Cramton, Peter, Robert Gibbons, and Paul Klemperer [1987], "Dissolving a Partnership Efficiently", *Econometrica* 55(3), May: 615 – 632.

[6] Crémer, Jacques and R. McLean [1985], "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent", *Econometrica* 55: 345 – 361.

[7] Crémer, Jacques and R. McLean [1988], "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions", *Econometrica* 56: 1247 – 1258.

[8] Farrell, Joseph [1987], "Information and the Coase Theorem", *Journal of Economic Perspectives* 1(2), Fall: 113 – 129.

[9] Fudenberg, Drew and Jean Tirole [1991], *Game Theory*, Cambridge, MA: MIT Press.

[10] Greenwood, Jeremy and R. Preston McAfee [1991], "Externalities and Asymmetric Information", *Quarterly Journal of Economics*: 103 - 121.

[11] Laffont, Jean-Jacques [1988], *Fundamentals of Public Economics*, Cambridge, MA: MIT Press.

[12] Laffont, Jean-Jacques and Eric Maskin [1979], "A Differential Approach to Expected Utility Maximizing Mechanisms" in *Aggregation and the Revelation of Preferences* (ed., J.-J. Laffont), Amsterdam: North-Holland: 289 – 308.

[13] Laffont, Jean-Jacques and Jean Tirole [1993], *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, forthcoming.

[14] Lewis, Tracy and David Sappington [1989], "Countervailing Incentives in Agency Problems", *Journal of Economic Theory* 49: 294 – 313.

[15] Lucas, Robert E., Jr. [1988], "On the Mechanics of Economic Development", *Journal of Monetary Economics.*

[16] Madison, James, Alexander Hamilton and John Jay [1787], *The Federalist Papers.* Reprinted 1987, Isaac Kramnick, ed., London: Penguin Books.

[17] Maggi, Giovanni and Andres Rodriguez [1993], "On Countervailing Incentives", Mimeo, Stanford University.

[18] Myerson, Roger B. and Mark A. Satterthwaite [1983], "Efficient Mechanisms for Bilateral Trading", *Journal of Economic Theory* 29: 265 – 281.

[19] McAfee, R. Preston and Philip Reny [1992], "Correlated Information and Mechanism Design", *Econometrica*, March: 395 – 421.

[20] Pratt, John W. and Richard Zeckhauser [1987], "Incentive-Based Decentralization: Expected-Externality Payments Induce Efficient Behaviour in Groups", in George R. Feiwel, ed., *Arrow and the Ascent of Modern Economic Theory*, New York: New York University Press.

[21] Rubinfeld, Daniel [1987], "The Economics of the Local Public Sector", in Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, volume 2. Amsterdam: North-Holland.

[22] Weingast, Barry [1993], "The Economic Role of Political Institutions", IPR Working Paper 46, Institute for Policy Reform, Washington, D.C.

Appendix A

Proof of Theorems 5 and 6 of Chapter 1

Theorem 6 will be proved first, as it will be used to prove Theorem 5.

Theorem 6 *Let $\succeq$ be a binary relation on L_0. Then the following are equivalent,*

(1) $\succeq$ satisfies B.1 and B.2 for $L = L_0$.

(2) $\exists$ an affine function $u : Y \rightarrow \mathcal{R}$ and $N \geq 1$ non-empty, closed, convex sets $C_i, i = 1, \ldots, N$, of finitely additive probability measures on Σ such that $\forall f, g \in L_0, f \succeq g$ if and only if $(\min_{p \in C_i} \int u \circ f dp)_{i=1}^N \geq_L (\min_{p \in C_i} \int u \circ g dp)_{i=1}^N$, where if $p(E) > 0$ for some $E \in \Sigma$, $p \in C_1$ then there exists an i such that $p(E) > 0$, for all $p \in C_i$, and where $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_N$.[1]

Furthermore, the function u is unique up to a positive affine transformation, and, if and only if A.6 holds, the set C_1 is unique.

Proof of Theorem 6:

We will first prove that (1) implies (2), then that the uniqueness properties of the representation in (2) are satisfied, and finally that (2) implies (1). The proof of (1) implies (2) is the most involved. We will use theorem 1 applied to each $\succeq_i$ and B.2 to

[1] For $a, b \in \mathcal{R}^N$, $a \geq_L b \Leftrightarrow [b_i > a_i \Rightarrow \exists k < i$ such that $a_k > b_k]$. This is the reflexive relation induced by lexicographic ordering.

derive the basic form of the representation. Then, to show that the superset relations between the sets of beliefs hold, we will appeal to a construction of suitable sets C_i in Chateauneuf (1991). Finally we use a lemma and B.1 to show that the measures in the C_i must satisfy the conditions stated in (2).

