Tolerance of Ambiguity and Entrepreneurial Innovation*

Luca Rigotti† Matthew Ryan‡ Rhema Vaithianathan§

February, 2007

Abstract

We build a general equilibrium model of occupation choice, where the risks inherent in a technology embodying a new innovation are only vaguely known (ambiguous), whereas the old technology has known risks. Using the Arrow-Hurwicz criterion to characterise decision-making in the face of ambiguity, we distinguish agents according to their degree of ambiguity aversion. Equilibrium requires that all agents choose optimal occupations, labour markets clear, and labour contracts within firms are Pareto efficient. In equilibrium, we show that occupations are ordered by ambiguity: the most ambiguity tolerant agents own innovative firms, the least ambiguity tolerant have occupations in old-technology firms, and a middle group supply labour to the innovative sector. By considering a simple dynamic version of this economy, in which agents learn about the new technology, we deduce properties of the diffusion of the innovation. When the new technology is superior to the old, we obtain the S-shaped diffusion profile commonly observed in data. More surprisingly, when the new technology is inferior, our model can produce a boom-bust profile reminiscent of the dot-com crash. Even though bad news predominates, any information reduces ambiguity, which encourages adoption. In the early phases of diffusion, the ambiguity-reduction effect of new information dominates, promoting significant adoption of the inferior technology. Later on, the bad-news effect takes over, leading to iTs ultimate demise.

JEL Codes: D5, D8, L2, J24, M13, O31

Keywords: Occupational Choice, Entrepreneurship, Ambiguity, Optimism/Pessimism, Belief Functions, Innovation, Diffusion.

---

*We thank Maitreesh Ghatak and two anonymous referees for comments that significantly improved the paper. Rigotti also thanks the Department of Economics at the University of Auckland for generous hospitality during January 2005.

†Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708, USA. Email: rigotti@duke.edu

‡Department of Economics, University of Auckland, PB 92019 Auckland, New Zealand. Email: m.ryan@auckland.ac.nz

§Department of Economics, University of Auckland, PB 92019 Auckland, New Zealand. Email: r.vaithianathan@auckland.ac.nz

1 Introduction

Entrepreneurial innovators have long been acknowledged as a central driver of economic progress in capitalist systems. Empirical evidence that propensity to start a business increases with wealth (Evans and Jovanavic, 1989; Blanchflower and Oswald, 1999) has lead to a theoretical focus on credit market imperfections as the principal barrier to entrepreneurship.[1] A number of occupational choice models illustrate the effect of credit-rationing in a general equilibrium setting. In these models, credit constraints prevent individuals setting up businesses, or taking up profitable opportunities. In Banerjee and Newman (1993), credit market imperfections (and a sufficiently unequal wealth distribution) can make entrepreneurship entirely disappear from a capitalist economy, with self-employment in cottage industries dominating. Similarly, in Aghion and Bolton (1997) credit-market rationing exists in the long-run, where profitable investment opportunities remain unexploited by entrepreneurs, as long as there is sufficient wealth inequality to start with.[2]

In this literature, the profitability of innovative projects is modelled as a matter of fact rather than opinion: all agents in the economy know these projects are profitable.[3] This view leaves no room for the existence of innovations which turn out to be failures. In this paper, we provide a theory of entrepreneurship which allows for the existence and (temporary) adoption of unprofitable new technologies.[4] This theory can then be used to explore how the distribution of entrepreneurial types in an economy can facilitate both good and bad innovation. The entrepreneur can be as destructively creative as he is creatively destructive.

We build an occupational choice model where not all individuals share the same preferences for innovation. To make this setting stark, credit constraints are ruled out, and therefore the only barriers to entrepreneurship are in the individuals' ability to tolerate the inherent doubts and uncertainties surrounding a new technology.[5] The function of the entrepreneur is to perceive new

[1] In the standard theory, even if there is a temporary slump in output while general technology is transformed, as discussed by Aghion and Howitt (1992), in the long-run, innovation always has A positive effect on growth.

[2] Ghatak, Morelli and Sjöström (2001), however, point out that a second-best argument in favour of credit constraints might also exist. When there are moral-hazard problems between firms' owners and workers, credit market rationing may in fact be welfare improving. In their model, the returns to entrepreneurship include a higher than market return to capital, motivating poor, young agents to work hard and save in order to use their own capital in the second period to establish a business.

[3] Banchflower and Oswald (1998) have model where where some people are alert to opportunities and others are not.

[4] In the long run, everyone will know whether the new technology is inferior, and will abandon it if that is the case.

[5] This theory is not meant as an alternative to the credit constraint story; it is meant to add richness to the existing models of entrepreneurship. It is also consistent with some recent empirical evidence. The role of tolerance for ambiguity and entrepreneurship has been well documented in the management literature (see Bhidé, 2000, for a comprehensive survey). In the economics literature, Vissing-Jorgenson (2002) finds some evidence that entrepreneurial optimism might explain the equity premium puzzle. Similarly, Huang (2005) shows that tolerance of ambiguity might be an important component in explaining inter-country differences in adopting innovations. Huang (2005) uses data on the psychological propensities of IBM employees from 50 countries which incorporates questions designed to measure respondents' ambiguity tolerenace. Huang correlates ambiguity aversion with the growth of industries located where stock prices are less informative. He finds that countries characterised by high levels of ambiguity tolerance demonstrate more rapid growth in "opaque" industries.

opportunities as much as exploiting them. As Schumpeter (1947, p.152) argues: "It is in most cases only one man, or a few men, who see the new possibility". In particular, entrepreneurship is not just awareness of the existence of a new technology, but the perception that this new technology represents an "opportunity", rather than a dangerous risk. We explore how the distribution of ambiguity tolerance in the economy determines the path of diffusion. Our primary interest is not in the steady state, but in the transition path – since the important function of a Schumpeterian entrepreneur is in influencing the transition.[6]

We consider an environment in which a lack of objective information about an invention justifies a variety of *ex ante* opinions on its economic potential. Such situations are called *ambiguous* since the probability of events is not uniquely determined.[7] The *Arrow-Hurwicz criterion* offers a tractable model of behavior in the face of ambiguous information. Ambiguity is characterised by a *set* of probabilities consistent with the available information. Individuals evaluate prospects by calculating a weighted average of the highest and lowest expected return with respect to probabilities in this set. Ambiguity tolerance is measured by the relative weight placed on the highest expectation. Therefore, a single parameter describes a decision-maker's ambiguity tolerance in a way reminiscent of a coefficient of risk tolerance.

Our model formalises the connection between agents' ambiguity tolerance, their occupational choice and their willingness to innovate. This allows us to relate the "entrepreneurial spirit" of an economy (characterised by the distribution of ambiguity tolerance amongst its members) with the rapidity with which the economy adopts a new technology. The model reveals how too much entrepreneurial spirit might lead to excessive adoption of bad technologies – resulting in a boom-crash profile for some industries.

The model goes as follows. In the economy there are two (freely available) technologies that can be used for the production of a single consumption good. One technology is a new innovation and there is significant ambiguity about its productivity. The other technology is traditional and its productivity is not ambiguous. For either technology, production requires the input of two labour units. Within each firm, these two units must share revenue in a Pareto efficient fashion.

Our first result shows that within firms using the innovative technology, the Pareto efficient sharing rule takes a simple and familiar form. The more ambiguity tolerant agent pays the other a fixed wage, subject to the limited liability constraint that wages are paid only to the extent that realised revenue allows. The role of "owner" and "worker" within a firm, therefore, are endogenously generated by the agents' ambiguity tolerance.

Firm formation is also endogenous. Each agent decides which sector to join, and in the case of the innovative sector, whether to be an owner or a worker. In particular, there are essentially three occupational choices: own a firm using the new ambiguous technology, work in such a firm, or seek employment in the non-ambiguous industry. Our second result shows that

[6]As Schumpeter points out: "The mechanism of economic change in capitalist society pivots on entrepreneurial activity" (*ibid.*, p.150).

[7]Ellsberg (1961) has first illustrated how the distinction between risk and ambiguity is important for individual decision making. Fox and Tversky (1995) report experimental results which show that people are less willing to bet on vague events if they are comparing it with an alternative with better known odds. See Camerer (1995) for a survey of the experimental evidence. Other economic applications of ambiguity aversion include Mukerji (1998) and Mukerji and Tallon (2001).

an equilibrium of our economy always exists, and is essentially unique. Equilibrium is also easily characterised. The least ambiguity-averse types own firms in the innovative industry, the most ambiguity averse operate firms in the traditional industry, while the middle group work in the innovative industry. In other words, occupations have a unique ambiguity ranking in equilibrium. Ambiguity, therefore, is a factor influencing entrepreneurship.[8]

This result is reminiscent of Kihlstrom and Laffont (1979), who recognised that attitude to risk can predict occupational choice – with the more risk averse being the worker. In our model, however, wage contracts do not necessarily provide full insurance against income ambiguity, while in Kihlstrom and Laffont (1979) wages are safe by assumption.

Our final results concern the evolution of ambiguity in a dynamic version of our economy and its impact on diffusion. Over time, information about the true productivity of the innovative technology is publicly revealed, and all individuals update their beliefs and re-evaluate their occupation choices. As ambiguity declines, the degree to which choices are governed by ambiguity attitudes subsides, and in the steady-state one technology comes to dominate the market. The time profile of the number of firms operating in the innovative industry generates a diffusion path for the new technology.

A stylized fact of such diffusion paths is that they rise very slowly in the initial phases, then gradually accelerate before slowing again as a point of saturation is reached. This gives rise to the well-documented S-shaped profile (Griliches, 1957; Mansfield, 1968; Gort and Klepper, 1982; Jovanovich and MacDonald, 1994; Jovanovich and Lach, 1989). Our model is consistent with this stylized fact.

Gort and Klepper suggest that part of the explanation for the S-shaped profile could be the process of uncertainty resolution. Jensen (1982), Balcer and Lipman (1984), and Vettas (1998) also express variants on this theme. In Jensen for instance, diffusion occurs because agents begin with heterogeneous priors and learn about the true profitability of the technology. As long as the technology is superior, every agent (however pessimistic initially) eventually comes to appreciate this superiority. The diffusion profile has an S-shape provided there is a Uniform distribution of priors over the population initially.

While our explanation of diffusion is also about the resolution of uncertainty, it differs from earlier theories since we allow for the presence of ambiguity. Information plays two distinct roles in fuelling diffusion. Firstly, as in Jensen, it allows agents to update their beliefs. Secondly, it reduces ambiguity by narrowing the range of possible beliefs that could be held. When the new technology is superior and the distribution of ambiguity tolerance is right-skewed, we generate an S-shaped diffusion path.[9]

The ambiguity reducing effect which is unique to our model generates a striking new result when the new technology is instead inferior – the potential for a boom-bust diffusion profile. These are diffusion paths which are at first positively sloped, despite the inferior technology generating a preponderance of bad news. The initial upswing comes about because the ambiguity-reducing effect of early news outweighs the information effect of bad news. However, once enough

[8] This factor is different from liquidity constraints (Banerjee and Newman, 1993; Holtz-Eakin, Joulfaian and Rosen 1994a,b; Aghion and Bolton, 1997; Piketty 1997; and Ghatak, Morelli and Sjöström, 2001) and entrepreneurial ability (Lloyd-Ellis and Bernhardt, 2000) which have been considered previously in the literature.

[9] If worker income is free of default risk, even a Uniform distribution suffices.

uncertainty has washed out of the system, the bad news effect takes over and the industry begins its terminal decline.

Importantly, these boom-bust profiles for inferior innovations occur in our model for the same right-skewed distribution of types that gives rise to the S-shaped profile for successful innovations. Moreover, the rapidity and scale of the initial growth phase depends on the entrepreneurial spirit of the economy. More entrepreneurial economies may experience more rapid and widespread diffusion of bad innovations, and correspondingly more severe crashes.

Our approach differs from theories which explain the bust of good innovations. Barbarino and Jovanovic (2005) model investment in an industry where there is the possibility of "overshooting" in the market – at which point prices fall to zero and the market dies. Geanakoplos (2002) incorporates liquidity in a model with margin-trading and agents who are heterogenous in their level of optimism regarding an asset. He points out that when the price of an asset falls, the composition of the agents who hold the asset changes since the optimists are "wiped out" and become illiquid. This has an additional effect on the price of the asset – resulting in a more severe crash. These papers focus on the crash of potentially good innovations rather than the boom and bust of bad innovations. In our model, the new innovation is intrinsically unprofitable, yet positive diffusion takes place.

The remainder of the paper is organised as follows. Section 2 discusses the Arrow-Hurwicz decision framework in more detail. Section 3 describes the model, and contains two key results: the Pareto efficient sharing rule for innovative firms, and the characterisation of equilibrium. Section 4 introduces a dynamic version of our economy. We discuss the mechanism for updating sets of priors, and describe the model's implications for the diffusion of new technologies. Section 5 concludes. Three Appendices contain technical details omitted from the text.

2 Decision-making under ambiguity

We consider situations in which risks are imprecisely known (ambiguous), but determined by objective information.[10] The decision-making framework is that of Anscombe and Aumann (1963), in which actions determine roulette lotteries conditional on the unknown state of the world. There exists *ambiguous* public information about the state. This information is quantified by assigning an interval of probabilities to each event. All individuals are risk neutral, but vary in their tolerance for ambiguity.

2.1 Ambiguity

Let $\mathcal{L}$ denote a set of *lotteries*: random variables with outcomes in $\mathcal{R}_+$. (We do not describe the underlying probability space directly, since only the distribution over $\mathcal{R}_+$ matters for preferences.) The objects of choice are *(Anscombe-Aumann) acts* $f : \Theta \to \mathcal{L}$, with Θ is a finite set of states. There is *ambiguous* information about Θ, described by a closed and convex set of probabilities $\Pi \subseteq \Delta(\Theta)$, where $\Delta(\Theta)$ is the unit simplex in $\mathcal{R}_+^{\Theta}$.

[10] Some decision-theorists refer to these as situations of *imprecise risk* – for example, Jaffray (1991) and Philippe (2002).

For example, if Π is a singleton, information is precise and we have a situation of pure risk. In this case, acts are two-stage roulette lotteries, so each outcome occurs with a known probability. Conversely, if $\Pi = \Delta(\Theta)$, then we have pure uncertainty. We allow that Π may fall anywhere in between these two well-known extremes. The following Ellsberg-like example illustrates an intermediate case.

Example I. Lottery 1 awards outcome M with probability θ_1 and outcome m with probability $1-\theta_1$. Lottery 2 yields the same outcomes, but with probabilities θ_2 and $1-\theta_2$ respectively. There are 100 lottery tickets in an urn, with one to be drawn at random. It is known that no less than 20 and no more than 75 of the tickets are for Lottery 1, while the remaining are Lottery 2 tickets. Therefore, $\Theta = \{\theta_1, \theta_2\}$ and the probability of drawing a ticket for Lottery 1 lies in the interval $\left[\frac{1}{5}, \frac{3}{4}\right]$.

Figure 1 illustrates Example I using a tree. At the initial node, Nature chooses $\theta \in \Theta = \{\theta_1, \theta_2\}$. For each $\theta \in \Theta$, the same two outcomes are possible, indicated by the end nodes M and m. The probability of drawing a Lottery 1 ticket lies in the interval $[v_1, 1 - v_2]$, where v_i is the lowest possible probability of drawing a ticket for Lottery i. (In this case, $v_1 = \frac{1}{5}$ and $v_2 = \frac{1}{4}$.) Ambiguity obtains because information about which θ_i will be used to award the outcomes is objective but vague; it is captured by a probability interval. Here, Π is the set $\{\pi \in \Delta(\Theta) \mid \pi(\theta_1) \in [v_1, 1 - v_2]\}$.

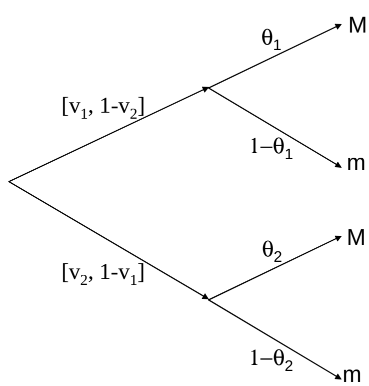


Figure 1: A two-stage lottery with ambiguity in the first stage

Dempster (1967) described a general class of situations in which objective information may lead to ambiguity of this sort.[11] Let (S, Σ, p) be a measure space and $\Gamma : S \twoheadrightarrow \Theta$ a measurable

[11] See Srivastava and Mock (2002) for a straightforward introduction, and Mukerji (1997) for a useful discussion in the context of economic decision-making.

correspondence, mapping each element of S to a non-empty subset of Θ. We shall refer to S as the set of *fundamental states*, while Θ contains the *payoff-relevant states*. We call Γ the *information correspondence*. If the fundamental state $s \in S$ is realised, the available information implies that some payoff-relevant state in $\Gamma(s)$ also occurs, but nothing is known about the relative likelihoods of states in $\Gamma(s)$. If Γ is a *function*, then (S, Σ, p) and Γ together induce a probability on Θ and we have pure risk. If $\Gamma(s) = \Theta$ for every $s \in S$, then we have pure uncertainty. In between are situations such as Example I and those in Ellsberg (1961), in which event probabilities are known with varying degrees of precision or ambiguity.

Given (S, Σ, p) and Γ, we define an associated *belief function* $\underline{v} : 2^{\Theta} \to [0,1]$ as follows:

$$\underline{v}(E) \;=\; p(\{s \in S \mid \Gamma(s) \subseteq E\}).$$

The probability of event E therefore lies in the interval $[\underline{v}(E), 1 - \underline{v}(E^c)]$, and

$$\Pi = \left\{ \pi \in \Delta(\Theta) \;\middle|\; \pi(E) := \sum_{\theta \in E} \pi(\theta) \in [\underline{v}(E), 1 - \underline{v}(E^c)] \quad \forall E \subseteq \Theta \right\}.$$

Economists, adopting terminology from cooperative game theory, often refer to Π as the *core* of the belief function $\underline{v}$.[12]

2.2 Decision-making criterion

Given the set Π, how should an individual evaluate an act f? This evaluation will depend on the "size" of Π and the individual's tolerance for the associated ambiguity. We consider decision-makers whose behavior conforms to the *Arrow-Hurwicz criterion*.[13] Letting $\overline{L}$ denote the expected value of the lottery $L \in \mathcal{L}$, this criterion evaluates the act f according to the utility function:

$$U(f;\ \lambda) \;=\; \lambda \left[\max_{\pi \in \Pi} \ \sum_{\theta \in \Theta} \pi(\theta)\, \overline{f(\theta)} \right] \;+\; (1-\lambda) \left[\min_{\pi \in \Pi} \ \sum_{\theta \in \Theta} \pi(\theta)\, \overline{f(\theta)} \right] \tag{1}$$

Since our primary interest is on the effects of heterogeneity in ambiguity tolerance, we have assumed that all decision-makers are risk neutral for simplicity. The parameter $\lambda \in [0,1]$ is the *ambiguity tolerance index*. A decision-maker with $\lambda = 0$ [$\lambda = 1$] focusses all attention on the most pessimistic [optimistic] $\pi \in \Pi$.[14] Observe that the identity of the most pessimistic [optimistic] $\pi \in \Pi$ depends on the particular f being evaluated, so (1) is not linear in the vector $\left(\overline{f(\theta)}\right)_{\theta \in \Theta}$ of state-contingent expected returns.

[12] There is some potential for confusion, as non-economists reserve the term "core" for a different property of belief functions – see Shafer (1976, p.40).

[13] See Arrow and Hurwicz (1972) and Luce and Raiffa (1957, p.282).

[14] Models of choice leading to the Arrow-Hurwicz criterion have been axiomatised by Jaffray (1989, 1991, 1994), Hendon *et al.* (1994) and Jaffray and Wakker (1994). For the extension to infinite state spaces, see Philippe, Debs and Jaffray (1999). There is also a literature on belief functions as *subjective* constructs in the characterisation of choice under (pure) uncertainty. See Jaffray and Phillipe (1997) and Ghirardato (2001). Recently, Viero (2006) has axiomatized a similar model in an Anscombe-Aumann setting in which the probability of the "roulette-lotteries" is objective but vague.

A graph may help to clarify how utility function (1) works. In Figure 2, we take $\Theta = \{\theta_1, \theta_2\}$ and the axes measure the expected value in each state. Given some $\pi \in \Delta(\Theta)$, we may construct a line through $\overline{f}$ representing all points with expected value equal to $\overline{f}$ according to π. As π varies over Π, we obtain a family of such lines. The locus of points whose *lowest* expected value over Π is equal to the lowest expected value of $\overline{f}$ is described by a kinked line, with the kink occurring at certainty.[15] Similarly for the locus of points which have *highest* expected value equal to the highest expected value of $\overline{f}$. These two kinked lines are the two thin dashed lines through $\overline{f}$ in Figure 2: the one that kinks further to the South corresponds to an indifference curve when $\lambda = 0$ and the other to an indifference curve when $\lambda = 1$. The indifference curve that corresponds to the utility function in (1) is a convex combination of these two, and it is drawn as a thick kinked line.[16] Therefore, these indifference curves do not have the constant marginal rate of substitution that we associate with risk-neutrality (unless $\lambda = \frac{1}{2}$). From Figure 2, one can also see that preferences are *convex* in this two-state world as long as $\lambda \leq \frac{1}{2}$.

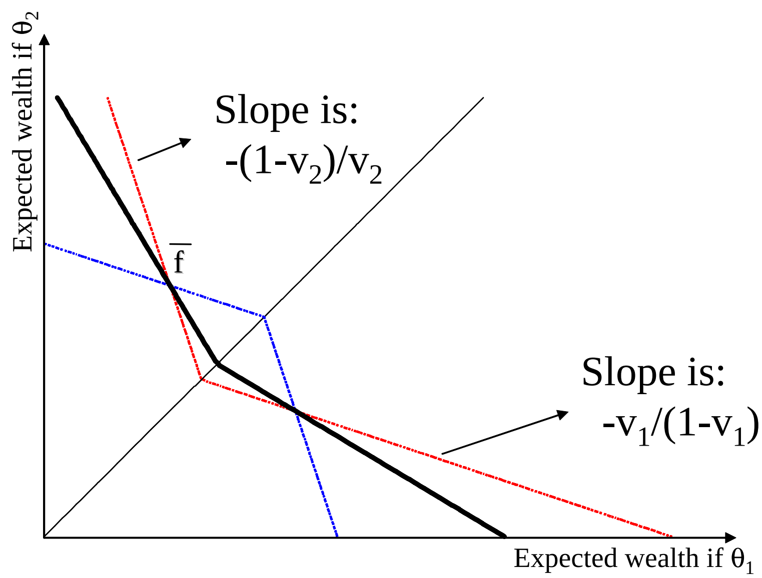


Figure 2: Arrow-Hurwicz preferences

We conclude this section with an important observation about (1). The value of λ captures an intrinsic property of the decision-maker, and is logically independent of the domain of choice. It is analogous to an index of risk tolerance. In this sense, λ is a "deep" parameter. By contrast, in a subjective expected utility (SEU) world, each decision-maker would make decisions by choosing a subjective prior over Π, but this prior would be specific to the parameter space Θ and the information about Θ leading to Π. This is important for the testability of the model. We

[15] This is because, with only two states, the minimising probability is the same for any two bundles on the same side of the certainty line.

[16] As drawn, the thick indifference curve is for some $\lambda \in \left(0, \frac{1}{2}\right)$.

shall see that our model predicts the S-shaped diffusion profile typical of successful innovations provided the distribution of λ over the population takes a particular form. To test this prediction, one could estimate the empirical distribution of λ from laboratory experiments. The empirical distribution of priors in an SEU world, on the other hand, must be evaluated in the context of the model to be tested. Estimates obtained in the lab may not readily transfer to the real-world situations of interest.[17]

3 An entrepreneurial economy

In this section we build a simple model of occupational choice. There is a continuum of agents, described by a differentiable and strictly increasing distribution function H on $[0,1]$. We interpret $H(z)$ as the density of agents with ambiguity tolerance $\lambda \leq z$. It is further assumed that $H(0) = 0$. Our assumptions imply that the total density of agents is unity, and that the distribution of ambiguity attitudes is atomless.

There is a single consumption good, and two technologies for producing it: an established technology ($T1$) and a new innovation ($T2$). Each technology requires the input of two (full-time) agents. We assume for simplicity that the technologies are freely available – one may think of them as different techniques for deploying the human capital of the firm (rather than technologies embodied in capital goods). A firm is formed when two agents decide to join forces to produce using one of the available technologies. It is important to note that the only constraint on firm structure is the necessity of two units of labour for production to take place – there is no predetermined role of "owner" or "worker". However, these roles will emerge endogenously in equilibrium.

A firm's output depends on the technology it employs plus some stochastic factors. Technology $T1$ is safe, and generates expected output $2K > 0$ (the reason for the "2" will become clear shortly), independently of the state. The output of a $T2$ firm, on the other hand, is a random variable $R(\theta) \in \mathcal{L}$ that depends on θ. We shall assume that $R(\theta)$ takes value $M > 0$ with probability θ, and $m \in (0, M)$ with probability $1 - \theta$. There is ambiguous public information about the state, embodied in a closed and convex set of probabilities Π. We let

$$\underline{p} = \min_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\,\theta$$

$$\overline{p} = \max_{\pi \in \Pi} \sum_{\theta \in \Theta} \pi(\theta)\,\theta$$

and

$$\Delta p = \overline{p} - \underline{p}.$$

Thus, $\overline{p}$ is the upper, and $\underline{p}$ the lower, bound on the probability that a $T2$ firm realises output level M. The quantity Δp is the difference between these two extremes. If $\Delta p > 0$ then there is genuine ambiguity about the prospects of $T2$ firms.

[17] Thus, when Jensen (1982, Theorem 5) shows that diffusion profiles in his Bayesian model of innovation are S-shaped under the assumption of a Uniform distribution of priors, it is hard to know how to interpret or test this result.

Each agent must decide which sort of firm to join (a $T1$ firm or a $T2$ firm), and the partners in each firm must decide how to share the realised output. We make two assumptions about sharing rules. First, we impose the natural *institutional* constraint that sharing satisfy limited liability: agents' consumption cannot be strictly negative.[18] The limited liability clause implies that m is the highest sure level of consumption that can be guaranteed to any agent in a $T2$ firm.[19] Second, we impose the *equilibrium* condition that sharing rules must be Pareto efficient within each partnership.

3.1 Pareto optimal sharing rules

In this section, we study Pareto optimal sharing rules for firms that use technology $T2$. Technology $T1$ produces $2K$ in expectation in each state, so any sharing rule is Pareto optimal in such firms.

Since the distribution of ambiguity tolerance is atomless, we ignore the possibility that both agents in a firm share the same λ. Consider two agents, one of type λ' and the other of type $\lambda'' > \lambda'$, operating a $T2$ firm. Let s'_M $[s'_m]$ be the output granted to the type λ' agent when M $[m]$ is realised. Define s''_M and s''_m similarly. Limited liability implies $s'_y, s''_y \in [0, y]$ and $s'_y + s''_y = y$ for each $y \in \{M, m\}$. The four numbers $(s'_M, s'_m, s''_M, s''_m)$ determine the sharing rule for the firm. The Pareto optimal rules depend on the ambiguity tolerance parameters λ' and λ''.[20] However, provided $\lambda'' \leq \frac{1}{2}$, the optimal rule takes a simple and familiar form.

Proposition 3.1 *Suppose $\Delta p > 0$ and the agents have ambiguity tolerance parameters λ' and λ'' with $\lambda' < \lambda'' \leq \frac{1}{2}$. Then sharing is Pareto efficient iff $s'_m = \min\{s'_M, m\}$.*

Proof: See Appendix A. □

Proposition 3.1 asserts that it is Pareto efficient for the less ambiguity tolerant partner to be offered a fixed "wage" $w = s'_M \in [0, M]$, which is subject to default if $w > m$. The partner with the higher tolerance for ambiguity becomes the residual output claimant. It is therefore natural to describe the former agent as the "worker" and the latter as the "owner". We shall adopt this terminology in what follows.

In light of Proposition 3.1, *we henceforth assume that* $H\left(\frac{1}{2}\right) = 1$, so every agent has $\lambda \leq \frac{1}{2}$. To better understand the significance of this assumption, let us define

$$p^\lambda = \lambda \overline{p} + (1 - \lambda)\underline{p}$$

[18] For a setting with multiple priors but no limited liability restrictions, see Kelsey and Spanjers (2004).

[19] It would be straightforward to elaborate the model so that agents also have heterogeneous endowments of the consumption good, and must use these to meet obligations to their partners if necessary (as in Kihlstrom and Laffont, 1979). Wealthier owners may then be able to pay lower wages in equilibrium, because of the lower likelihood of default. Formally, this effect is similar to a credit constraint in a model where owners must invest capital in order to start their businesses, and credit markets are imperfect (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Ghatak, Morelli and Sjöström, 2001).

[20] Analogously, Pareto efficient risk-sharing in a world without ambiguity depends on the risk tolerance of each party.

and

$$q^\lambda = \lambda \underline{p} + (1 - \lambda)\,\overline{p}.$$

Consider an agent with utility function (1) who is a partner in a $T2$ firm and receives s_M $[s_m]$ when M $[m]$ is realised. If $s_M > s_m$, she behaves as if she attaches probability p^λ to the event that M is realised. Otherwise, she behaves as if the probability of M is q^λ. Observe that $p^\lambda = q^\lambda$ if and only if $\lambda = \frac{1}{2}$ or $\Delta p = 0$, in which case

$$p^\lambda = q^\lambda = \frac{1}{2}\underline{p} + \frac{1}{2}\overline{p}.$$

This is the probability on M obtained from a Uniform prior on $\left[\underline{p}, \overline{p}\right]$.[21] Moreover, $p^\lambda \leq q^\lambda$ iff $\lambda \leq \frac{1}{2}$, so the utility (1) is quasi-concave in (s_M, s_m) iff $\lambda \leq \frac{1}{2}$. (See Appendix A for further elaboration.) Thus, assuming $H\left(\frac{1}{2}\right) = 1$ amounts to restricting ambiguity attitudes to the range from extreme pessimism ($\lambda = 0$) to "ambiguity neutrality" ($\lambda = \frac{1}{2}$). It is analogous to excluding *risk-seeking* preferences in a model of behaviour in the face of pure risk – a common enough assumption.

3.2 Equilibrium

In view of Proposition 3.1 and our assumption that $H\left(\frac{1}{2}\right) = 1$, we may assume that all owners of $T2$ firms offer the same wage w to their workers. The $T2$ labour market will not support more than one wage, since all $T2$ firms are identical.[22] Since there is no ambiguity about expected output in $T1$ firms, we shall assume without loss of generality that $T1$ partnerships split realised output equally in all states. Thus, each partner in a $T1$ firm has expected consumption K.[23] An equilibrium consists of a wage w and an allocation of agents to occupations, such that all agents undertake utility-maximising occupations and the $T2$ labour market clears. We shall now specify these equilibrium conditions more precisely.

Each agent has three occupational options: own a $T2$ firm; work in a $T2$ firm; or join a $T1$ firm. Exactly one occupation must be chosen: we do not allow agents to divide their time among a portfolio of jobs. Let $\mathcal{O} = \{O, W, 1\}$ index these occupations in the obvious manner. An allocation function for the economy is a μ-measurable function $\phi : \left[0, \frac{1}{2}\right] \to \mathcal{O}$, where μ is the measure associated with distribution function H. Thus, for each $o \in \mathcal{O}$, $\mu\left[\phi^{-1}(o)\right]$ is the mass of agents assigned to occupation o.

[21] In other words, in this two-outcome set-up, an agent with $\lambda = \frac{1}{2}$ behaves as an SEU maximiser who applies Laplace's "principle of insufficient reason".

[22] This is important for the implementability of the Pareto efficient sharing rule. There is no need for sharing to be determined by within-firm bargaining, which would require each bargaining agent to know the type of the other – an implausible assumption. Instead, we rely on markets to set an equilibrium value for the $T2$ wage. Optimal occupational selection by individuals ensures that revenue is shared efficiently within firms – see Proposition 3.2.

[23] Hence the reason for specifying expected $T1$ output as $2K$.

Given w, an agent with ambiguity tolerance parameter λ obtains utility[24]

$$\lambda \max_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) \left[\overline{\max\{R(\theta)-w,\,0\}}\right] + (1-\lambda) \min_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) \left[\overline{\max\{R(\theta)-w,\,0\}}\right]$$

from owning a $T2$ firm, and utility

$$\lambda \max_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) \left[\overline{\min\{w, R(\theta)\}}\right] + (1-\lambda) \min_{\pi\in\Pi} \sum_{\theta\in\Theta} \pi(\theta) \left[\overline{\min\{w, R(\theta)\}}\right]$$

from working in such a firm. As remarked, everyone obtains K from an occupation in the $T1$ sector. For any agent of type λ, $BR(w;\lambda) \subseteq \mathcal{O}$ is the set of occupations that maximise her utility given w.

Formally, an equilibrium is defined as follows:

Definition 1 *The couple (w,ϕ) is an* equilibrium *if*

(i) $\phi(j) \in BR(w;\lambda) \qquad \forall j \in \left[0, \frac{1}{2}\right]$*; and*

(ii) $\mu\left[\phi^{-1}(O)\right] = \mu\left[\phi^{-1}(W)\right]$.

Condition (i) says that each type is assigned to a utility-maximising occupation. Condition (ii) says that the $T2$ labour market clears. Note that it is possible that only one technology is used in equilibrium.

The following Proposition verifies that equilibria exist and all "macro" variables are uniquely determined.

Proposition 3.2 *Suppose $\Delta p > 0$.*

i. For any $w < \frac{1}{2}(M+m)$, there are unique values $\tilde{\lambda}(w)$ and $\hat{\lambda}(w)$ in $\left[0, \frac{1}{2}\right]$ such that that $\tilde{\lambda}(w) \leq \hat{\lambda}(w)$ and

$$BR(w;\lambda) = \begin{cases} \{O\} & \text{if } \lambda > \hat{\lambda}(w) \\ \{W,1\} & \text{if } \tilde{\lambda}(w) < \lambda < \hat{\lambda}(w) \text{ and } w = K \leq m \\ \{W\} & \text{if } \tilde{\lambda}(w) < \lambda < \hat{\lambda}(w) \text{ and } (w > m \text{ or } w \neq K) \\ \{1\} & \text{if } \lambda < \tilde{\lambda}(w) \end{cases}$$

ii. An equilibrium (w,ϕ) exists and is essentially unique (i.e., $\mu\left[\phi^{-1}(o)\right]$ is unique for each $o \in \mathcal{O}$ and w is unique if $\mu\left[\phi^{-1}(1)\right] < 1$).

[24] By $\overline{\max\{R(\theta)-w,0\}}$ we mean the lottery that delivers $\max\{M-w,0\}$ with probability θ and $\max\{m-w,0\}$ with probability $1-\theta$. The expression $\overline{\min\{R(\theta),w\}}$ should be interpreted similarly. Likewise, we shall use $\overline{\max\{R-w,0\}}$ and $\overline{\min\{R,w\}}$ to denote the corresponding acts.

The proof of Proposition 3.2 is instructive, so we include (most of) it in text below. But first we pause to interpret the Proposition.

Part (i) says that, given $\Delta p > 0$ and any "reasonable" value for w,[25] there are unique numbers $\tilde{\lambda}(w)$ and $\hat{\lambda}(w)$ such that all types $\lambda > \hat{\lambda}(w)$ choose to own $T2$ firms and all types $\lambda < \tilde{\lambda}(w)$ choose an occupation in the $T1$ sector. Intermediate types – those strictly between $\tilde{\lambda}(w)$ and $\hat{\lambda}(w)$ – either strictly prefer to work in $T2$ firms or are indifferent between this option and an occupation in the $T1$ sector. In other words, occupations are ordered by the degree of ambiguity tolerance required to undertake them: ownership of a $T2$ firm attracts only the most ambiguity tolerant, while the least tolerant take jobs in the $T1$ sector. A middle group supplies labour to $T2$ firms.[26]

Part (ii) guarantees equilibrium existence, at least when there is genuine ambiguity about the return to technology $T2$ (i.e., $\Delta p > 0$). Moreover, the size of each sector is the same in any equilibrium, and the equilibrium wage in the $T2$ sector is also unique (unless the $T2$ sector is empty).