(1) $\Rightarrow$ (2): From B.2 we know that the representation is lexicographic in the $\succeq_i$. Applying theorem 1 to each $\succeq_i$ we have that $f \succeq_i g$ if and only if $\min_{p \in C_i} \int u_i \circ f dp \geq \min_{p \in C_i} \int u_i \circ g dp$, for a non-empty, closed, convex set C_i and an affine $u_i : Y \to \mathcal{R}$ which is unique up to a positive affine transformation. As B.2 requires all $\succeq_i$ to agree on constant acts, we can take $u_i = u_1, i = 1, \ldots, N$. Thus we have the basic representation.

Now we prove the superset condition holds. Consider the space B of all bounded, Σ-measurable real functions on S. By Lemmas 3.1-3.4 of Gilboa and Schmeidler (1989), there exists, for each i, $I_i : B \to \mathcal{R}$ such that $I_i(u \circ y^*) = u(y)$ for $y^* \in L_c$ with outcome $y \in Y$; $f \succeq_i g$ if and only if $I_i(u \circ f) \geq I_i(u \circ g)$ for $f, g \in L_0$; I_i monotonic, superadditive, homogeneous of degree 1, and C-independent. Thus I_i satisfies the conditions of the Fundamental lemma[2] in Chateauneuf (1991), and thus, by his constructive proof, we can take C_i to be the set $\{p | p$ is an additive probability measure on Σ; $\int b dp \geq I_i(b), \forall b \in B$ such that $I_i(b) > 0\}$. As B.2 requires $f \succeq_{i+1} y^*$ if $f \sim_i y^*$, $I_k(u \circ f) \geq I_i(u \circ f)$ if $k > i$. Thus for all $b \in B$, $I_k(b) \geq I_i(b)$ if $k > i$. From the definition of C_i, we see that this implies $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_N$. To complete the proof of (1) $\Rightarrow$ (2) we make use of the following result:

[2] This lemma says that for $I : V \to \mathcal{R}$, where V is the set of all Σ-measurable functions from S to the positive reals, the following two conditions are equivalent:

Condition 1. I satisfies:

(i) for all $\alpha \geq 0, \beta \geq 0, x \in V : I(\alpha x + \beta 1^*) = \alpha I(x) + \beta$, where 1^* is a function which takes on the value 1 in all states.

(ii) $x, y \in V \Rightarrow I(x + y) \geq I(x) + I(y)$.

(iii) If $x \geq y$ on S, then $I(x) \geq I(y)$.

Condition 2.

There exists a unique closed, convex set C of additive probabilities on Σ, such that

(iv) $I(x) = \min_{p \in C} \int x dp$, for all $x \in V$.

To apply this lemma to I_i we can simply rescale u so that u takes on only positive values and consider the restriction of I_i to V. Monotonicity, superadditivity, homogeneity, and C-independence ensure Condition 1 is satisfied.

Lemma 6.1

An event $E \in NNE \Leftrightarrow p(E) > 0$ *for some* $p \in C_1$.

Proof of Lemma 6.1: $(\Rightarrow) : p(E) = 0, \forall p \in C_1$ implies $p(E) = 0, \forall p \in C_i$, which implies E null.

$(\Leftarrow)$: Consider $f, g \in L_0$ such that $u(f(s)) = u(g(s)) = k$ on S/E and $k > u(f(s)) > u(g(s))$ on E. $\min_{p \in C_i} \int u \circ f dp \neq \min_{p \in C_i} \int u \circ g dp$ if and only if $p(E) > 0$ for some $p \in C_i$. Thus $p(E) > 0$ for some $p \in C_1$ implies E not null. *QED*

For any $E \in NNE$, Lemma 6.1 tells us that $p(E) > 0$ for some $p \in C_1$. For any such E, consider f, g such that $u(f(s)) = u(g(s)) = k$ on S/E and $u(f(s)) > u(g(s)) > k$ on E. For each C_i, if there exists $p \in C_i$ such that $p(E) = 0$ then $\min_{p \in C_i} \int u \circ f dp = \min_{p \in C_i} \int u \circ g dp$. Since B.1 requires that $f \succ g$, there must be some $i \in \{1, \ldots, N\}$ such that $p(E) > 0$, for all $p \in C_i$.