If $\Delta p = 0$ – a case excluded from Proposition 3.2(ii) – matters are even more straightforward. Since there is no ambiguity, all agents seek occupations in whichever industry offers the highest expected output. Therefore, equilibrium exists in this case also, and is unique in all "macro" aspects unless $T1$ and $T2$ have identical expected returns. In particular, there is a unique w that splits expected $T2$ returns equally, although many other sharing rules are also Pareto optimal in the absence of ambiguity.

Proof of Proposition 3.2. We start with a proof of (i). Observe that expected output $\overline{R(\theta)}$ may be decomposed into the owner's and worker's expected shares as follows:

$$\overline{R(\theta)} \;=\; \overline{\max\{R(\theta) - w,\, 0\}} \;+\; \overline{\min\{R(\theta)\,,\, w\}}.$$

Each expected share is *comonotone* with expected output,[27] so

$$U(R;\,\lambda) \;=\; U(\max\{R - w,\, 0\}\,;\,\lambda) + U(\min\{R,\, w\}\,;\,\lambda)\,. \tag{2}$$

The right-hand side of (2) is the sum, for an agent of type λ, of the utility from owning a $T2$ firm and the utility from working in one. By direct calculation, these utilities are, respectively:

$$U(\max\{R - w,\, 0\}\,;\,\lambda) = \begin{cases} p^{\lambda} M + \left(1 - p^{\lambda}\right) m - w & \text{if } w \le m \\ p^{\lambda}(M - w) & \text{if } w > m \end{cases} \tag{3}$$

[25] Since $M > m > 0$ by assumption, if $w \geq \frac{1}{2}(M + m)$, the worker receives *all* output when m is realised and *more than half* of the output when M is realised, so no agent would rationally choose to own a $T2$ firm at such a wage.

[26] Proposition 3.2 allows that any of these three groups may be of measure zero, since it does not exclude $\tilde{\lambda}(w) = 0$, $\tilde{\lambda}(w) = \hat{\lambda}(w)$ or $\hat{\lambda}(w) = \frac{1}{2}$.

[27] Two functions $F, G : \Theta \to \mathbb{R}$ are *comonotone* provided

$$\left[F(\theta) - F(\theta')\right]\left[G(\theta) - G(\theta')\right] \;\geq\; 0$$

for all $\theta, \theta' \in \Theta$. Observe that the expected return of each participant in a $T2$ firm is weakly increasing in expected total output.

and

$$U(\min\{R, w\}; \lambda) = \begin{cases} w & \text{if } w \leq m \\ p^{\lambda} w + \left(1 - p^{\lambda}\right) m & \text{if } w > m \end{cases} \tag{4}$$

For given w, both functions are linear in λ. The former is also strictly increasing in λ, while the latter is non-decreasing. From equation (2), the *average* of these two functions is equal to $\frac{1}{2}U(R; \lambda)$, so $\frac{1}{2}U(R; \lambda)$ is strictly increasing in λ.

The difference between (3) and (4) is

$$U(\max\{R - w, 0\}; \lambda) - U(\min\{R, w\}; \lambda) = \begin{cases} m - 2w + p^{\lambda}(M - m) & \text{if } w \leq m \\ -m + p^{\lambda}[(M + m) - 2w] & \text{if } w > m \end{cases} \tag{5}$$

This difference is strictly increasing in λ, since p^{λ} is strictly increasing in λ ($\Delta p > 0$) and

$$w < \frac{1}{2}(M + m).$$

Hence, for each $w < \frac{1}{2}(M + m)$, there exists a *unique* real number $a(w)$ (not necessarily in $\left[0, \frac{1}{2}\right]$) such that

$$U(\max\{R - w, 0\}; a(w)) = U(\min\{R, w\}; a(w)) = \frac{1}{2}U(R; a(w)).$$

For any $z > a(w)$,

$$U(\max\{R - w, 0\}; z) > U(\min\{R, w\}; z)$$

and for any $z < a(w)$,

$$U(\max\{R - w, 0\}; z) < U(\min\{R, w\}; z).$$

Finally, consider the piecewise linear function

$$G_w(z) = \max\{U(\max\{R - w, 0\}; z), U(\min\{R, w\}; z)\}.$$

This gives the maximum return available from $T2$ occupations. It is strictly increasing above $a(w)$ and weakly increasing below it. Figure 3 illustrates.

To complete the proof of (i), we compare K with $G_w(z)$. There are two cases to consider.

Case I: $K \leq w \leq m$. In this case, $G_w(z) = w \geq K$ when $z \leq a(w)$, and $G_w(z) > w \geq K$ otherwise: wage w is not exposed to default, and no type λ strictly prefers to be employed in the $T1$ sector. See Figure 4.

Therefore, $\tilde{\lambda}(w) = 0$ and $\hat{\lambda}(w) = \max\left\{0, \min\left\{a(w), \frac{1}{2}\right\}\right\}$. If $w = K$, then all types with $\lambda < \hat{\lambda}(w)$ are indifferent between working in a $T2$ firm and an occupation in $T1$; otherwise, such types have a strict preference for working in a $T2$ firm.

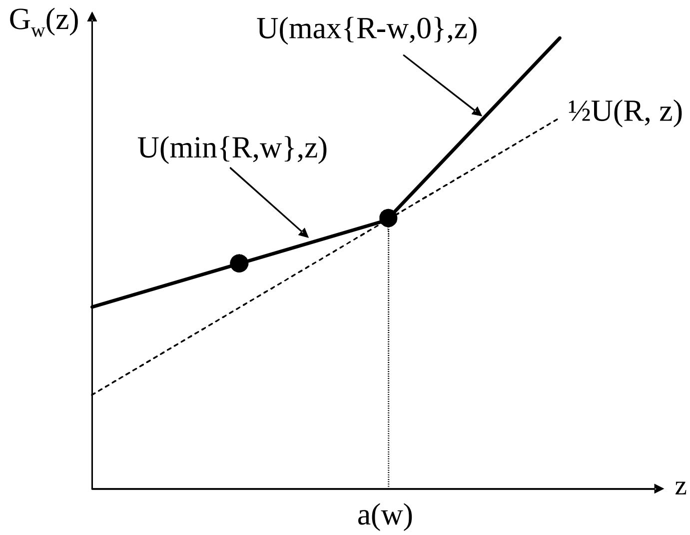


Figure 3: The piecewise linear function $G_w(z)$

Case II: $w > m$ **or** $w < K$**.** In this case, $G_w(z)$ is either strictly increasing or else $w < K$. It follows that there exists a *unique* $b(w)$ such that $K = G_w(b(w))$. Moreover, $G_w(z) > K$ for $z > b(w)$ and $G_w(z) < K$ for $z < b(w)$. Figure 5 illustrates a scenario with $w > m$.

If $b(w) > a(w)$, then $\hat{\lambda}(w) = \tilde{\lambda}(w) = \max\left\{0, \min\left\{b(w), \frac{1}{2}\right\}\right\}$, while if $b(w) \leq a(w)$ we have $\hat{\lambda}(w) = \max\left\{0, \min\left\{a(w), \frac{1}{2}\right\}\right\}$ and $\tilde{\lambda}(w) = \min\left\{\max\{0, b(w)\}, \frac{1}{2}\right\}$.

This concludes the proof of (i).

We now sketch the proof of part (ii), relegating details to Appendix B. From (3) and (4) we observe that

$$w > w' \quad \Rightarrow \quad U(\min\{R, w\}; z) > U\left(\min\left\{R, w'\right\}; z\right) \text{ for all } z \tag{6}$$

and

$$w > w' \quad \Rightarrow \quad a(w) > a\left(w'\right) \tag{7}$$

It follows that an increase in w will strictly and continuously reduce net excess demand in the $T2$ labour market.[28] Since excess demand is clearly positive when $w = 0$ and negative as $w \to \frac{1}{2}(M + m)$, an equilibrium can always be struck. The uniqueness properties follow from (6) and (7). Figure 6 illustrates an equilibrium. □

Figure 6 is reminiscent of the analysis in Kihlstrom and Laffont (1979). Khilstrom and Laffont obtain a division between owners and workers based on relative risk tolerance, rather than

[28] Excess labour demand can be multi-valued, so these statements are somewhat loose, but can be made precise without altering their spirit.

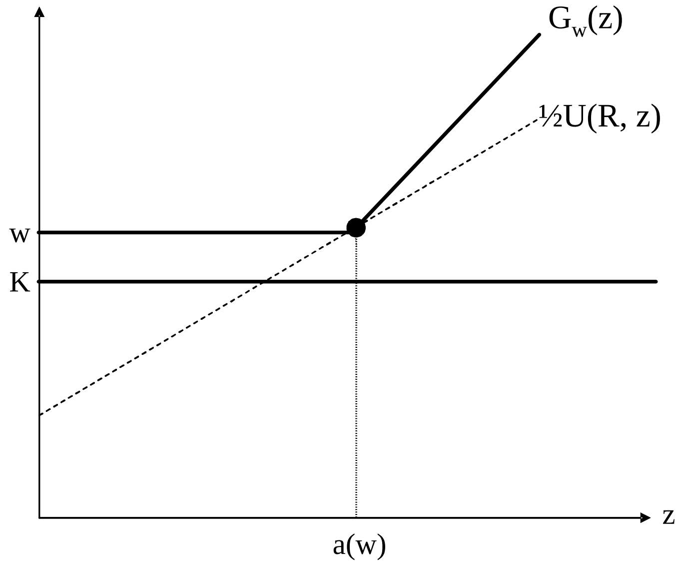


Figure 4: Case I with $w > K$

ambiguity tolerance.[29] However, there are also important points of difference between Kihlstrom and Laffont's model and the one presented here. First, Kihlstrom and Laffont exogenously assume that workers are paid a fixed wage. The form of our wage contracts is endogenously determined by the Pareto efficiency condition on output sharing. Fixed wages are not, in general, Pareto efficient ways to share risk in Kihlstrom and Laffont's model.

Second, Kihlstrom and Laffont assume owners cannot default on wages – they cannot promise more than they can pay in all contingencies, and must dip into private wealth to meet the wage bill if necessary. We assume limited liability. When studying innovation, as we do next, it is important to acknowledge that workers in a new start-up also face uncertainty. If the enterprise fails, they will not receive the remuneration they were promised.

3.3 Robustness

Much of the foregoing analysis makes use of the facts that Θ contains only two states and the $T2$ technology can realise only two output levels (M or m). The reader may reasonably wonder whether our results are robust to relaxation of these assumptions. Appendices C–E take up these issues.

If Θ contains more than two states, it may no longer be true that agents with $\lambda = \frac{1}{2}$ are ambiguity neutral, nor that preferences are convex for $\lambda \leq \frac{1}{2}$. However, as we show in Appendix

[29] In a similar vein, Banerjee and Newman (1993) generate a pattern of occupation choice based on wealth. Agents with wealth above a certain level w^{**} employ agents with wealth below a certain level $w^{*} < w^{**}$. Agents with wealth between w^{*} and w^{**} become self-employed.

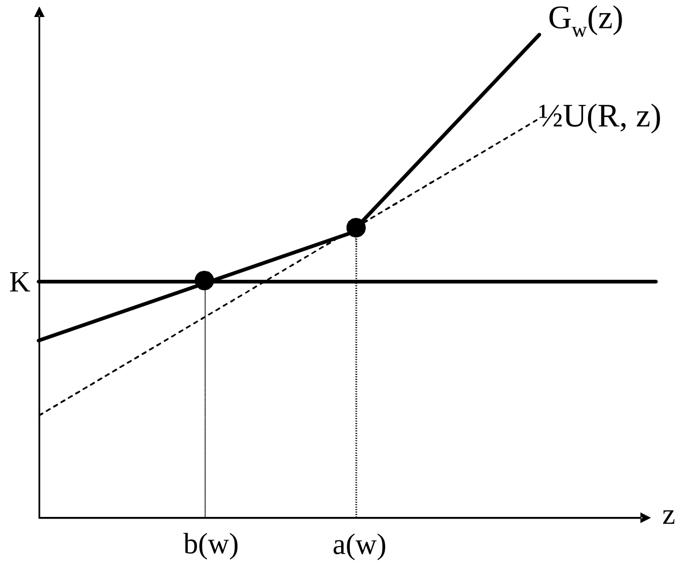


Figure 5: Case II with $w > m$

C, there always exists some $\overline{\lambda} > 0$ such that preferences are convex when $\lambda \leq \overline{\lambda}$. We also illustrate a class of belief functions for which $\overline{\lambda} = \frac{1}{2}$ and agents with $\lambda = \frac{1}{2}$ are ambiguity-neutral.

Appendix D examines the extent to which Proposition 3.1 generalises when Θ contains $n \geq 2$ elements and the $T2$ technology may realise $J \geq 2$ different output levels. While a fully general characterisation is beyond the scope of the paper, we show that if it is feasible to deliver the utility of less ambiguity tolerant agent through a fixed wage (i.e., if her utility is no greater than the lowest output level that can be realised by the $T2$ technology), then it is Pareto efficient to do so. Even if it is not feasible to pay a fixed wage, it will always by Pareto efficient to employ a *comonotone* sharing rule that brings the less ambiguity tolerant agent's remuneration as close to her certainty equivalent as possible. Appendix D also describes a class of situations where it can be shown to be Pareto efficient to pay a the less ambiguity tolerant agent a fixed wage, subject to the limited liability constraint (i.e., there is some w such that the less ambiguity tolerant agent receives the minimum of w and realised output).[30]

Finally, Appendix E discusses the existence and structure of equilibrium in the n state, J output case. Provided K is no greater than the smallest possible output realisation for technology $T2$, results are essentially unaffected. For K larger than the minimum possible output realisation for $T2$, matters are more complicated. However, these are also the cases for which the nature of optimal sharing contracts is imperfectly understood, since the reservation utility of a $T2$ agent

[30] Lemma 3 in Rigotti and Shannon (2006) shows how the Second Welfare Theorem can be used to characterise Pareto optimal allocations between ambiguity-averse consumers. It produces the desired outcome when their setup is specialised to linear utility. Bose, Ozdenoren and Pape (2006) show that full insurance auctions can be optimal if the seller is more ambiguity averse than the buyers.

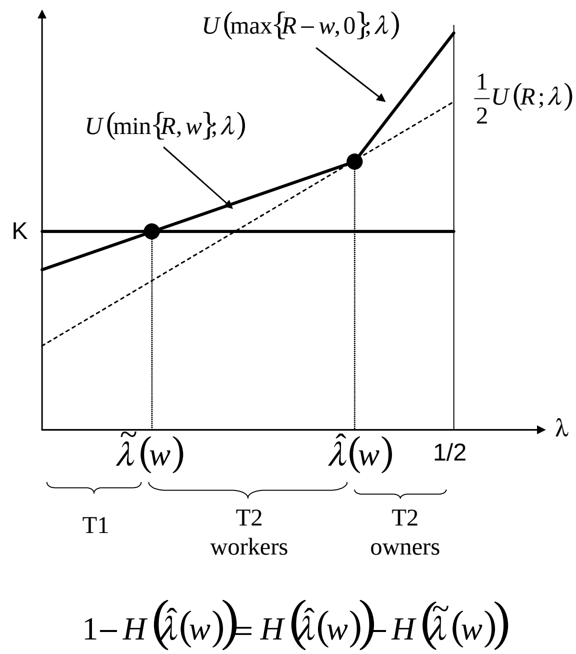


Figure 6: Equilibrium

will be at least K. As argued in Appendix E, for highly innovative industries, quite complex remuneration schemes are common and it is an important avenue for further research to study the dynamics of innovation for such industries. The present model, based as it is on a competitive, price-taking labour market, is designed to help us understand less drastic innovations within a more settled industry. It is therefore appropriate to confine attention to cases where worker remuneration is exposed to relatively low risk of default so that it is reasonable to assume single-price labour contracts.

4 Ambiguity and innovation

Over time, agents will learn more about the ambiguous technology $T2$. Realised output from $T2$ firms provides information with which agents can update their beliefs about θ. We asume that this information is public. Once updating takes place, the equilibrium values of w and $\mu\left[\phi^{-1}(1)\right]$ may change: "good news" is likely to attract new entrants to the $T2$ sector, and conversely for "bad news". By tracking the dynamics of this process, we are able to trace out a diffusion profile for the innovative technology. We shall do so both for the case of a successful and an unsuccessful innovation.

In our model there are no entry costs, technology choices are freely reversible, all information is publicly available, and individuals cannot save (the good is assumed to be perishable). We therefore do not have any dynamic trade-offs of the sort considered in Jovanovic (1982), Mookehrjee and Ray (1991), Sjöstrom (1991), Vettas (1998), or Banerjee and Newman (1993). There is only respect in which decisions may be temporally dependent in our model. If, in the current period's equilibrium, $\mu\left[\phi^{-1}(1)\right] = 1$ so there is measure zero of $T2$ firms, then agents may not have an opportunity to learn about $T2$. In this case, some types may have incentives to "experiment" with $T2$, even though they anticipate an expected return less than K by doing so, as this will generate information that will alter the future values of equilibrium variables.

Since dynamic decision models are problematic outside the subjective expected utility paradigm, we shall neutralise this potential source of temporal dependence as well. There are two strategies for doing so, both inspired by Jensen (1982). First, in the case of a "successful" innovation, we are interested in the growth rate of the $T2$ sector and expect that, on average, it will grow monotonically. In section 4.2.1 we track its growth along an "averaged" sample path that always embodies information precisely consistent with the true state of the world. Along such a path, the mass of $T2$ firms does indeed grow monotonically, so agents never have to make decisions in the problematic scenarios in which the $T2$ sector has "died". The second strategy is to assume that information about $T2$ is revealed in every period, whether or not $\mu\left[\phi^{-1}(1)\right] < 1$. One might imagine, for example, the existence a zero measure of agents with exogenous preference for experimentation, or a laboratory that publishes the results of tests on the new technolgy each period.

With all temporal dependencies thereby neutralised, the diffusion paths are described using a comparative static exercise in which we simulate the arrival of information, and track the consequent impact on equilibrium from belief updating. Since agents begin with a *set* of priors, the nature of this inference problem requires further discussion.

4.1 Updating Π

A decision-maker who implements an act f receives, in state θ, an outcome according to lottery $f(\theta) \in \mathcal{L}$. We assume that decision-makers do not observe the lottery itself, only the realised outcome. From a statistical point of view, each act may be thought of as a *model* – a parametric class of distributions – with common parameter space Θ. A decision-maker who implements the act f runs a statistical *experiment* whose outcome may be used to update the prior set Π.

Unless Π is a singleton, this is a non-standard inference problem, and there does not exist a consensus in the statistics or decision theory literatures about the most suitable updating rule. Two inference procedures have received most attention: *Dempster's rule of combination* and the *robust Bayesian* approach. We shall use the former in what follows, so we give a brief discussion of its merits and demerits here. See Wasserman (1990) for a more detailed exposition. Dempster's rule of combination is tailored to sets Π that may be derived from a belief function $\underline{v}$. It combines the prior information $\underline{v}$ with a *likelihood-based belief function* that encapsulates the new information produced by the experiment.

Let (S, Σ, p) be the measure space and $\Gamma : S \twoheadrightarrow \Theta$ the information correspondence giving rise to the prior belief function $\underline{v}$. Suppose action f is taken and outcome $y \in \mathcal{R}_+$ is realised. Define the likelihood function

$$l(\theta) \;=\; g^f(y \,|\, \theta),$$

where $g^f(\cdot \,|\, \theta)$ is the density function associated with the lottery $f(\theta)$. Then Dempster's rule gives rise to the following posterior belief function, denoted $\underline{v}_y^f$: for each $E \subseteq \Theta$

$$\underline{v}_y^f(E) \;=\; 1 \;-\; \left[\frac{\int_S \max_{\theta\in\Gamma(s)\cap E^c} l(\theta)\; dp}{\int_S \max_{\theta\in\Gamma(s)} l(\theta)\; dp}\right] \tag{8}$$

Suppose, for example, that $\Theta = \{\theta_1, \theta_2\}$ with $0 \le \theta_1 < \theta_2 \le 1$. Let v_i denote the lower probability of θ_i, and let v_i^y denote the updated value for v_i given a single realisation $y \in \{M, m\}$ of the random variable $R(\theta)$. The updating rule (8) implies the following posterior values:

$$v_1^M \;=\; \frac{v_1\theta_1}{v_1\theta_1 + (1-v_1)\,\theta_2} \;\le\; v_1$$

$$v_2^M \;=\; \frac{v_2\theta_2 + (1-v_1-v_2)(\theta_2-\theta_1)}{v_1\theta_1 + (1-v_1)\,\theta_2} \;\ge\; v_2$$

$$v_1^m \;=\; \frac{v_1(1-\theta_1) + (1-v_1-v_2)(\theta_2-\theta_1)}{v_2(1-\theta_2) + (1-v_2)(1-\theta_1)} \;\ge\; v_1$$

$$v_2^m \;=\; \frac{v_2(1-\theta_2)}{v_2(1-\theta_2) + (1-v_2)(1-\theta_1)} \;\le\; v_2.$$

One easily verifies that

$$1 - (v_1^y + v_2^y) \;\le\; 1 - (v_1 + v_2)$$

for each $y \in \{M, m\}$. Hence, $\Delta p = [1-(v_1+v_2)](\theta_2-\theta_1)$ shrinks (weakly) under updating: new information reduces ambiguity.

The key features of this updating process are the following: "good news" (realisation M) weakly increases both $\underline{p}$ and $\overline{p}$, "bad news" (realisation m) does the opposite; and any news weakly decreases Δp. In other words, the whole interval $[\underline{p}, \overline{p}]$ shifts up or down in the obvious direction, and also shrinks in size. Information therefore has two effects in our model: it brings agents' beliefs closer together, and causes convergence of beliefs towards the truth. The following example provides a simple numerical illustration of this updating process. Note how the effect of the two early "bad" draws has a larger impact on the more ambiguity tolerant individuals.

Example II. Suppose $M = 1$ and $m = 0.01$. Consider initial values $v_1 = 0.1$ and $v_2 = 0.1$. We also take $\theta_1 = 0.2$ and $\theta_2 = 0.7$, and assume θ_1 is the "true" state. With these values we have

$$\underline{p} \;=\; v_2\theta_2 \;+\; (1-v_2)\,\theta_1 \;=\; 0.25$$

and

$$\overline{p} \;=\; (1-v_1)\,\theta_2 \;+\; v_1\theta_1 \;=\; 0.65.$$

Figure 7 plots a sequence of random draws from $R(\theta_1)$, indicated by the bars. The path of p^λ for $\lambda \in \{0, 0.5, 1\}$ is also illustrated. The value of p^λ at time t reflects information up to $t-1$, where updating follows Dempster's rule. Figure 7 also has a line at 0.2 to indicate the true value of θ.

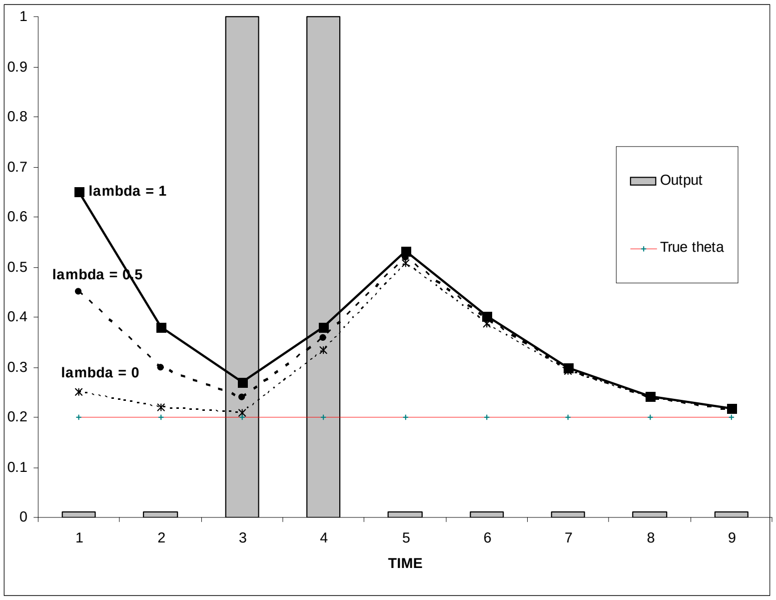


Figure 7: Updating with Dempster's rule

The alternative to Dempster's rule, the robust Bayesian approach, involves updating every $\pi \in \Pi$ using Bayes' Rule. Wasserman (1990, Section 3.2) provides some axiomatic support for Dempster's rule, but its primary attraction is computational simplicity. In general, there do not exist straightforward means of calculating upper and lower posterior expectations when using the robust Bayesian approach. If the set of priors is the core of a belief function, the upper and lower posterior probabilities of events are readily determined (Wasserman, 1990b, Theorem 4.1), but the posterior set may not be *m-closed* (i.e., may not coincide with the set of measures dominating its lower probability – see Wasserman and Kadane, 1990, p.1333). In particular, it may not be the core of any belief function. This means that knowledge of the lower and upper probabilities is not sufficient for the computation of upper and lower posterior expectations. While Jaffray (1994, Proposition 1) does provide an ingenious method for computing posterior upper and lower expectations when the prior is the core of a belief function, his method is based on a fixed-point algorithm, and the lack of an explicit updating formula greatly reduces the usefulness of robust Bayesian updating for dynamic analysis.

Finally, the robust Bayesian approach has qualitatively similar properties to Dempster's procedure, its main distinguishing characteristic being that it resolves ambiguity more slowly (Wasserman, 1990, Theorem 7). In terms of the results presented below, the only difference we would expect from using robust Bayesian inference is a slowing of the belief dynamics. The nature of the dynamics (as opposed to the speed of adjustment) is dictated by the process of equilibration in the labour market and the empirical distribution of ambiguity tolerance in the population.

4.2 Good innovations

Suppose that our model economy is observed over several discrete periods, indexed by $t \in \{1, 2, ...\}$. If (w_t, ϕ_t) is the equilibrium in period t, let $\hat{\lambda}_t \equiv \hat{\lambda}(w_t)$ and define $\tilde{\lambda}_t$ to be the unique solution to

$$H\left(\tilde{\lambda}_t\right) = \mu\left[\phi_t^{-1}(1)\right].$$

Thus, $\hat{\lambda}_t$ is the marginal type between those who own $T2$ firms and those who work in them; while $\tilde{\lambda}_t$ is the marginal type between those working in the $T1$ sector and those working in $T2$ firms. In particular, we assume without loss of generality that ϕ_t assigns each type $\lambda \in \left[0, \tilde{\lambda}_t\right]$ to the $T1$ sector. Thus, we may use the triple $\left(\hat{\lambda}_t, \tilde{\lambda}_t, w_t\right)$ to characterise the period t equilibrium.

For each period t (regardless of whether the $T2$ sector is active), an independent draw is made from the random variable $R(\theta)$ and *all* firms using technology $T2$ produce a quantity of output equal to this value. In other words, the random factors affecting $T2$ production are common across firms. Therefore, each period generates *one* observation from the random variable $R(\theta)$. It is assumed that everyone can observe this draw so the data is public. All agents therefore update their beliefs in the same way, and we obtain new equilibrium values $\left(\hat{\lambda}_{t+1}, \tilde{\lambda}_{t+1}, w_{t+1}\right)$. By tracking the value of $1 - H\left(\hat{\lambda}_t\right)$ over time, we trace out a dynamic path for the size of the $T2$ sector (i.e., the mass of firms using technology $T2$). This allows us to assess the diffusion profile of the innovative technology.

We employ a two-state version of our model, with $\Theta = \{\theta_1, \theta_2\}$ and $0 < \theta_1 < \theta_2 < 1$ such that

$$\theta_2 M + (1 - \theta_2) m > 2K > \theta_1 M + (1 - \theta_1) m.$$

Thus, $T2$ is superior in state θ_2, and $T1$ is superior in state θ_1. We first consider the case in which θ_2 is the true state, and hence the diffusion profile of a "good" innovation. If $K \leq m$ we can determine this diffusion profile analytically, at least for a "averaged" sample path. When $K > m$ we shall proceed using numerical simulations.

4.2.1 The case $K \leq m$

In this case, we have an explicit expression for $\hat{\lambda}$:

Proposition 4.1 *Suppose $\Delta p > 0$ and $K \leq m$. If $\delta \in \left(0, \frac{1}{2}\right)$ in equilibrium, then $w = K$ and*

$$\hat{\lambda} = \frac{2K - m - \underline{p}(M - m)}{(M - m)\Delta p} \tag{9}$$

Conversely, if (9) is strictly between the median of the λ distribution and $\frac{1}{2}$, then $\delta \in \left(0, \frac{1}{2}\right)$ in equilibrium.

Proof. If $\delta \in \left(0, \frac{1}{2}\right)$ in equilibrium then both sectors operate. Everyone in the $T1$ sector receives utility K, so the marginal $T2$ worker (that is, the one with type $\tilde{\lambda}$) must derive utility K from working in a $T2$ firm. Since $K \leq m$, this means $w = K$. Therefore, from equation (5), $\hat{\lambda}$ satisfies

$$m \; - \; 2K \; + \; \left(\underline{p} + \hat{\lambda}\Delta p\right)(M - m) \; = \; 0 \tag{10}$$

from which (9) follows.

Conversely, if (9) is strictly between the median of the λ distribution and $\frac{1}{2}$, then a fraction strictly between zero and one half of the population derives utility above K from owning a $T2$ firm paying wage $w = K$. We may therefore construct an equilibrium with $w = K$ and $\tilde{\lambda}$ satisfying

$$1 - H\left(\hat{\lambda}\right) \; = \; H\left(\hat{\lambda}\right) - H\left(\tilde{\lambda}\right)$$

with $\hat{\lambda}$ equal to (9). Note that $\tilde{\lambda} \in \left(0, \hat{\lambda}\right)$. Since equilibrium is unique (Proposition 3.2(ii)) the result is proved. $\square$

This result is very convenient, as it not only provides a simple formula for the computation of $\hat{\lambda}$, but also reveals that equilibrium wages in the $T2$ sector are unaffected by new information (at least until the demise of the $T1$ sector). The only parts of expression (9) that vary over the course of the diffusion process are

$$\underline{p} \; = \; \theta_1 + v_2 \Delta\theta$$

and

$$\Delta p \; = \; (1 - v_1 - v_2)\,\Delta\theta$$

since v_1 and v_2 will evolve over time. In particular:

$$\Delta_t \hat{\lambda}_t \; = \; \frac{\hat{\lambda}_t \Delta_t v_{1,t} - \left(1 - \hat{\lambda}_t\right)\Delta_t v_{2,t}}{(1 - v_{1,t} - v_{2,t})} \tag{11}$$

where $v_{k,t}$ is the value of v_k at time t and "Δ_t" indicates time differencing.

Let $\delta_t = 1 - H\left(\hat{\lambda}_t\right)$ denote the mass of $T2$ firms. Then

$$\Delta_t \delta_t \; = \; -h\left(\hat{\lambda}_t\right)\Delta_t \hat{\lambda}_t$$

We wish to examine the sign of $\Delta_t^2 \delta_t \equiv \Delta_t\left(\Delta_t \delta_t\right)$ along the diffusion path for a successful innovation (i.e., along which $\Delta_t \hat{\lambda}_t < 0$). Since

$$\Delta_t^2 \delta_t \; = \; -\left\{h'\left(\hat{\lambda}_t\right)\left(\Delta_t \hat{\lambda}_t\right)^2 + h\left(\hat{\lambda}_t\right)\Delta_t^2 \hat{\lambda}_t\right\}$$

an S-shaped diffusion profile requires

$$\frac{h'\left(\hat{\lambda}_t\right)}{h\left(\hat{\lambda}_t\right)} \; < \; \frac{-\Delta_t^2 \hat{\lambda}_t}{\left(\Delta_t \hat{\lambda}_t\right)^2}$$

initially and

$$\frac{h'\left(\hat{\lambda}_t\right)}{h\left(\hat{\lambda}_t\right)} \;>\; \frac{-\Delta_t^2\hat{\lambda}_t}{\left(\Delta_t\hat{\lambda}_t\right)^2}$$

subsequently. For example, with a *Uniform* density for λ, this requires that $\Delta_t^2\hat{\lambda}_t$ is initially negative and subsequently positive.

To approximate the average $\hat{\lambda}_t$ path, we shall adopt the approach of Jensen (1982) and evaluate $\hat{\lambda}_t$ along a single, artificially constructed "sample path", for which the proportion of M observations is always equal to the true θ value. We also assume initial parameter values such that $\delta_1 \in \left(0, \frac{1}{2}\right)$. As we will see, $\Delta\delta_t \geq 0$ along the constructed path, so the $T2$ sector grows monotonically. Let σ_t denote the number of "successes" (M realisations) observed up to time t inclusive. Taking θ_2 to be the true parameter value (i.e., a superior innovation), we assume that

$$\sigma_t \;=\; t\theta_2 \qquad t = 1, 2, ...$$

Of course, this will prevent σ_t from being integer-valued, so we shall treat the updating formulae to follow as if they are valid for all $\sigma_t \in [0, t]$.

Writing v_1 and v_2 for $v_{1,1}$ and $v_{2,1}$ respectively, the updating rule (8) implies

$$v_{1,t} \;=\; \frac{v_1 f_t\left(\theta_1\right)}{\left(1 - v_1\right) f_t\left(\theta_2\right) + v_1 f_t\left(\theta_1\right)}$$

and

$$v_{2,t} \;=\; 1 \;-\; \frac{\left(1 - v_2\right) f_t\left(\theta_1\right)}{\left(1 - v_1\right) f_t\left(\theta_2\right) + v_1 f_t\left(\theta_1\right)}$$

where

$$f_t\left(\theta\right) \;=\; \theta^{\sigma_t}\left(1 - \theta\right)^{t - \sigma_t} \;=\; \theta^{t\theta_2}\left(1 - \theta\right)^{t\left(1 - \theta_2\right)}$$

is the likelihood of observing $t\theta_2$ successes in t trials when θ is the true parameter value. In particular, $f_t\left(\theta\right)$ is maximised at $\theta = \theta_2$. Furthermore:

$$\frac{f_t\left(\theta_1\right)}{f_t\left(\theta_2\right)} \;=\; \left[\frac{\theta_1^{\theta_2}\left(1 - \theta_1\right)^{\left(1 - \theta_2\right)}}{\theta_2^{\theta_2}\left(1 - \theta_2\right)^{\left(1 - \theta_2\right)}}\right]^t \;<\; 1$$

and declines monotonically to zero as $t \to \infty$, so $\lim_{t\to\infty} v_{2,t} = 1$ as one would hope. We also observe that

$$1 - v_{1,t} - v_{2,t} \;=\; \left[\frac{f_t\left(\theta_1\right)}{\left(1 - v_1\right) f_t\left(\theta_2\right) + v_1 f_t\left(\theta_1\right)}\right]\left(1 - v_1 - v_2\right)$$

is monotonically declining in t.