Uniqueness: that u is unique up to a positive affine transformation follows directly from the vNM representation theorem (von Neumann and Morgenstern, 1947). If A.6 fails then any closed, convex set C_1 will do in combination with a constant u. Suppose A.6 holds. We adapt an argument of Gilboa-Schmeidler (1989) to our setting. Assume there exist $C_1' \neq C_1''$, non-empty, closed, and convex such that $(\min_{p \in C_i'} \int u \circ f dp)_{i=1}^{N}$, for some C_i', $i = 2, \ldots, N$ such that $C_1' \supseteq C_2' \supseteq \ldots \supseteq C_N'$ and $(\min_{p \in C_i''} \int u \circ f dp)_{i=1}^{N}$, for some C_i'', $i = 2, \ldots, N$ such that $C_1'' \supseteq C_2'' \supseteq \ldots \supseteq C_N''$ represent $\succeq$ on L_0 in the manner of the theorem. Without loss of generality, assume there exists $p' \in C_1'/C_1''$. By a separation theorem [Dunford and Schwartz, 1957, V.2.10], there exists $a \in B$ such that $\int a dp' < \min_{p \in C_1''} \int a dp$. Without loss of generality assume $a = u \circ f$ for some $f \in L_0$. Let $y \in Y$ be such that $u(y) = \min_{p \in C_1''} \int u \circ f dp$. Since $C_1'' \supseteq C_2'' \supseteq \ldots \supseteq C_N''$, this implies that $f \succeq y^*$ where y^* is the constant act which results in y. But $u(y) = \min_{p \in C_1''} \int u \circ f dp > \min_{p \in C_1'} \int u \circ f dp$, which implies $y^* \succ f$, a contradiction. Thus C_1 is unique if and only if A.6 holds.

$(2) \Rightarrow (1)$: We define $f \succeq_i g \Leftrightarrow \min_{p \in C_i} \int u \circ f dp \geq \min_{p \in C_i} \int u \circ g dp$. B.2 is then easily verified (recall that $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_N$). The fact that $p(E) > 0$ for some $E \in \Sigma$, $p \in C_1$ implies there is an i such that $p(E) > 0, \forall p \in C_i$, means that all non-null events are given positive weight in some element of $(\min_{p\in C_i}\int u\circ f dp)_{i=1}^N, \forall f \in L_0$. Suppose that $u(f(s)) \geq u(g(s)), \forall s \in S$. Since f and g are Σ-measurable, $u\circ f - u\circ g$ is Σ-measurable and thus $\{s : u(f(s)) - u(g(s)) > 0\} \in \Sigma$. $f \succ g$ if and only if $\{s : u(f(s)) - u(g(s)) > 0\}$ is not null. Therefore B.1 holds. *QED*

Theorem 5 *Let $\succeq$ be a binary relation on L_0. Then the following are equivalent,*

(1) $\succeq$ satisfies A.1 - A.3, A.5 and B.1 for $L = L_0$.

(2) There exists an affine function $u : Y \to \mathcal{R}$ and a non-empty, closed, convex set C of finitely additive probability measures on Σ satisfying $[p(E) = 0$ if and only if $\forall p \in C, p(E) = 0]$ such that $\forall f, g \in L_0, f \succeq g$ if and only if $\min_{p\in C}\int u\circ f dp \geq \min_{p\in C}\int u\circ g dp$.

Furthermore, the function u is unique up to a positive affine transformation and, if and only if A.6 holds, the set C is unique.

Proof of Theorem 5: First note that B.1 implies A.4 (Monotonicity).

(1) $\Rightarrow$ (2): A.1-A.5 imply B.2 with $N = 1$ by Theorem 1. B.1 and B.2 with $N = 1$ imply (2) by Theorem 6.

Uniqueness: Follows by the same arguments (vNM theorem, separation theorem) as uniqueness in Theorems 1 and 6.

(2) $\Rightarrow$ (1): (2) implies $\succeq$ satisfies A.1-A.5 by Theorem 1. (2) implies (by Lemma 6.1) that for all $p \in C$, $[p(E) > 0, \forall$ non-null $E \in \Sigma]$ which implies B.1 (weak admissibility) since $\{s : u(f(s)) - u(g(s)) > 0\}$ is Σ-measurable. *QED*

Appendix B

Appendicies for Chapter 3

B.1 Proof of Theorem 1

We first provide the basic intuition for why we can limit attention to just these two constraints (the first pertaining to the type with the greatest possible local net benefit from producing, $\overline{v}$, the second pertaining to the type with the lowest local net benefit, $\underline{v}$.)