Proposition 4.2 *When $v_{1,t}$ and $v_{2,t}$ are specified as above, $\Delta_t\hat{\lambda}_t < 0$ and $\Delta_t^2\hat{\lambda}_t < 0$.*

Proof. By direct calculation:

$$\Delta_t v_{2,t} = \frac{(1-v_1)(1-v_2)\left[f_t(\theta_1)\Delta_t f_t(\theta_2) - f_t(\theta_2)\Delta_t f_t(\theta_1)\right]}{\left[(1-v_1) f_t(\theta_2) + v_1 f_t(\theta_1)\right]^2}$$

Since

$$\Delta_t f_t(\theta) = C(\theta) f_t(\theta),$$

where $C(\theta) = \ln\left[\theta^{\theta_2}(1-\theta)^{(1-\theta_2)}\right]$,

$$\frac{\Delta_t v_{2,t}}{(1-v_{1,t}-v_{2,t})} = \left\{\frac{(1-v_1)(1-v_2)\left[C(\theta_2)-C(\theta_1)\right]}{(1-v_1-v_2)}\right\}\frac{f_t(\theta_2)}{(1-v_1) f_t(\theta_2) + v_1 f_t(\theta_1)} > 0,$$

where the inequality uses $C(\theta_2) > C(\theta_1)$. Similarly:

$$\frac{\Delta_t v_{1,t}}{(1-v_{1,t}-v_{2,t})} = \left\{\frac{v_1(1-v_1)\left[C(\theta_1)-C(\theta_2)\right]}{(1-v_1-v_2)}\right\}\frac{f_t(\theta_2)}{(1-v_1) f_t(\theta_2) + v_1 f_t(\theta_1)} < 0.$$

Thus:

$$\begin{aligned}
\Delta_t \hat{\lambda}_t &= \frac{\hat{\lambda}_t \Delta_t v_{1,t}}{(1-v_{1,t}-v_{2,t})} - \frac{\left(1-\hat{\lambda}_t\right)\Delta_t v_{2,t}}{(1-v_{1,t}-v_{2,t})} \\
&= \left\{\frac{(1-v_1)\left[C(\theta_1)-C(\theta_2)\right]}{(1-v_1-v_2)}\right\}\left[\frac{f_t(\theta_2)}{(1-v_1) f_t(\theta_2) + v_1 f_t(\theta_1)}\right]\left[v_1\hat{\lambda}_t + (1-v_2)\left(1-\hat{\lambda}_t\right)\right] \\
&< 0
\end{aligned}$$

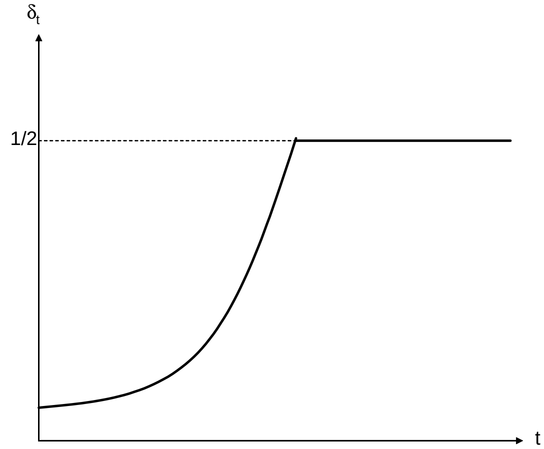
which proves the first part of the Proposition. It is also clear that both square-bracketed terms are positive and increasing with t (recall that $1-v_2 > v_1$), which proves the second part. □

Proposition 4.2 implies that, "on average", $\Delta_t^2\hat{\lambda}_t < 0$ while ambiguity persists (i.e., while $1-v_{1,t}-v_{2,t} > 0$), implying a convex diffusion profile when λ is Uniformly distributed. Of course, since $\theta_2 M + (1-\theta_2)m > 2K$ by assumption, the $T1$ sector will be abandonned before $1-v_{1,t}-v_{2,t}$ hits zero (consider (10) as $\underline{p}$ approaches θ_2), so the actual diffusion path will have the shape illustrated in Figure 4.2.1.

If, instead, h is unimodal, the negatively sloped portion to the right of the mode will accentuate the convex portion of the diffusion profile, as

$$\frac{h'\left(\hat{\lambda}_t\right)}{h\left(\hat{\lambda}_t\right)} < 0 < \frac{-\Delta_t^2\hat{\lambda}_t}{\left(\Delta_t\hat{\lambda}_t\right)^2}$$

when $h' < 0$. Conversely, the positively sloped portion to the left of the mode may introduce some concavity into the latter stages of the diffusion profile. In other words, moving from a Uniform to a unimodal distribution will give an even more pronounced S-shape, and tend to smooth the non-differentiability as δ_t hits $\frac{1}{2}$.



4.2.2 The case $K > m$

Matters are now more complicated, as w_t and $\hat{\lambda}_t$ are no longer independently determined. The relationship between accumulated information and $\hat{\lambda}_t$ is too complex for analytic results to be feasible, so we proceed numerically.

Suppose "good news" – realisation M – is observed at t. This increases p^λ (the probability that $R(\theta) = M$ as assessed by a λ type agent) for every λ, so it is intuitive that $\hat{\lambda}_{t+1}$ and $\tilde{\lambda}_{t+1}$ will be *lower* than their values in the preceding period: the $T2$ sector grows.[31] If news is "good on average", then the size of the $T2$ sector tends to increase over time. The time profile of this diffusion depends on two things: the updating process and the form of H.

Our model is consistent with an S-shaped diffusion profile, but a Uniform H will no longer suffice.[32] Instead, we need some right skewness to obtain the necessary convexity in the early phases of the diffusion profile.

Consider Figure 8, which depicts the density function h associated with a particular distribution function H. Suppose $\tilde{\lambda}$ and $\hat{\lambda}$ start at $\tilde{\lambda}_1$ and $\hat{\lambda}_1$ respectively, and the true θ is such that $T2$ has a higher expected return than $T1$. Since the "average" news received will be "good" (or rather, better than the average expectation), $\tilde{\lambda}_t$ and $\hat{\lambda}_t$ will tend leftwards over time. In the early phases of diffusion, as $\tilde{\lambda}_t$ travels down the thin right-hand tail of h, the process of adoption is slow. Few agents are optimistic enough to own or work in such an uncertain enterprise. However,

[31] Suppose $\tilde{\lambda}_{t+1} \geq \tilde{\lambda}_t$, which means there are weakly fewer $T2$ firms in existence at $t+1$. Since p^λ has increased, both $T2$ occupations are more attractive (at the original w_t) than they were previously. For $\tilde{\lambda}_{t+1} \geq \tilde{\lambda}_t$ we must therefore have $w_{t+1} < w_t$. But an increase in p^λ and a fall in w will increase the density of types for whom $T2$ ownership is optimal – see (5). This is inconsistent with a fall in the density of $T2$ firms in equilibrium. Therefore, $\tilde{\lambda}_{t+1} < \tilde{\lambda}_t$, and hence $\hat{\lambda}_{t+1} < \hat{\lambda}_t$ as well.

[32] This produced a concave diffusion profile. These results are not reported here, but are available on request.

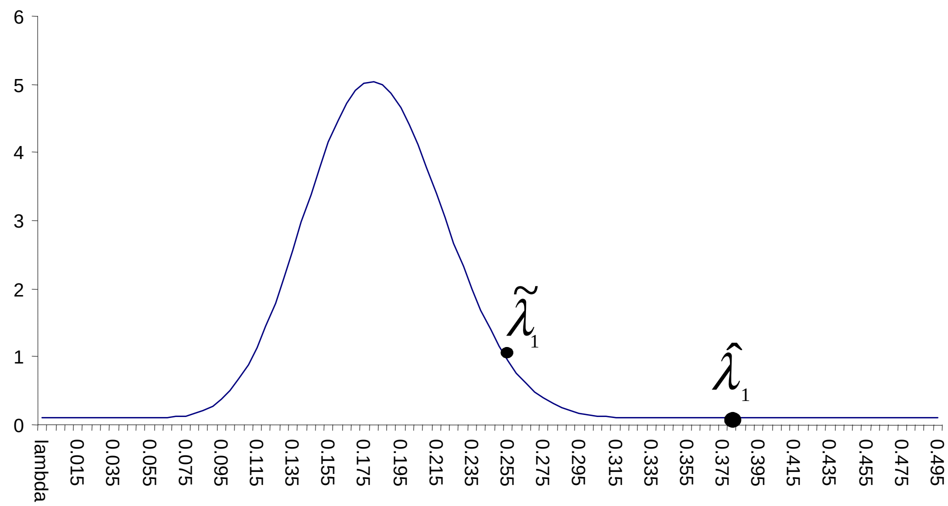


Figure 8: Right Skewed Distribution

once $\tilde{\lambda}_t$ reaches the steeply sloping portion of h, labour floods into the $T2$ sector and the pace of diffusion will increase. Hence, during this phase, many workers quit $T2$ firms to establish their own start-ups, and are replaced by labour migrating from the $T1$ sector. Finally, when $\tilde{\lambda}_t$ reaches the peak of the density function, diffusion will slow. This process generates an S-shaped diffusion path.

The results reported below use $M = 70$, $m = 10$, $\theta_1 = 0.45$, $\theta_2 = 0.5$ and $K = 19.15$.[33] We also assume that Π at $t = 1$ is such that $v_1 = v_2 = 0.1$, so there is initially substantial ambiguity.

Suppose that h is the PDF of a random variable X such that $2X$ follows a Beta(a, b) distribution. By varying the values of a and b we obtain a wide range of PDF's supported on $\left[0, \frac{1}{2}\right]$. For example, when $a = b = 1$, h is Uniform on $\left[0, \frac{1}{2}\right]$. If $a > b$, then h is left-skewed, and conversely if $a < b$. For the purposes of the computations reported below, we introduce a third parameter, c, and let

$$h \ = \ (1 - c)\, h' \ + \ c,$$

where h' is the PDF of a random variable X with $2X$ distributed as Beta(a, b). The parameter c determines a vertical shift of the PDF h', with a compensating scaling so that h integrates to unity. This allows us to add enough weight under the tails to avoid convergence problems with our numerical solutions.

[33] Hence

$$\theta_2 M + (1 - \theta_2)\, m \ > \ 2K \ > \ \theta_1 M + (1 - \theta_1)\, m$$

as required.

To simulate the diffusion of a successful innovation, we generate $n = 100$ runs of $T = 25$ periods of independent draws from the random variable $R(\theta_2)$. For each run, we calculate the equilibrium values $\left(\hat{\lambda}_t, \tilde{\lambda}_t, w_t\right)$ at each t, using Dempster's rule to update beliefs after each observation on $R(\theta_2)$. Thus, each run generates a path for $1 - H\left(\hat{\lambda}_t\right)$, the mass of $T2$ firms in the economy at t.[34] We report the average path, averaging over all runs for which $1 - H\left(\hat{\lambda}_T\right) > 0$. That is, we report averages conditional on the "success" of the innovation, as indicated by its survival through to period T. This is important, since the innovative technology is permanently eliminated in many runs, despite its superiority. Empirical evidence on the diffusion of "successful" innovations likewise interprets "success" to mean that the technology has been widely adopted by the end of the sample period. Betamax may have been superior to VHS, but it would not be included in a sample of "successful" technologies.

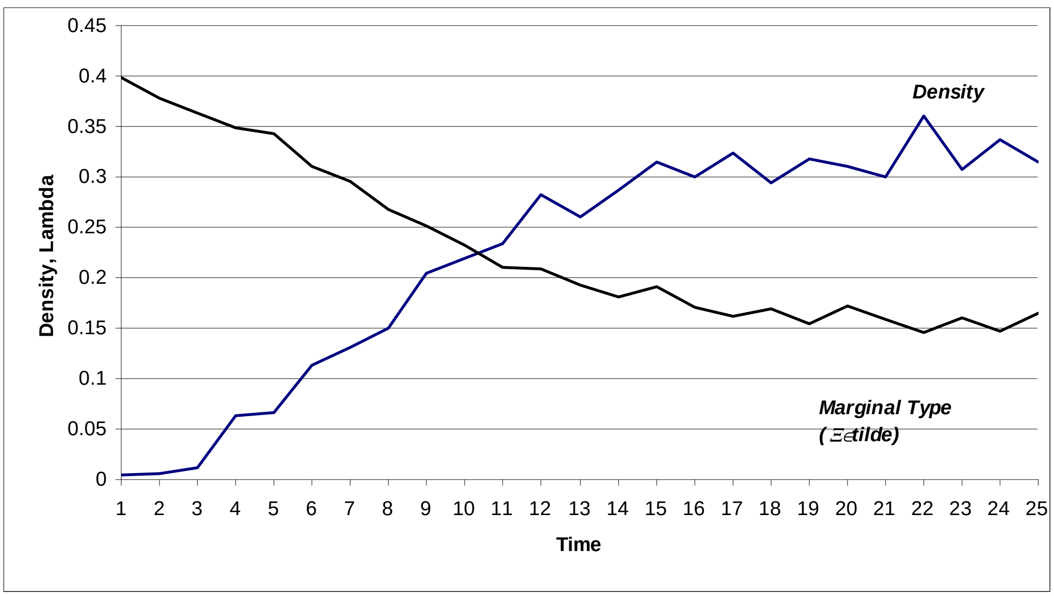


Figure 9: Diffusion with Right Skewed Distribution

We now consider a right skewed distribution. In Figure 9 we plot the conditional average diffusion path, when the distribution of types is given in Figure 8.[35] This diffusion profile has a pronounced S-shape. On the same graph, we also plot the conditional average of $\tilde{\lambda}_t$.

The density profile follows three distinct phases. In the first phase ($1 \leq t \leq 3$) diffusion is extremely slow. We can understand this with reference to the long right-hand tail on the distribution function in Figure 8, and the fact that in these three periods, $\widetilde{\lambda} > 0.35$. The scarcity

[34] Note that it lies in the interval $\left[0, \frac{1}{2}\right]$.

[35] In this case, $a = 25$, $b = 15$ and $c = 0.1$.

of types who are sufficiently ambiguity tolerant to be owners and workers in the innovative industry retards the initial adoption of the new product. As time proceeds, ambiguity is lessened and $\widetilde{\lambda}$ falls. There is then a flood of entrants as $\widetilde{\lambda}_t$ moves up the steep part of the distribution ($3 < t < 12$). In the third phase ($t \geq 12$) diffusion is again slow as the flow of new entrants is not as rapid.

4.3 Bad Innovations

We next repeat the exercise, with $a = 25$, $b = 15$, $c = 0.1$, $T = 50$ and all other parameter values as in the preceding section, but using the random variable $R(\theta_1)$ to generate our output sequences. This allows us to examine the adoption path of an inferior technology. Figure 10 presents the results for two cases.[36] The *Ambiguous Information* case is the usual one: the probability interval for θ_2 is initially set at $[0.1,\ 0.9]$, types are distributed according to H and behave as per the model. For this case, we observe a boom-bust phenomenon.[37] The *Precise Information* case assumes that all agents begin with a *unique* prior probability of θ_2 equal to 0.5 (the mid-point of the interval) and update using Bayes' Rule. In this case, all agents start in $T2$ and depart as more information arrives.[38]

What can explain the increasing level of adoption early on in the Ambiguous Information case, despite the preponderance of "bad news"? Figure 11 provides some insight into this puzzle. It plots the average values for p^0 and $p^{\frac{1}{2}}$ in each period. Recall that these give the probability of $R(\theta) = M$ as assessed by the most ambiguity averse ($\lambda = 0$) and most ambiguity tolerant ($\lambda = \frac{1}{2}$) type respectively. Because the news is "bad on average" (relative to expectations), the average value of $p^{\frac{1}{2}} = \frac{1}{2}\underline{p} + \frac{1}{2}\overline{p}$ – the mid-point of the interval $\left[\underline{p},\ \overline{p}\right]$ – tracks gradually downwards. However, in the early phases of the updating process, the ambiguity-reducing effect dominates the path of $p^0 = \underline{p}$, which rises to meet $p^{\frac{1}{2}}$, before following the latter downwards.

Owners of $T2$ firms, who are the most optimistic group in the economy and hence whose beliefs follow a path similar to that of $p^{\frac{1}{2}}$, slowly reduce their output expectations as "bad news" accumulates. However, their more pessimistic workers – whose beliefs are tied more closely to $\underline{p}$ – are emboldened by the reduction in ambiguity to accept lower wages, as the default risk appears to loom less large in their minds. If the former effect (on owners) is sufficiently dominated by the latter effect (on workers), there is an overall boost to the $T2$ sector. However, once the the ambiguity is flushed out of the system, workers' and owners' beliefs both tend south and $T2$ begins its inevitable decline.

[36] Figure 10 gives *unconditional* averages, since inferior technologies are eventually abandonned with probability one.

[37] Only the start of the bust is shown in Figure 10. The downward path continues in subsequent periods, but we wish to focus on the more surprising up-swing early on, so we have truncated the Figure at $t = 50$.

[38] Of course, starting from a prior that places all agents in $T2$ precludes the possibility of an initial upswing in diffusion. However, with a unique prior, all agents must be allocated to the *same* industry in equilibrium, unless each industry has an identical expected return. This prevents us from producing a smooth path starting from an intermediate mass of $T2$ firms. However, the downward trend in Figure 10 is quite general in the sense that, on average, the posterior of every agent will be more pessmistic than the prior after every observation. This will make $T1$ occupations relatively more attractive over time.

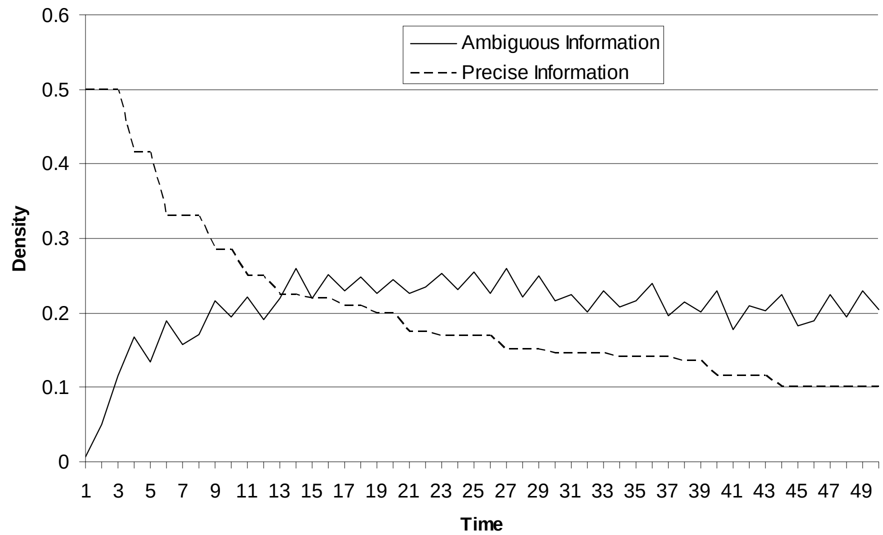


Figure 10: Diffusion of a Bad Innovation with Right Skewed Distribution

To see what relative optimism does to the boom-crash profile, we simulated the model again using $a = 15$, $b = 25$ and $c = 0.1$. This corresponds to a left-skewed density function which is the mirror image of that in Figure 8. Such an economy may be said to exhibit greater "entrepreneurial spirit". This results in a more pronounced initial boom in take-up of the bad innovation.[39] The average annual *per capita* income over the 50 periods for the more entrepreneurial (left-skewed) economy is \$18.86, with a variance of 0.41. For the less entrepreneurial economy, the average income is \$19.03 with a variance of 0.29. In the steady state, where all firms use technology $T1$, average income would be K (= \$19.15). Herein lies a cautionary tale. Innovators embrace ambiguity, and they do not always back the right horses. The optimal degree of innovation in an economy may be a complex matter to judge, but there is no reason to expect that more entrepreneurs is necessarily better.