Can it be true that the individual rationality constraint for firm 1 binds for types $\overline{v}$ and $\underline{v}$, but is violated for other types? First, consider a type with negative net benefits, $v < 0$. If the individual rationality constraint was violated, type v could simply pretend to be the type with the greatest net costs, $\underline{v}$, and get $U(\underline{v}) + (v - \underline{v}) > U(\underline{v})$ if $\underline{v}$ was producing and $U(\underline{v}) > \max(v, 0)$ otherwise. But then incentive compatibility would be violated at v, so this could not happen. Similarly, can it be true that individual rationality is violated for a type with positive net benefits from producing, $v > 0$? If this was the case, type v could pretend to be $\overline{v}$ and would get $U(\overline{v}) - (\overline{v} - v) > \max(v, 0)$ if $\overline{v}$ was producing and $U(\overline{v}) > \max(v, 0)$ otherwise. But, again, incentive compatibility would be violated, so this could not happen. Thus, incentive compatibility implies that we need to just consider the types $\overline{v}$ and $\underline{v}$ when checking whether individual rationality is satisfied for firm 1. Since firm 2's individual rationality constraint does not vary with v, the same is trivially true of it.

Constraint $(\overline{A.1})$ says that total expected social surplus, adjusted for information costs which arise from the incentive compatibility constraint, must be as big as the sum of what both localities would expect to get in the absence of coordination if firm 1 was the high type (but firm 2, of course, does not know this valuation). Constraint (A.1) gives the analogous condition for the lowest types.

Now we proceed with the formal proof. We first show the "only if" part of the theorem: Suppose that the direct revelation mechanism $\langle p(\cdot), t(\cdot)\rangle$ satisfies (IC), (IR1), and (IR2). Incentive compatibility (IC) implies

$$\begin{aligned} U_1(v) &= vp(v) + t(v) \\ &\geq vp(\hat{v}) + t(\hat{v}). \end{aligned}$$

Thus, by the envelope theorem, $\frac{dU_1(v)}{dv} = p(v)$ almost everywhere. So,

$$U_1(v) = U_1(v^*) + \int_{v^*}^{v} p(z)dz, \forall v, v^*,$$

which implies that

$$t(v) = t(v^*) + v^*p(v^*) - vp(v) + \int_{v^*}^{v} p(z)dz \;\; \forall v, v^*,$$

Taking expectations over v, we get:

$$\begin{aligned} \int_{\underline{v}}^{\overline{v}} t(z)f(z)dz &= T(v^*) + v^*p(v^*) - \int_{\underline{v}}^{\overline{v}} zp(z)f(z)dz \\ &\quad + \int_{v^*}^{\overline{v}} [1 - F(z)]p(z)dz - \int_{\underline{v}}^{v^*} F(z)p(z)dz. \end{aligned}$$

Now we can rewrite (IR2) as

$$\begin{aligned} U_j &= w^* \int_{\underline{v}}^{\overline{v}} p(z)f(z)dz - \int_{\underline{v}}^{\overline{v}} t(z)f(z)dz && \text{(B.1)} \\ &= w^* \int_{\underline{v}}^{\overline{v}} p(z)f(z)dz - v^*p(v^*) - t(v^*) && \text{(B.2)} \\ &\quad + \int_{\underline{v}}^{\overline{v}} zp(z)f(z)dz - \int_{v^*}^{\overline{v}} [1 - F(z)]p(z)dz \end{aligned}$$

$$\begin{aligned} &+\int_{\underline{v}}^{v^*} F(z)p(z)dz \\ \geq\ & w^*(1-F(0)). \end{aligned}$$

Now, since $\frac{dU_1(v)}{dv} = p(v)$ and $0 \leq p(v) \leq 1$ and $p(v)$ is non-decreasing in v, $U_1(v) - \max(v, 0) - w^*(1 - F(0))$, reaches a global minimum at either $v = \underline{v}$ or $v = \overline{v}$. Thus, if (IR1) is satisfied at these two points, it is satisfied everywhere.

So, recalling the assumption that $\underline{v} < 0 < \overline{v}$, (IR1) implies:

$$\begin{aligned} U_1(\underline{v}) = \underline{v}p(\underline{v}) + t(\underline{v}) &\geq 0 \text{ and} \\ U_1(\overline{v}) = \overline{v}P_i(\overline{v}) + t(\overline{v}) &\geq \overline{v}. \end{aligned}$$

This places restrictions on the size of transfers:

$$t(\underline{v}) \geq -\underline{v}p(\underline{v}) \text{ and} \tag{B.3}$$

$$t(\overline{v}) \geq \overline{v} - \overline{v}p(\overline{v})). \tag{B.4}$$

From (IR2),

$$\begin{aligned} vp(v) + t(v) \leq\ & w^* \int_{\underline{v}}^{\overline{v}} p(z)f(z)dz + \int_{\underline{v}}^{\overline{v}} zp(z)f(z)dz \\ & - \int_{v}^{\overline{v}} [1 - F(z)]p(z)dz + \int_{\underline{v}}^{v} F(z)p(z)dz - w^*(1 - F(0)). \end{aligned}$$

If we bring all of the v terms to the left-hand side we have,

$$\begin{aligned} vp(v) + t(v) + \int_{v}^{\overline{v}} [1 - F(z)]p(z)dz - \int_{\underline{v}}^{v} F(z)p(z)dz & \\ \leq \int_{\underline{v}}^{\overline{v}} (z + w^*)p(z)f(z)dz. & \end{aligned}$$

In particular, for $v = \underline{v}$ and $v = \overline{v}$:

$$t(\underline{v}) \leq -\underline{v}p(v) - w^*(1 - F(0)) + \int_{\underline{v}}^{\overline{v}} \left(z + w^* - \frac{1 - F(z)}{f(z)}\right) p(z)f(z)dz$$

and

$$t(\overline{v}) \;\leq\; -\overline{v}p(\overline{v}) - w^*(1-F(0)) + \int_{\underline{v}}^{\overline{v}} \left(z + w^* + \frac{F(z)}{f(z)}\right) p(z)f(z)dz.$$

For these inequalities to be compatible with the ones in (B.3) and (B.4), we need

$$(\underline{A.1}) \qquad \int_{\underline{v}}^{\overline{v}} p(z) \left(z + w^* - \frac{(1-F(z))}{f(z)}\right) f(z)dz \geq w^*(1-F(0)),$$

and

$$(\overline{A.1}) \qquad \int_{\underline{v}}^{\overline{v}} p(z) \left(z + w^* + \frac{F(z)}{f(z)}\right) f(z)dz \geq \overline{v} + w^*(1-F(0)).$$

Now, we prove the other direction ("if"). We proceed by considering any nondecreasing $p(v)$ and constructing transfers which satisfy (IC), (IR1), and (IR2), using the assumption that $(\overline{A.1})$ and $(\underline{A.1})$ hold.

Consider the following transfer:

$$t(v) = c + \int_{\underline{v}}^{v} p(z)dz + \underline{v}p(\underline{v}) - vp(v),$$

for some constant c. Thus,

$$t(v) - t(v^*) = v^*p(v^*) - vp(v) - \int_{\underline{v}}^{v^*} p(z)dz + \int_{\underline{v}}^{v} p(z)dz.$$

Rearranging terms yields:

$$\begin{aligned} U_1(v) &= U_1(v^*) + \int_{v^*}^{v} p(z)dz \\ &\geq U_1(v^*) + [v - v^*]p(v^*) \\ &= vp(v^*) + t(v^*), \end{aligned}$$

where the inequality in the second line follows from the assumption that $p(v)$ is nondecreasing. As this holds for any v, v^*, we have shown that incentive compatibility (IC) is satisfied.

From the "only if" part of the proof we know that (IC) implies that (IR1) is satisfied everywhere if it is satisfied at $\underline{v}$ and $\overline{v}$. Furthermore we know that (IR1) at $\underline{v}$ and $\overline{v}$ is equivalent to

$$\begin{aligned} t(\underline{v}) &\geq -\underline{v}p(\underline{v}) \text{ and} \\ t(\overline{v}) &\geq \overline{v} - \overline{v}p(\overline{v})). \end{aligned}$$

For the transfer we have defined,

$$\begin{aligned} t(\underline{v}) &= c \\ &\text{and} \\ t(\overline{v}) &= c + \int_{\underline{v}}^{\overline{v}} p(z)dz + \underline{v}p(\underline{v}) - \overline{v}p(\overline{v}) \end{aligned}$$

For these transfers to be compatible with the previous two inequalities, we require

$$\begin{aligned} c &\geq -\underline{v}p(\underline{v}) \text{ and} \\ c &\geq -\underline{v}p(\underline{v}) + \overline{v} - \int_{\underline{v}}^{\overline{v}} p(z)dz. \end{aligned}$$

Let $c = \max[-\underline{v}p(\underline{v}), -\underline{v}p(\underline{v}) + \overline{v} - \int_{\underline{v}}^{\overline{v}} p(z)dz]$. Then (IR1) is satisfied.