5 Concluding remarks

Innovation by its very nature is an activity where the odds of success are unknown and the ability to tolerate ambiguity is what marks out the entrepreneurial personality. We have shown how this situation can be modelled by introducing ambiguity in a simple framework. A set of probabilities is consistent with the available information, and individuals evaluate prospects by calculating a weighted average of the highest and lowest expected return with respect to probabilities in this

[39] Details of these results are available on request.

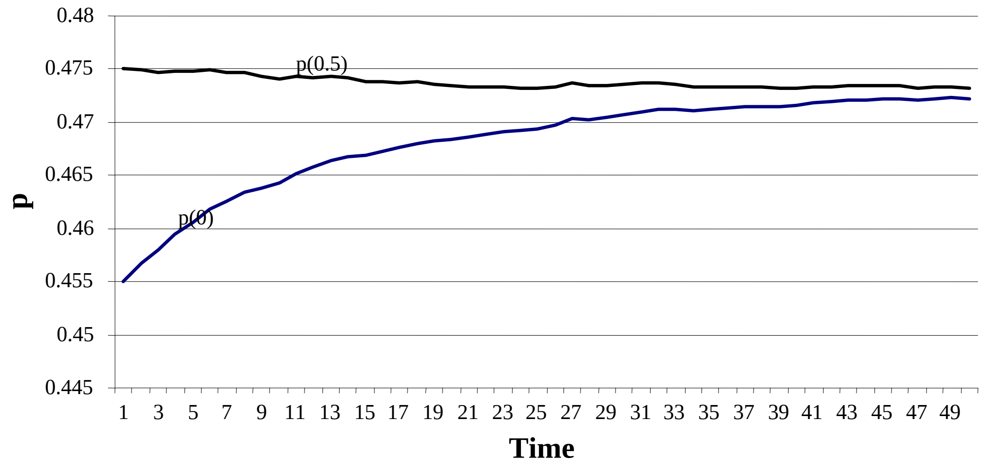


Figure 11: Evolution of probabilities

set. Ambiguity tolerance is measured by the relative weight placed on the highest expectation.

In a model of occupational choice we can show that Pareto efficient sharing rules require the more ambiguity tolerant agent to pay the other a fixed wage (subject to limited liability). Therefore, ambiguity tolerance makes the worker and owner roles within a firm endogenous. Furthermore, in the equilibrium of our economy, the most ambiguity tolerant agents own innovative firms, an intermediate group supplies labour to the innovative firms, while the least ambiguity tolerant enter the non-ambiguous sector.

We also provide a unique slant on the question of the interplay of entrepreneurial propensities and innovation in a general equilibrium model where occupational choice and innovation are endogenous. This allows us to understand how diffusion paths are tied to entrepreneurial propensities. As information about the true productivity of the innovative technology is revealed ambiguity diminishes and this determines the diffusion profile of the innovation.

When the innovation is good, the standard S-shaped diffusion profile is generated, at least for a right-skewed distribution of types. When the innovation is bad, however, booms followed by crashes are possible. The boom happens because, for ambiguity-averse agents, "any news is good news" since it reduces ambiguity and may support initial entry into the sector. A crash will eventually happen as the truth about the technology is revealed. The first part of this process could not occur if information was precise. Therefore, ambiguity is helps understand a feature of diffusion paths that eludes existing models.

APPENDIX A

Proof of Proposition 3.1. Consider an agent of type λ who receives $s_M \in [0, M]$ when revenue M is realised, and $s_m \in [0, m]$ otherwise. Denote by s the state-contingent lottery corresponding to this sharing rule: $s(\theta)$ is a lottery delivering s_M with probability θ and s_m with probability $1 - \theta$. If $\lambda \leq \frac{1}{2}$, then one easily verifies that[40]

$$U(s;\,\lambda) \;=\; \min_{p \in \left[p^{\lambda}, q^{\lambda}\right]} p s_M \;+\; (1-p)\, s_m \tag{12}$$

where

$$p^{\lambda} \;=\; \lambda \overline{p} + (1-\lambda)\, \underline{p}$$

and

$$q^{\lambda} \;=\; \lambda \underline{p} + (1-\lambda)\, \overline{p}.$$

Observe that (12) is quasi-concave in (s_M, s_m). Since p^{λ} (respectively, q^{λ}) is increasing (respectively, decreasing) in λ, we also note that $p^{\lambda''} > p^{\lambda'}$ and $q^{\lambda''} < q^{\lambda'}$ when $\overline{p} > \underline{p}$ and $\lambda' < \lambda'' \leq \frac{1}{2}$.

We first prove necessity. Let $U\left(s';\,\lambda'\right) = u'$, so u' is the certainty equivalent of s' for type λ'. Suppose, contrary to what we wish to show, it is *not* the case that $s'_m = \min\{s'_M,\, m\}$. Then *either* $s'_m > s'_M$ *or* $s'_m < \min\{m, s'_M\}$. Hence, $u' < \max\{s'_m, s'_M\}$ and

$$\alpha\left(u',\, u'\right) \;+\; (1-\alpha)\left(s'_M,\, s'_m\right) \;\leq\; (M,\, m) \tag{13}$$

for $\alpha \in (0, 1)$ sufficiently close to zero. In other words, it is possible to move the λ'-type's contract marginally in the direction of her certainty equivalent without violating feasibility. We shall show that this yields a Pareto improvement.

Choose $\alpha \in (0, 1)$ small enough to satisfy (13), and define

$$\left(s_M^{\alpha'},\, s_m^{\alpha'}\right) \;=\; \alpha\left(u',\, u'\right) \;+\; (1-\alpha)\left(s'_M,\, s'_m\right)$$

$$\left(s_M^{\alpha''},\, s_m^{\alpha''}\right) \;=\; (M,\, m) \;-\; \left(s_M^{\alpha'},\, s_m^{\alpha'}\right).$$

Recalling that that $p^{\lambda''} > p^{\lambda'}$ and $q^{\lambda''} < q^{\lambda'}$, we have

$$U\left(s^{\alpha'};\;\lambda'\right) \;=\; \min_{p \in \left[p^{\lambda'}, q^{\lambda'}\right]} p s_M^{\alpha'} \;+\; (1-p)\, s_m^{\alpha'} \;=\; u'$$

[40]If $\lambda > \frac{1}{2}$, then $p^{\lambda} \geq q^{\lambda}$ (with equality iff $\overline{p} = \underline{p}$) and

$$U(s;\,\lambda) \;=\; \max_{p \in \left[q^{\lambda}, p^{\lambda}\right]} p s_M \;+\; (1-p)\, s_m.$$

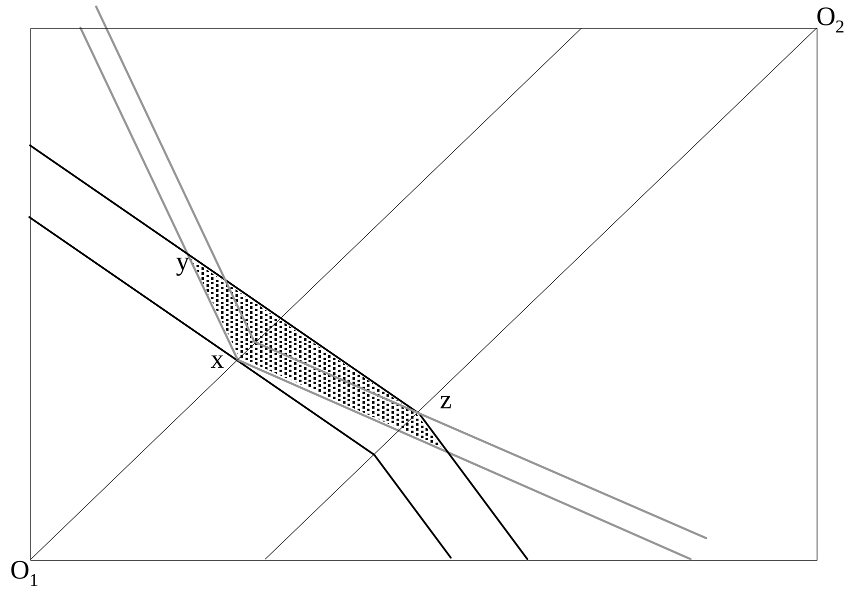


Figure 12: Pareto efficiency in an Edgeworth Box

and (Aubin, 1998, Proposition 4.4)

$$\begin{aligned}\lim_{\alpha\downarrow 0}\frac{U\left(s^{\alpha''};\ \lambda''\right)\ -\ U\left(s'';\ \lambda''\right)}{\alpha} &= \min_{p\in\left[p^{\lambda''},q^{\lambda''}\right]} ps_M^{\alpha'}\ +\ (1-p)\,s_m^{\alpha'}\ -\ u' \\ &> \min_{p\in\left[p^{\lambda'},q^{\lambda'}\right]} ps_M^{\alpha'}\ +\ (1-p)\,s_m^{\alpha'}\ -\ u' \\ &= 0.\end{aligned}$$

Therefore, when $\alpha \in (0,1)$ is small enough, $\left(s_M^{\alpha'},\, s_m^{\alpha'}\right)$ and $\left(s_M^{\alpha''},\, s_m^{\alpha''}\right)$ Pareto improve on the original revenue sharing contracts. This proves the necessity part of the Proposition.

To see the sufficiency part, consider Figure 12. Here we have two agents with differing degrees of ambiguity tolerance. The grey indifference curves represent the relatively more ambiguity averse individual; the black indifference curves represent the relatively more ambiguity tolerant. Since no two individuals can have the same indifference curves, the picture represents a generic situation. One can easily see that a point like y cannot be Pareto optimal, since there is a large area (shaded) of allocations which are preferred by both individuals. This reasoning also applies to a point like z, which is situated on the certainty line of individual 2. Therefore, all that is left are points on the certainty line of individual 1. At a point like x, in fact, there are no mutually profitable trades available.

APPENDIX B

Proof of Proposition 3.2(ii). We construct an excess labour demand correspondence $\Gamma : \left[0,\, \frac{1}{2}(M+m)\right] \twoheadrightarrow [-1,1]$ for the $T2$ labor market, and confirm that $0 \in \Gamma(w^*)$ for some

$$w^* \in \left[0,\, \frac{1}{2}(M+m)\right].$$

Define Γ as follows:

$$\Gamma(w) \;=\; \begin{cases} \left\{1 - 2H\left(\hat{\lambda}(w)\right) + H\left(\tilde{\lambda}(w)\right)\right\} & \text{if } U\left(\min\{R, w\}\,;\, \hat{\lambda}(w)\right) \;>\; K \\ \left\{1 - 2H\left(\hat{\lambda}(w)\right) + H(\lambda) \,\middle|\, \lambda \in \left[\tilde{\lambda}(w)\,,\, \hat{\lambda}(w)\right]\right\} & \text{otherwise} \end{cases}.$$

Standard arguments confirm that Γ has closed graph, and that

$$\Gamma^{-1}(z) \;=\; \left\{w \in \left[0,\, \frac{1}{2}(M+m)\right] \,\middle|\, z \in \Gamma(w)\right\}$$

is convex. Applying von Neumann's Intersection Lemma (Border, 1985, p.75) to the graph of Γ and the set

$$\left[0,\, \frac{1}{2}(M+m)\right] \times \{0\},$$

it follows that there exists a $w^* \in \left[0,\, \frac{1}{2}(M+m)\right]$ such that $0 \in \Gamma(w^*)$. Using the monotonicity of H, we deduce that there also exists a unique $\tilde{\lambda}^* \in \left[\tilde{\lambda}(w^*)\,,\, \hat{\lambda}(w^*)\right]$ such that there is an equilibrium (w^*, ϕ^*) with $\phi^*(\lambda) = \{O\}$ for all $\lambda \in \left[\hat{\lambda}(w^*)\,,\, \frac{1}{2}\right]$, $\phi^*(\lambda) = \{W\}$ for all $\lambda \in \left(\tilde{\lambda}^*,\, \hat{\lambda}(w^*)\right)$, and $\phi^*(\lambda) = \{1\}$ for all other λ.

It remains to prove the uniqueness part of the Proposition. Let $\left(w^*, \hat{\lambda}^* = \hat{\lambda}(w^*)\,, \tilde{\lambda}^*\right)$ describe an equilibrium as above, and assume it is *not* the case that $\hat{\lambda}^* = \tilde{\lambda}^* = \frac{1}{2}$ (i.e., assume that a non-zero density of $T2$ firms operate in equilibrium). It suffices to show that there is no other w satisfying $0 \in \Gamma(w)$. Since $T2$ firms operate in the equilibrium $\left(w^*, \hat{\lambda}^*, \tilde{\lambda}^*\right)$, we have $\tilde{\lambda}^* < \hat{\lambda}^* < \frac{1}{2}$. But the utility difference (5) is strictly decreasing in w and strictly increasing in λ, so any change to the wage rate must upset labour market equilibrium in the $T2$ sector.

APPENDIX C

There are two approaches one may take to ensure robustness of our basic model to the case in which the cardinality of Θ, denoted $|\Theta|$, is strictly larger than 2. One is to restrict the class of belief functions so that, for any value of $|\Theta|$, preferences are ambiguity neutral when $\lambda = \frac{1}{2}$ and convex for $\lambda \leq \frac{1}{2}$. The other is to impose a tighter bound on the range of acceptable λ values. We shall consider these two approaches in reverse order.

Let $\Pi \subseteq \Delta(\Theta)$ denote the core of a belief function over the set $\Theta = \{\theta_1, ..., \theta_n\}$. The following facts are well-known (see, for example, Shapley, 1971): there exist $\{\pi_\rho \mid \rho \in P\}$, where P is the set of permutations of Θ (i.e., the set of one-to-one mappings $\rho : \{1, ..., n\} \to \Theta$), such that

$$\Pi \;=\; \text{co}(\{\pi_\rho \mid \rho \in P\}),$$

where co(A) denotes the convex hull of A. Furthermore, defining

$$F^\rho \;=\; \{f : \Theta \to \mathcal{R} \mid f(\rho(1)) \leq \cdots \leq f(\rho(n))\}$$

for each $\rho \in P$,

$$\min_{\pi \in \Pi} \pi \cdot f \;=\; \pi_\rho \cdot f$$

for every $f \in F^\rho$, where $\pi \cdot f = \sum_{\theta \in \Theta} \pi_\rho(\theta) f(\theta)$. Given $\rho \in P$, we let $\rho^* \in P$ denote the "reverse" permutation: $\rho^*(i) = \rho(n - i + 1)$ for each $i \in \{1, ..., n\}$. It follows that, for any $f \in F^\rho$,

$$U(f;\ \lambda) \;= [\lambda \pi_{\rho^*} + (1 - \lambda)\pi_\rho] \cdot f$$

Proposition 5.1 *Let*

$$\Pi^\lambda \;=\; \text{co}(\{\lambda \pi_{\rho^*} + (1 - \lambda)\pi_\rho \mid \rho \in P\}).$$

There exists $\overline{\lambda} > 0$ *such that, for any* $\lambda \leq \overline{\lambda}$,

$$U(f;\ \lambda) \;=\; \min_{\pi \in \Pi^\lambda} \pi \cdot f$$

for any f.

Proof. We need to show that

$$[\lambda \pi_{\rho^*} + (1 - \lambda)\pi_\rho] \cdot f \;\leq\; \left[\lambda \pi_{\hat{\rho}^*} + (1 - \lambda)\pi_{\hat{\rho}}\right] \cdot f \qquad (14)$$

for every $\rho \in P$, every $f \in F^\rho$ and every $\hat{\rho} \in P$. Since (14) is trivial when $f \equiv 0$, and since scaling f by some strictly positive constant will not affect its membership of F^ρ or inequality (14), it suffices to restrict attention to functions in

$$D \;=\; \left\{ f : \Theta \to \mathcal{R} \;\middle|\; \sum_{\theta \in \Theta} |f(\theta)| = 1 \right\}.$$

Indeed, as D is polyhedral, it is enough to consider the extreme points of D, $\text{ext}(D)$. That is, we must verify (14) for every $\rho \in P$, every $f \in F^\rho \cap \text{ext}(D)$ and every $\hat{\rho} \in P$. Thus, we have finitely many inequalities of the form (14) to verify.