Since (IR2) is not a function of v, if it is satisfied anywhere it is satisfied everywhere. Using the analysis in the "only if" part, we can write (IR2) in terms of $\underline{v}$:

$$t(\underline{v}) \leq -\underline{v}p(\underline{v}) - w^*(1 - F(0)) + \int_{\underline{v}}^{\overline{v}} \left(z + w^* - \frac{1 - F(z)}{f(z)}\right) p(z)f(z)dz.$$

For the transfers we have constructed,

$$\begin{aligned} t(\underline{v}) &= c \\ &= \max[-\underline{v}p(\underline{v}), -\underline{v}p(\underline{v}) + \overline{v} - \int_{\underline{v}}^{\overline{v}} p(z)dz]. \end{aligned}$$

There are two possible cases to check to see if this $t(\cdot)$ satisfies (IR2). First, suppose $c = -\underline{v}p(\underline{v})$. Then $(\underline{A.1})$ implies that (IR2) is satisfied. Second, suppose that $c = -\underline{v}p(\underline{v}) + \overline{v} - \int_{\underline{v}}^{\overline{v}} p(z)dz$. Then $(\overline{A.1})$ implies that (IR2) is satisfied. *QED*

B.2 Proof of Proposition 1

In Section 3.2, we showed that the inequalities in (3.10) and (3.11) give the constraints on our problem. We will proceed by examining three cases, corresponding to possible values of $\tilde{v}$.

Case 1: Suppose that $\tilde{v} > 0$. Then rearranging (3.11) yields

$$w^* \left[\frac{F(0) - F(\tilde{v})}{F(\tilde{v})} \right] \geq \tilde{v}.$$

As $\underline{v} < 0 < \overline{v}$, the term in brackets is less than one in magnitude. Therefore, $\tilde{v} = |\tilde{v}| < |w^*|$.

Case 2: Suppose that $\tilde{v} < 0$. Rearranging (3.10) yields

$$w^* \left[\frac{F(0) - F(\tilde{v})}{1 - F(\tilde{v})} \right] \geq -\tilde{v}.$$

Again our assumptions imply that the term in brackets is less than one in magnitude. Thus, $-\tilde{v} = |\tilde{v}| < |w^*|$.

Case 3: Suppose $\tilde{v} = 0$. Since $w^* \neq 0$, $|\tilde{v}| < |w^*|$.

QED

B.3 Obtaining Ex Post Efficient Outcomes

When can we reach the first-best outcome? If $w^* > 0$, we can get the first-best outcome (produce if $v + w^* \geq 0$) if and only if:

$$\max(\underline{v}, -w^*)(1 - F(\max(\underline{v}, -w^*))) + w^*(F(0) - F(\max(\underline{v}, -w^*))) \geq 0. \tag{B.5}$$

We look at the two relevant cases. First, we consider the case where $\max(\underline{v}, -w^*) = -w^*$. Here,

$$-w^*(1 - F(-w^*)) + w^*(F(0) - F(-w^*)) = w^*(F(0) - 1).$$

Thus we cannot achieve the first-best outcome unless $F(0) = 1$ — i.e., unless no firm will ever produce on their own.

In the second case, $\max(\underline{v}, -w^*) = \underline{v}$. Here,

$$\underline{v}(1 - F(\underline{v})) + w^*(F(0) - F(\underline{v})) = \underline{v} + w^*F(0).$$

Thus, we require

$$w^* \geq \frac{-\underline{v}}{F(0)} \tag{B.6}$$

in order to achieve the first-best. If externalities are negative ($w^* < 0$), we can get the first best if and only if

$$-\min(-w^*, \overline{v})F(\min(-w^*, \overline{v})) + w^*(F(0) - F(\min(-w^*, \overline{v}))) \geq 0.$$

By analogous reasoning, this happens if and only if

$$w^* \leq \frac{-\overline{v}}{1 - F(0)} \tag{B.7}$$

and/or $F(0) = 0$.

Note that if the v are distributed uniformly and $\underline{v} < 0 < \overline{v}$, then the conditions for achieving the first-best outcome become

$$|w^*| \geq \overline{v} - \underline{v}.$$

Thus, for $v \sim \mathrm{U}[-1,1]$, $|w^*| \geq 2$ is required for the first-best outcome to be implementable through decentralized bargaining.

B.4 Obtaining the Second-Best Outcome

We are interested in the circumstances under which the second-best outcome improves on the autonomous allocation (i.e., when does the second-best mechanism involve a cut-off type $\tilde{v} \neq 0$?)

We consider first the case of a positive externality ($w^* > 0$; constraint $(\underline{A.1})$ will bind here.) Thus we want to know if there exists a $\tilde{v} < 0$ such that:

$$K(\tilde{v}) \equiv \tilde{v}(1 - F(\tilde{v})) + w^*(F(0) - F(\tilde{v})) \geq 0.$$

We observe that $K(0) = 0$ and

$$\frac{dK(\tilde{v})}{d\tilde{v}} = 1 - F(\tilde{v}) - \tilde{v}f(\tilde{v}) - w^* f(\tilde{v}).$$

Since

$$\text{sign}\left(\frac{dK(\tilde{v})}{d\tilde{v}}\right) = \text{sign}\left(\frac{1 - F(\tilde{v})}{f(\tilde{v})} - \tilde{v} - w^*\right)$$

and $\frac{1-F(\tilde{v})}{f(\tilde{v})} - \tilde{v} - w^*$ is non-increasing in $\tilde{v}$ by our hazard-rate assumption (3.9), such a $\tilde{v} < 0$ exists if and only if:

$$\left.\frac{dK(\tilde{v})}{d\tilde{v}}\right|_{\tilde{v}=0} \leq 0.$$

This is equivalent to

$$w^* \geq \frac{1 - F(0)}{f(0)}. \qquad \text{(B.8)}$$

We now take up the case of a negative externality ($w^* < 0$; constraint $(\overline{A.1})$ will bind here.) Here, we want to know if there exists a $\tilde{v} > 0$ such that:

$$J(\tilde{v}) \equiv -\tilde{v}F(\tilde{v}) + w^*(F(0) - F(\tilde{v})) \geq 0.$$

Analogous to the case above, this happens when

$$\left.\frac{dJ(\tilde{v})}{d\tilde{v}}\right|_{\tilde{v}=0} \geq 0$$

since our hazard rate assumption (3.8) implies that $J(\cdot)$ is single-peaked. This is equivalent to

$$w^* \leq \frac{-F(0)}{f(0)}. \tag{B.9}$$

In the special case of uniform distributions, we find that we must have $w^* \geq \bar{v}$ or $w^* \leq \underline{v}$ in order to improve on the autonomous allocation ($v \geq 0$). Thus, for $v \sim \mathrm{U}[-1, 1]$, it is necessary that

$$|w^*| \geq 1$$

holds for any improvements to be implementable.

B.5 Relaxing Budget Balance

In keeping with our focus on decentralization, we have assumed that budgets must be balanced. However, if there is cross-subsidization of the various activities of governments, budget balance may be too restrictive. Here we show that the qualitative results go through in the more general case in which there is a social cost λ to raising revenues (but no restriction on budget deficits). This section closely follows the general approach of Laffont and Tirole [1993].

We also note that the problem solved in this section is very similar to the problem faced in decentralized bargaining where firm 2 can make a take-it-or-leave-it offer to firm 1. The objective function for that problem is the same as the one here with $\lambda = 0$ and an extra $-U_1$ term subtracted off (since firm 2 dislikes paying transfers to firm 1). The solution to the bargaining problem is the same as the solution to the problem solved in this section with $\frac{\lambda}{1+\lambda} = 1$.

The central government's problem is

$$\max \int_{\underline{v}}^{\overline{v}} [(1+\lambda)(z+w^*)p(z) - \lambda U_1(z) - \lambda U_2)]f(z)dz$$

subject to

$$\text{(IC)} \qquad \frac{dU_1(v)}{dv} = p(v),$$

$$\text{(IR1)} \qquad U_1(v) \geq \max(v, 0), \ \forall v,$$

$$\text{(IR2)} \qquad U_2 \geq w^*(1 - F(0)),$$

$$\text{(Monotonicity)} \qquad p(v) \text{ non-decreasing.}$$

The first thing to observe is that U_2 is not a function of v and enters the objective function with a negative sign. Therefore, (IR2) will always bind at the optimum and U_2 is just a constant which can be ignored. We ignore the monotonicity constraint for now, and solve using optimal control. Letting $\mu(v)$ be the Pontryagin multiplier on the (IC) constraint we can write the Hamiltonian as,

$$H = [(1+\lambda)(v+w^*)p(v) - \lambda U_1(v)]f(v) + \mu(v)p(v). \tag{B.10}$$

Applying the Maximum Principle we have

$$\dot{\mu}(v) = -\frac{\partial H}{\partial U_1} = \lambda f(v).$$

and

$$\frac{\partial H}{\partial p} = (1+\lambda)(v+w^*)f(v) + \mu(v).$$

Assume for the moment that the (IR) constraint binds only at $\underline{v}$. Transversality then requires that $\mu(\overline{v}) = 0$. This gives

$$\mu(v) = \lambda(F(v) - 1).$$

So the conditions on production become

$$p(v) = \begin{cases} 1 & \text{if } v + w^* + \frac{\lambda}{1+\lambda}\frac{(F(v)-1)}{f(v)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, if the (IR) constraint binds only at $\overline{v}$, transversality requires that $\mu(\underline{v}) = 0$. This gives

$$\mu(v) = \lambda F(v)$$

and the conditions on production are

$$P(v) = \begin{cases} 1 & \text{if } v + w^* + \frac{\lambda}{1+\lambda}\frac{F(v)}{f(v)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that these are the same formulas obtained in the corresponding cases in our model with budget balance, except for the fact that λ is now exogenous rather than endogenous.