Note that (14) may be written

$$A(\rho, \hat{\rho}, f) \; + \; B(\rho, \hat{\rho}, f)\, \lambda \; \leq \; 0$$

with

$$A(\rho, \hat{\rho}, f) \; = \; (\pi_\rho - \pi_{\hat{\rho}}) \cdot f$$

and

$$B(\rho, \hat{\rho}, f) \; = \; (\pi_{\rho^*} - \pi_\rho) \cdot f \; - \; \left(\pi_{\hat{\rho}^*} - \pi_{\hat{\rho}}\right) \cdot f.$$

When $f \in F^\rho$, $A(\rho, \hat{\rho}, f) \leq 0$ and $B(\rho, \hat{\rho}, f) \geq 0$, with $B(\rho, \hat{\rho}, f) = 0$ iff $A(\rho, \hat{\rho}, f) = 0$. Thus,

$$\overline{\lambda} \; = \; \min_{\substack{\rho, \hat{\rho} \in P \\ f \in F^\rho \cap \text{ext}(D)}} \left\{ \frac{|A(\rho, \hat{\rho}, f)|}{B(\rho, \hat{\rho}, f)} \,\middle|\, B(\rho, \hat{\rho}, f) > 0 \right\} \; > \; 0$$

does the needful. □

Proposition 5.1 implies that an Arrow-Hurwicz decision-maker with sufficiently low ambiguity tolerance acts as a maxmin expected utility decision-maker with respect to the set Π^λ of probabilities (see Gilboa and Schmeidler, 1989).[41] Such preferences are convex (*ibid.*, Axiom A.5). If we restrict the range of λ values in the economy to $\left[0, \overline{\lambda}\right]$ this will guarantee that all agents have convex preferences (when the cardinality of Θ equals 2 $\overline{\lambda}$ equals $\frac{1}{2}$).

The second approach to obtaining convex preferences is to restrict the class of allowable belief functions so that an agent with $\lambda = \frac{1}{2}$ acts in an ambiguity-neutral fashion. This can be done as follows.

Definition 2 *A set*

$$\Pi \; = \; \text{co}(\{\pi_\rho \mid \rho \in P\}) \subseteq \Delta(\Theta)$$

is called barycentrically symmetric *if it has the following properties:*
(i) the vertices $\{\pi_\rho \mid \rho \in P\}$ *are either all distinct or all identical;*
(ii) for each $\rho \in P$*,*

$$\min_{\pi \in \Pi} \pi \cdot f \; = \; \pi_\rho \cdot f$$

for every $f \in F^\rho$*;*
(iii) for each $\rho \in P$*, the barycentre of* $\text{co}(\{\pi_\rho, \pi_{\rho^*}\})$ *coincides with the barycentre of* Π*.*

Note that Π' is barycentrically symmetric if and only if $\Pi' = \Pi + \{z\}$ for some barycentrically symmetric Π with barycentre $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ and some $z \in \mathcal{R}^n$ with $\sum_{i=1}^n z_i = 0$.

[41] In fact, it is straightforward to show that Π^λ is *m-closed* – it contains all the probabilities that dominate the *lower probability* defined for all $E \subseteq \Pi^\lambda$ as follows: $\underline{p}(E) = \inf_{\pi \in \Pi^\lambda} \pi(E)$ – when $\lambda \leq \overline{\lambda}$, so these preferences also conform to the *Choquet expected utility* model (Schmeidler, 1989).

Definition 3 *A belief function is barycentrically symmetric if it has a barycentrically symmetric core.*

Obviously, *any* belief function is barycentrically symmetric when $n = 2$. Figure 13 illustrates a barycentrically symmetric set Π for $n = 3$.

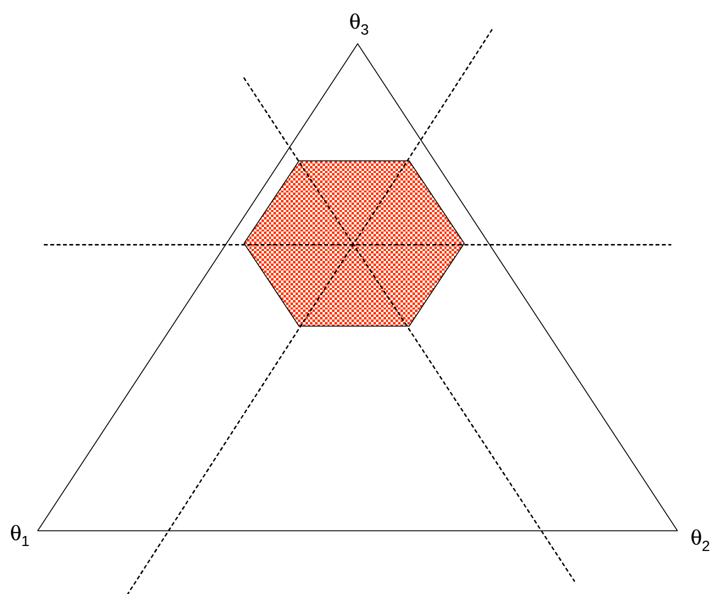


Figure 13: A barycentrically symmetric set for $n = 3$

This class of belief functions and corresponding sets of probabilities has not, to the best of our knowledge, been studied elsewhere. It is related to – though distinct from – two other well-known classes: the *ε-contaminated priors* familiar from robust Bayesian statistics (Berger, 1996), and the *symmetric, coherent capacities* of Wasserman and Kadane (1992, 1996). As with the latter class of belief functions, baricentric symmetry may be interpreted as generalising the notion of a *uniform prior*.

For any barycentrically symmetric set $\Pi = \mathrm{co}(\{\pi_\rho \mid \rho \in P\})$ with barycentre b:

$$\frac{1}{2}\left[\max_{\pi\in\Pi} \pi \cdot f\right] \; + \; \frac{1}{2}\left[\min_{\pi\in\Pi} \pi \cdot f\right] \;\; = \;\; \left[\frac{1}{2}\pi_{\rho^*} + \frac{1}{2}\pi_\rho\right] \cdot f \;\; = \;\; b \cdot f$$

for any $\rho \in P$ and any $f \in F^\rho$. In other words,

$$U\left(f;\; \frac{1}{2}\right) \;\; = \;\; b \cdot f,$$

so an agent with $\lambda = \frac{1}{2}$ is ambiguity neutral. It is also easy to see that for any $\lambda \leq \frac{1}{2}$ and any f:

$$U(f;\ \lambda) \ = \ \min_{\pi \in \Pi^\lambda} \ \pi \cdot f$$

where

$$\Pi^\lambda \ = \ \mathrm{co}(\{\lambda \pi_{\rho^*} + (1-\lambda)\, \pi_\rho \mid \rho \in P\}).$$

Therefore, preferences are convex for any $\lambda \in \left[0, \frac{1}{2}\right]$. Figure 14 illustrates when $n = 3$.

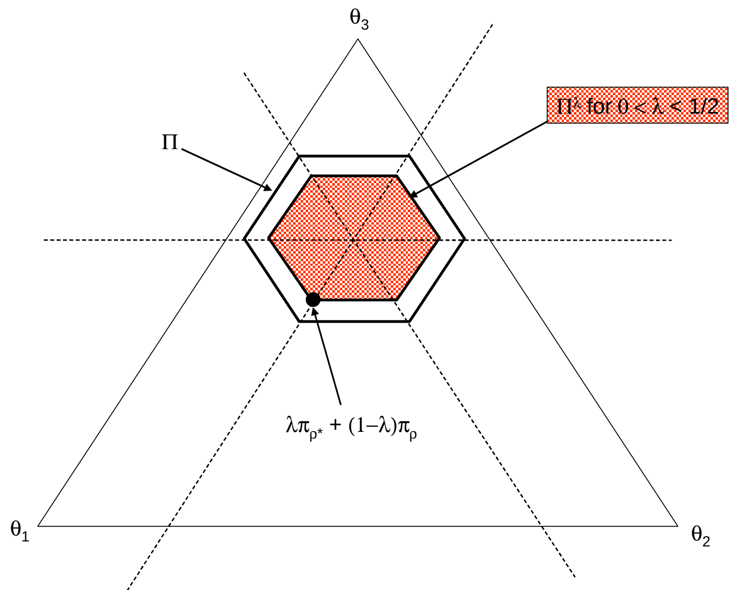


Figure 14: Permutation ρ in the Figure is the identity.

We shall now provide a characterisation of a class of barycentrically symmetric belief functions for $n \geq 3$. Suppose that $\Pi = \mathrm{co}(\{\pi_\rho \mid \rho \in P\})$ is a barycentrically symmetric set with barycentre $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$. Then there exists $\varepsilon \in \left[-\frac{1}{n}, \frac{1}{n}\right]^n$ with

$$\varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_n,$$

and $\varepsilon_i + \varepsilon_{n-i+1} = 0$ for each i,[42] such that

$$\pi_\rho\left(\theta_{\rho(i)}\right) \ = \ \frac{1}{n} + \varepsilon_{n-i+1} \ = \ \frac{1}{n} - \varepsilon_i$$

[42] In particular, if n is odd, then

$$\varepsilon_{\frac{n+1}{2}} = 0.$$

for every $\rho \in P$. Observe that $\pi_\rho \in \Delta(\Theta)$ and

$$\frac{1}{2}\pi_\rho + \frac{1}{2}\pi_{\rho^*} = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$$

for every $\rho \in P$ as required.

The *lower probability*, $\underline{v}$, associated with Π takes the form:

$$\underline{v}(E) = \frac{|E|}{n} + \sum_{i=1}^{|E|} \varepsilon_i.$$

Note that $\underline{v}(E)$ depends only on $|E|$. We therefore define

$$\underline{v}_k = \frac{k}{n} + \sum_{i=1}^{k} \varepsilon_i = \frac{k}{n} - \sum_{i=n-k+1}^{n} \varepsilon_i$$

to be the lower probability of an event of cardinality k.

Is $\underline{v}$ a belief function? To answer this question, we shall construct the *Möbius inverse* σ of $\underline{v}$ and check whether it is non-negative in all components (Shafer, 1976, §2 of Chapter 2 and footnote to p.48). Since $\underline{v}(E)$ depends only on the cardinality of E, so does σ. Let σ_k denote $\sigma(E)$ for any event E of cardinality k. By definition, $\sigma_1 = \underline{v}_1$ and

$$\sigma_k = \underline{v}_k - \sum_{h=1}^{k-1} \binom{k}{h} \sigma_h \tag{15}$$

for $k \in \{2, 3, ..., n\}$, where

$$\binom{k}{h} = \frac{k!}{h!\,(k-h)!}$$

is the number of subsets of size h in a set of size k.

While it is beyond the scope of the present paper to provide a complete characterisation of all ε values yielding non-negative solutions to the recursive system (15), we shall identify a particular subset of these ε values. To this end, let us suppose that there is some constant $\Delta\varepsilon \geq 0$ with

$$\varepsilon_k - \varepsilon_{k-1} = \Delta\varepsilon \qquad k = 2, 3, ..., n \tag{16}$$

and hence

$$\varepsilon_k = \varepsilon_1 + (k-1)\,\Delta\varepsilon \qquad k = 1, 2, ..., n.$$

Since $\varepsilon_1 + (n-1)\,\Delta\varepsilon = |\varepsilon_1|$, we have

$$\Delta\varepsilon = \frac{2\,|\varepsilon_1|}{n-1}.$$

In other words, there is a one-to-one relationship between $\Delta\varepsilon$ and ε_1, so there is precisely one free parameter in the specification of this class of belief functions.

We shall show that, under the additional assumption (16), every solution to the system (15) is non-negative. Note in particular that assumption (16) is *automatically satisfied when* $n = 3$.

From (15) for the cases $k = 1$ and $k = 2$ respectively, we obtain:

$$\sigma_1 \;=\; \frac{1}{n} \;+\; \varepsilon_1 \;\geq\; 0$$

and

$$\sigma_2 \;=\; \Delta\varepsilon \;\geq\; 0.$$

We claim that $\sigma_k = 0$ for all $k \geq 3$. The argument is by induction. Substituting $k = 3$ into (15) gives:

$$\sigma_3 \;=\; \sum_{h=0}^{2} \Delta\varepsilon \;-\left(\begin{array}{c} 3 \\ 2 \end{array}\right)\Delta\varepsilon \;=\; 0.$$

For the inductive step, assume $\sigma_k = 0$ for $k \in \{3, 4, ..., s-1\}$. Then

$$\begin{aligned} \sigma_s &= \underline{v}_s - \sum_{h=1}^{s-1}\left(\begin{array}{c} s \\ h \end{array}\right)\sigma_h \\ &= \underline{v}_s - s\sigma_1 - \left(\begin{array}{c} s \\ 2 \end{array}\right)\sigma_2 \\ &= \left[\sum_{i=0}^{s-1} i \;-\; \frac{s\,(s-1)}{2}\right]\Delta\varepsilon \end{aligned}$$

The square-bracketed term is zero (as an elementary argument by induction will confirm), so $\sigma_s = 0$ and the claim is proved.

To sum up so far, under assumption (16), any barycentrically symmetric set Π with barycentre $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ is the core of a (barycentrically symmetric) belief function. Now suppose $\Pi' = \Pi + \{z\} \subseteq \Delta(\Theta)$ for some barycentrically symmetric Π with barycentre $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ and some $z \in \mathcal{R}^n$ with $\sum_{i=1}^{n} z_i = 0$. We shall show:

Proposition 5.2 *The set Π' is the core of a belief function ($\underline{v}'$) if – and hence, only if – Π is the core of a belief function ($\underline{v}$).*

Proof. Since $\Pi' \subseteq \Delta(\Theta)$ the vector z satisfies:

$$\begin{aligned} \underline{v}'(E) &= \underline{v}_{|E|} + \sum_{i \in E} z_i \;\geq\; 0 \\ \Leftrightarrow \quad \sum_{i\in E} z_i &\geq -\underline{v}_{|E|} \qquad \text{for each } E \subseteq \Theta \end{aligned} \tag{17}$$

Because $\underline{v}$ is a belief function, it is *2-monotone*[43] and hence

$$-\underline{v}_{|E|} \;\leq\; -|E|\,\underline{v}_1.$$

[43] That is:

$$\underline{v}(A \cup B) + \underline{v}(A \cap B) \;\geq\; \underline{v}(A) + \underline{v}(B)$$

for any $A, B \subseteq \Theta$. See Shafer (1976, Theorem 2.1).

Therefore, the condition (17) is equivalent to

$$\min_i \ z_i \ \geq \ -\underline{v}_1 \tag{18}$$

Let σ denote the Möbius inverse of $\underline{v}$ and σ' the Möbius inverse of $\underline{v}'$. Then:

$$\sigma_1' \ = \ \underline{v}_1 \ + \ z_1 \ \geq \ 0.$$

We claim that for any E with $|E| \geq 2$, $\sigma'(E) = \sigma_{|E|}$ and hence $\sigma'(E) \geq 0$. We shall argue by induction on $|E|$. First, observe that when $|E| = 2$:

$$\begin{aligned}
\sigma'(E) &= \underline{v}_2 + \sum_{i \in E} z_i \ - \ \sum_{i \in E} \sigma'(\{i\}) \\
&= \sigma_2
\end{aligned}$$

Next, assume the claim is true for all $|E| < k$ and let $|E| = k$. Then:

$$\begin{aligned}
\sigma'(E) &= \underline{v}'(E) \ - \ \sum_{B \subset E} \sigma_{|B|} \ - \ \sum_{i \in E} z_i \\
&= \left[\sum_{B \subseteq E} \sigma_{|B|} \ + \ \sum_{i \in E} z_i \right] \ - \ \sum_{B \subset E} \sigma_{|B|} \ - \ \sum_{i \in E} z_i \\
&= \sigma_{|E|}.
\end{aligned}$$

This completes the proof. □

Therefore, under assumption (16), any barycentrically symmetric set Π is the core of a (barycentrically symmetric) belief function. In particular, since assumption (16) is automatically satisfied when $n = 3$, any barycentrically symmetric set is the core of a (barycentrically symmetric) belief function in the three-state case.

APPENDIX D

We now consider whether Proposition 3.1 generalises to a set-up with $|\Theta| \geq 2$ and in which a $T2$ firm may realise outputs in the set $\mathcal{M} = \{m_1, m_2, ..., m_J\}$ with

$$0 < m_1 < \cdots < m_J.$$

Define the vector $\mathbf{m} = (m_1, ..., m_J)$ and let $p_j(\theta)$ denote the probability of realisation m_j given $\theta \in \Theta$. Thus, $R(\theta)$ is now interpreted as the random variable that delivers m_j with probability $p_j(\theta)$. Assume that the $p(\theta_s) \in \Delta(\mathcal{M})$ are mutually distinct, with $p(\theta_s)$ *strictly first-order stochastically dominating* $p(\theta_{s'})$ whenever $s > s'$. In other words, the expected output function

$$\overline{R(\theta)} \;=\; \sum_{j=1}^{J} p_j(\theta)\, m_j \;=\; p(\theta) \cdot \mathbf{m}$$

orders states (strictly) from worst (θ_1) to best (θ_n):

$$\overline{R(\theta_1)} < \cdots < \overline{R(\theta_n)}.$$

It is further assumed that

$$\mathrm{co}(\{p(\theta) \mid \theta \in \Theta\}) \cap \mathrm{ri}\,[\Delta(\mathcal{M})] \neq \varnothing \tag{19}$$

where $\mathrm{ri}\,[\Delta(\mathcal{M})]$ denotes the *relative interior* of $\Delta(\mathcal{M})$. This latter assumption is without loss of essential generality – if it is violated, there exists a proper face of the simplex $\Delta(\mathcal{M})$ containing $\{p(\theta) \mid \theta \in \Theta\}$, so we may simply omit the "redundant" output level(s).

Suppose $s = (s_1, ..., s_J)$ is a vector of output-contingent payments satisfying $s_j \in [0, m_j]$ for each j. Restricting attention throughout to barycentrically symmetric belief functions,[44] an agent of type $\lambda \in \left[0, \frac{1}{2}\right]$ evaluates s as follows:

$$U(s;\lambda) \;=\; \min_{\pi \in \Pi^{\lambda}} \sum_{\theta \in \Theta} \pi(\theta) \left[\sum_{j=1}^{J} p_j(\theta)\, s_j \right]$$

with $\Pi^{\lambda'} \subseteq \Pi^{\lambda}$ if $\lambda < \lambda' \leq \frac{1}{2}$. Furthermore, we may write

$$U(s;\lambda) \;=\; \min_{q \in \mathcal{P}^{\lambda}} \sum_{j=1}^{J} q_j s_j$$

[44] The following arguments also apply (with minor adjustments) if the belief function is arbitrary and we instead restrict $\lambda \in \left[0, \overline{\lambda}\right]$ as defined in the preceding section.

where

$$\mathcal{P}^{\lambda} = \left\{ \sum_{\theta\in\Theta} \pi(\theta)\, p(\theta) \,\middle|\, \pi \in \Pi^{\lambda} \right\}$$
$$= \operatorname{co}\left\{ \sum_{\theta\in\Theta} \left[\lambda \pi_{\rho^*}(\theta) + (1-\lambda)\,\pi_{\rho}(\theta)\right] p(\theta) \,\middle|\, \rho \in P \right\} \subseteq \Delta(\mathcal{M})$$

Observe that[45]

$$\overline{p} = \sum_{\theta\in\Theta} \pi_{Id^*}(\theta)\, p(\theta)$$

and

$$\underline{p} = \sum_{\theta\in\Theta} \pi_{Id}(\theta)\, p(\theta)$$

are the probability vectors that maximise and minimise (respectively) expected output given Π. Thus, for any s satisfying $s_1 \leq s_2 \leq \cdots \leq s_J$,

$$p^{\lambda} = \lambda \overline{p} + (1-\lambda)\,\underline{p}$$

is the most pessimistic element of $\mathcal{P}^{\lambda}$.