We can now investigate the potential for coordination. To ensure that monotonicity is satisfied, assume as before that

$$\frac{d}{dv}\left(\frac{F(v)}{f(v)}\right) \geq 0$$

and

$$\frac{d}{dv}\left(\frac{1-F(v)}{f(v)}\right) \leq 0.$$

Define $\tilde{v}(w^*)$ as $\tilde{v}$ such that $v + w^* + \frac{\lambda}{1+\lambda}\frac{F(\tilde{v})}{f(\tilde{v})} = 0$ (i.e., the cut-off type if (IR1) binds only at $\overline{v}$). Then, the cut-off type if $\underline{v}$ only binds is greater than $\tilde{v}(w^*)$; denote it $\hat{v}(w^*)$.

So suppose that $\underline{v}$ binds. This implies

$$\begin{aligned} U_1(\underline{v}) &= 0 \\ U_1(\overline{v}) &= \int_{\hat{v}(w^*)}^{\overline{v}} 1dv \\ &= \overline{v} - \hat{v}(w^*). \end{aligned}$$

The (IR1) constraint will not bind at $\overline{v}$ if and only if $\hat{v}(w^*) < 0$. That is, if and only if

$$w^* > \frac{\lambda}{1+\lambda}\frac{1-F(0)}{f(0)}.$$

Now suppose that $\overline{v}$ binds. This implies

$$\begin{aligned} U_1(\overline{v}) &= \overline{v} \\ U_1(\underline{v}) &= \overline{v} - \int_{\tilde{v}(w^*)}^{\overline{v}} 1dv \\ &= \tilde{v}(w^*). \end{aligned}$$

The (IR1) constraint will not bind at $\underline{v}$ if and only if $\tilde{v}(w^*) > 0$. That is, if and only if

$$w^* < -\frac{\lambda}{1+\lambda}\frac{F(0)}{f(0)}.$$

Thus, the constraint binds at both points if

$$-\frac{\lambda}{1+\lambda}\frac{F(0)}{f(0)} \leq w^* \leq \frac{\lambda}{1+\lambda}\frac{1-F(0)}{f(0)},$$

and coordination cannot improve on the autonomous allocation.

Note that, other than the $\frac{\lambda}{1+\lambda}$ terms, this is the same condition that we derived in the model with budget balance. As λ approaches $+\infty$, this range approaches the range that we derived earlier. Intuitively, with budget balance there is an infinite cost to subsidizing the project when the constraint binds at both points. Here, however, there is some fixed cost λ.

The first-best outcome can be obtained if

$$w^* \leq -\overline{v} - \frac{\lambda}{1+\lambda}\frac{1}{f(\overline{v})}$$

or

$$w^* \geq -\underline{v} + \frac{\lambda}{1+\lambda}\frac{1}{f(\underline{v})}.$$

Notice that here, for the uniform case, if we let λ approach $+\infty$, we get stronger conditions than under budget balance. The intuition is that as the (IR1) constraint binds at only one end as the second-best approaches the first-best, the cost of subsidizing the project is no longer infinite. This reinforces the point that $\underline{\lambda}$ and $\overline{\lambda}$ in the budget balance model are endogenous and depend on the parameters of the model such as w^*, whereas λ here is exogenous.

Finally, for values of w^* which do not fall into either the autonomous or the first-best ranges, we have a second-best outcome which improves on the autonomous allocation. As before, the second-best outcome can be described by a cut-off type $\tilde{v}$ and can be implemented using second-best Pigouvian taxes and subsidies.

A key here is that reducing the amount of money required for the project at hand frees money for other (valued) projects elsewhere. So, having negative net subsidies (i.e., net taxes) is viable in the model here. Since when the gains from leaving the money in the projects here are small (as we are very close to the first-best) they are outweighed (socially) by diverting the funds to other projects it is more difficult to obtain efficient outcomes. In other words, in this setting, a marginal as well as a total cost/benefit evaluation is required. This is important in understanding why, for example, it is harder to get first-best investment when firm 2 makes a take-it-or-leave-it offer as compared to a mediator or government proposing a balanced budget scheme.