We may now prove:[46]

Proposition 5.3 *Consider two agents with respective ambiguity tolerance parameters* $\lambda' < \frac{1}{2}$ *and* $\lambda'' \in \left(\lambda', \frac{1}{2}\right]$. *Suppose these agents operate a T2 firm, agent* λ' *receiving* $s' \geq 0$ *and agent* λ'' *receiving* $s'' \geq 0$, *with* $s' + s'' = \mathbf{m}$. *If* $U(s';\lambda') = u'$ *and*

$$s' + \alpha\left(u'\mathbf{1} - s'\right) \leq \mathbf{m}$$

(where $\mathbf{1} \in \mathbb{R}^J$ *is a vector of ones) for* $\alpha > 0$ *sufficiently close to zero, then*

$$\lim_{\alpha\downarrow 0} \frac{U\left(s' + \alpha\left(u'\mathbf{1} - s'\right);\lambda'\right) - U\left(s';\lambda'\right)}{\alpha} = 0 \tag{20}$$

and

$$\lim_{\alpha\downarrow 0} \frac{U\left(s'' + \alpha\left(s' - u'\mathbf{1}\right);\lambda''\right) - U\left(s'';\lambda''\right)}{\alpha} \geq 0. \tag{21}$$

To understand the import of Proposition 5.3, consider a Pareto efficient sharing rule that maximises the utility of agent λ'' subject to agent λ' receiving utility at least u'. The Proposition implies that any such sharing rule can be adjusted by moving s' as far as is feasible in the direction of the λ' type's certainty equivalent $u'\mathbf{1}$ without violating Pareto efficiency. This result should feel familiar as it implies that interior Pareto optimal allocations between two ambiguity averse decision makers require that the most ambiguity averse is fully insured (see Rigotti and Shannon, 2006), and it is reminiscent of the optimality of full insurance auctions obtained by Bose, Ozdenoren and Pape (2006). From Proposition 5.3 one immediately deduces:

[45] We use Id to denote the identity permutation.

[46] The following is reminiscent of Proposition 1 in Bose, Ozdenoren and Pape (2006).

Corollary 5.1 *If $u' \leq m_1$ then it is Pareto efficient to pay λ' a fixed "wage" u' in all contingencies. If $J = 2$, then for* any *feasible utility level u', it is Pareto efficient to use a sharing rule of the form $s'_1 = \min\{s'_2, m_1\}$.*

Proof of Proposition 5.3: Equality (20) is obvious, since $U\left(s' + \alpha\left(u'\mathbf{1} - s'\right); \lambda'\right) = u'$ for any $\alpha \in [0, 1]$. To see (21), apply Aubin (1998, Proposition 4.4) and $\mathcal{P}^{\lambda''} \subseteq \mathcal{P}^{\lambda'}$ to obtain:

$$\begin{aligned}
\lim_{\alpha \downarrow 0} \frac{U\left(s'' + \alpha\left(s' - u'\mathbf{1}\right); \lambda''\right) \; - \; U\left(s''; \lambda''\right)}{\alpha} \quad &\geq \quad \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j s'_j \; - \; u' \\
&\geq \quad \min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j s'_j \; - \; u' \\
&= \quad 0.
\end{aligned}$$

□

What if $J > 2$ and $u' > m_1$? Can we still deduce the optimality of a sharing rule with

$$s'_j = \min\left\{s'_J, m_i\right\}$$

for all j? While a complete answer to this question is, to the best of our knowledge, unknown, we can offer the following partial answer:

Proposition 5.4 *Suppose $0 < \lambda' < \lambda'' \leq \frac{1}{2}$. If (s', s'') is a Pareto efficient sharing rule for a T2 firm operated by agents with respective ambiguity tolerances λ' and λ'', then there exists a comonotone Pareto efficient sharing rule that gives the same utility to the agent of type λ'.*[47]

Proposition 5.5 *Suppose $J = 3$ and $0 < \lambda' < \lambda'' \leq \frac{1}{2}$. Suppose further that the set*

$$\mathcal{P}^0 \;=\; \left\{ \sum_{\theta \in \Theta} \pi(\theta)\, p(\theta) \;\middle|\; \pi \in \Pi^0 \right\}$$

is a barycentrically symmetric subset of $\Delta(\mathcal{M})$. If (s', s'') is a Pareto efficient sharing rule for a T2 firm operated by agents with respective ambiguity tolerances λ' and λ'', then there exists a Pareto efficient sharing rule $(\hat{s}', \hat{s}'')$ with

$$\hat{s}'_j = \min\left\{\hat{s}'_J, m_i\right\} \qquad j = 1, 2, 3$$

that gives the same utility to the agent of type λ'.

[47] Recall that s' and s'' are *comonotone* provided

$$\left[s'_j - s'_k\right]\left[s''_j - s''_k\right] \geq 0$$

for all j, k. It follows that each of s' and s'' is comonotone with $\mathbf{m}$ – see Chateauneuf, Dana and Tallon (2000, Proposition 2.2).

Proposition 5.4 may be of independent interest. Several recent papers study Pareto efficient allocations for economies with MEU agents, but none of which we are aware contains this result. In fact, as the proof makes clear, Proposition 5.4 holds whether or not Π is barycentrically symmetric, provided $\mathcal{P}^{\lambda''} \subseteq \mathcal{P}^{\lambda'}$ and $q \gg 0$ for all $q \in \mathcal{P}^{\lambda'}$.

Proposition 5.5 generalises our Pareto efficient sharing result to $J = 3$ (arbitrary n), provided we restrict attention to belief functions that induce barycentrically symmetric sets of probabilities over $\mathcal{M}$.[48] In particular, if $\mathcal{P}^0$ is barycentrically symmetric, then so is $\mathcal{P}^\lambda$ for all $\lambda \in \left[0, \frac{1}{2}\right]$, as is obvious from Figure 14. Note that the condition $\lambda' > 0$, which does not appear in Proposition 3.1, is innocuous when the distribution of λ is atomless.

Proof of Proposition 5.4. We argue in two steps. First, adapting arguments from the proof of Chateauneuf, Dana and Tallon (2000, Proposition 3.1) one may show that Pareto efficient allocations are *necessarily* comonotone when agents a *strictly risk averse.* Next, we construct a Pareto efficient allocation for the economy with risk neutral agents as the limit of Pareto efficient allocations in nearby risk-averse economies.

Lemma 5.1 *Suppose agents λ' and λ'' in the statement of the Proposition are strictly risk averse, with strictly increasing and strictly concave vNM utility function $v : \mathcal{R}_+ \to \mathcal{R}_+$.*[49] *Then any Pareto efficient sharing rule is comonotone; i.e., each agent's payoff is comonotone with* $\mathbf{m}$.

Proof: Suppose that (s', s'') is Pareto efficient. Recall that

$$p^\lambda \;=\; \lambda'\overline{p} + \left(1 - \lambda'\right)\underline{p}$$

is the most pessimistic element of $\mathcal{P}^\lambda$ when an agent of type λ faces a payment vector comonotone with $\mathbf{m}$. This remains true if the agent is risk averse. Thus:

$$\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j v\left(s'_j\right) \;\leq\; \sum_{j=1}^{J} p_j^{\lambda'} v\left(s'_j\right)$$

$$\min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j v\left(s''_j\right) \;\leq\; \sum_{j=1}^{J} p_j^{\lambda''} v\left(s''_j\right)$$

We also observe that $p^\lambda \gg 0$ when $\lambda > 0$ by virtue of (19).

Consider a situation of "pure risk", in which both agents attach the unique probability $p^{\lambda''}$ to $\mathbf{m}$. We shall show that (s', s'') is Pareto efficient in this scenario. Suppose, to the contrary, that (s', s'') is Pareto dominated by the Pareto efficient allocation $(\hat{s}', \hat{s}'')$. Since $p^{\lambda''} \gg 0$, preferences

[48] For example, suppose $n = 3$, $\Pi \subseteq \Delta(\Theta)$ is barycentrically symmetric and

$$\mathrm{co}\left(\{p(\theta_i) \mid i = 1, 2, 3\}\right)$$

forms a triangle inside $\Delta(\mathcal{M})$ with edges parallel to those of $\Delta(\mathcal{M})$ itself. Then $\mathcal{P}^0$ is barycentrically symmetric.

[49] The result would continue to hold if we allowed agents to have different vNM utility functions, but the added generality is surplus to present requirements.

are strongly increasing, so we may assume that $(\hat{s}', \hat{s}'')$ *strictly* increases the expected utility of *both* agents. Furthermore, standard results on optimal risk sharing imply that: (i) $(\hat{s}', \hat{s}'')$ is comonotone; and (ii) $(\hat{s}', \hat{s}'')$ Pareto improves on (s', s'') when the agents assign unique probability $p^{\lambda'} \gg 0$ to $\mathbf{m}$ (see Chateauneuf, Dana and Tallon, 2000, Proposition 2.3). Therefore:

$$\left(\sum_{j=1}^{J} p_j^{\lambda'} v\left(s_j'\right), \sum_{j=1}^{J} p_j^{\lambda'} v\left(s_j''\right)\right) < \left(\sum_{j=1}^{J} p_j^{\lambda'} v\left(\hat{s}_j'\right), \sum_{j=1}^{J} p_j^{\lambda'} v\left(\hat{s}_j''\right)\right)$$

and

$$\left(\sum_{j=1}^{J} p_j^{\lambda''} v\left(s_j'\right), \sum_{j=1}^{J} p_j^{\lambda''} v\left(s_j''\right)\right) \ll \left(\sum_{j=1}^{J} p_j^{\lambda''} v\left(\hat{s}_j'\right), \sum_{j=1}^{J} p_j^{\lambda''} v\left(\hat{s}_j''\right)\right)$$

Since

$$\sum_{j=1}^{J} p_j^{\lambda'} v\left(\hat{s}_j'\right) = \min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j v\left(\hat{s}_j'\right)$$

and

$$\sum_{j=1}^{J} p_j^{\lambda''} v\left(\hat{s}_j''\right) = \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j v\left(\hat{s}_j''\right)$$

this violates the assumption that (s', s'') is Pareto efficient in the firm in which agent belief sets are $\mathcal{P}^{\lambda'}$ and $\mathcal{P}^{\lambda''}$. Therefore, (s', s'') is Pareto efficient when both agents attach the unique probability $p^{\lambda''}$ to $\mathbf{m}$.

It now follows, recalling Chateauneuf, Dana and Tallon (2000, Proposition 2.3) once again, that (s', s'') is comonotone. This completes the proof of Lemma 5.1. □

Let $\mathcal{E}$ denote the set of Pareto efficient allocations for the firm, and let $(s', s'') \in \mathcal{E}$. If (s', s'') is already comonotone there is nothing to prove, so suppose otherwise. For each $n \in \{1, 2, ...\}$ let $\mathcal{E}^n$ denote the set of Pareto efficient allocations for a $T2$ firm operated by agents with common vNM utility function $v^n(z) = z^{1-\frac{1}{2n}}$ and respective ambiguity tolerance parameters λ' and λ''. Note that v^n is strictly increasing and strictly concave for every $n \in \{1, 2, ...\}$, and $\lim_{n\to\infty} v^n(z) = z$. It follows by Lemma 5.1 that $(s', s'') \notin \mathcal{E}^n$ for any n. Let $\left(\hat{s}^{(n)\prime}, \hat{s}^{(n)\prime\prime}\right) \in \mathcal{E}^n$ with

$$\left(\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j v^n\left(s_j'\right), \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j v^n\left(s_j''\right)\right) < \left(\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j v^n\left(\hat{s}_j^{(n)\prime}\right), \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j v^n\left(\hat{s}_j^{(n)\prime\prime}\right)\right)$$

By compactness of the set of feasible allocations, it is without loss of generality to suppose that $\left(\hat{s}^{(n)\prime}, \hat{s}^{(n)\prime\prime}\right)$ converges to some feasible allocation $\left(\hat{s}^{(\infty)\prime}, \hat{s}^{(\infty)\prime\prime}\right)$ as $n \to \infty$. It follows that $\left(\hat{s}^{(\infty)\prime}, \hat{s}^{(\infty)\prime\prime}\right)$ is comonotone and

$$\left(\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j s_j', \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j s_j''\right) \leq \left(\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j \hat{s}_j^{(\infty)\prime}, \min_{q \in \mathcal{P}^{\lambda''}} \sum_{j=1}^{J} q_j \hat{s}_j^{(\infty)\prime\prime}\right).$$

Since $(s', s'') \in \mathcal{E}$, neither inequality can be strict, so $\left(\hat{s}^{(\infty)\prime}, \hat{s}^{(\infty)\prime\prime}\right) \in \mathcal{E}$ and

$$\min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j s'_j = \min_{q \in \mathcal{P}^{\lambda'}} \sum_{j=1}^{J} q_j \hat{s}_j^{(\infty)\prime}$$

This completes the proof of Proposition 5.4. □

Proof of Proposition 5.5: Invoking Proposition 5.4, we will assume that (s', s'') is comonotone, and hence each of s' and s'' is comonotone with $\mathbf{m}$. Let $u' = U\left(s'; \lambda'\right)$. If $u' \leq m_1$ we are done by Corollary 5.1, so assume $u' > m_1$.

Suppose that $s'_1 < m_1$. Adjust s' by increasing s'_1 an amount $\delta_1 > 0$ and decreasing each of s'_2 and s'_3 by $\delta_{2,3} > 0$ such that $U\left(s'; \lambda'\right)$ remains fixed and comonotonicity with $\mathbf{m} - s'$ is preserved. (This is always possible since we assume (19) and $\lambda' > 0$). Then, since s' is comonotone with $\mathbf{m}$ (both before and after the change):

$$p_1^{\lambda'} \delta_1 - \left(1 - p_1^{\lambda'}\right) \delta_{2,3} \; = \; 0$$

where

$$p^{\lambda'} \; = \; \lambda' \overline{p} + \left(1 - \lambda'\right) \underline{p}$$

Since s'' is also comonotone with $\mathbf{m}$ (both before and after the change) it follows that

$$p_1^{\lambda''} \delta_1 - \left(1 - p_1^{\lambda''}\right) \delta_{2,3} \; \leq \; 0 \quad \Leftrightarrow \quad - \left[p_1^{\lambda''} \delta_1 - \left(1 - p_1^{\lambda''}\right) \delta_{2,3}\right] \; \geq \; 0$$

where

$$p^{\lambda''} \; = \; \lambda'' \overline{p} + \left(1 - \lambda''\right) \underline{p}$$

and hence $p_1^{\lambda''} \leq p_1^{\lambda'}$. This implies that these changes (weakly) improve the welfare of the λ'' agent.

We may therefore assume that EITHER $s'_1 = s'_2 < m_1$ OR $s'_1 = m_1$. In the former case, we may achieve a weak mutual improvement in welfare by an analogous argument to that just given, this time increasing s'_1 and s'_2, decreasing s'_3, and using the fact that $p_3^{\lambda''} \geq p_3^{\lambda'}$. Since $u' > m_1$ we can continue this adjustment until $s'_1 = s'_2 = m_1$ or $s'_1 = s'_2 = s'_3$. In the latter case we are done.

It therefore remains to consider the case $s'_1 = m_1$.

Since s' is comonotone, it suffices to exclude $s'_2 < \min\{s'_3, m_2\}$. If this *were* the case, we could move s'_2 and s'_3 closer together while preserving both comonotonicity and the λ' agent's utility intact. In other words:

$$p_2^{\lambda'} \delta_2 - p_3^{\lambda'} \delta_3 \; = \; 0 \tag{22}$$

where $\delta_2 > 0$ is the increase in s'_2 and δ_3 is the magnitude of the decrease in s'_3. We shall show that such a move will (weakly) increase the utility of the other agent. To see why, recall that $\mathcal{P}^0$ is barycentrically symmetric. Since $J = 3$, this implies that p_2^{λ} is independent of λ, while p_3^{λ} is increasing in λ. Hence, from (22) we have:

$$p_2^{\lambda''} \delta_2 - p_3^{\lambda''} \delta_3 \; \leq \; 0 \quad \Leftrightarrow \quad - \left[p_2^{\lambda''} \delta_2 - p_3^{\lambda''} \delta_3\right] \; \geq \; 0.$$

This proves that agent λ'' is made (weakly) better off, which completes the proof. $\square$

APPENDIX E

We here consider the existence and structure of equilibrium in the n state, J output case.

Unless the $T1$ sector is inactive, the reservation utility of the marginal $T2$ participant must be K, irrespective of the values of n or J. If $K \leq m \equiv m_1$, then in the $n = J = 2$ case, *every* $T2$ firm pays its less optimistic member a fixed wage $w = K$. By virtue of Proposition 5.1, such an arrangement continues to be Pareto efficient for any n and any J. In the general case, and maintaining the assumption $K \leq m_1$, we may therefore construct an equilibrium as follows.

Set $w = K$ as the $T2$ wage. Assuming non-trivial ambiguity about $T2$ output, there will continue to exist a unique $\hat{\lambda}$ that divides types between those who prefer owning a $T2$ firm over working in one and those with the opposite preference (allowing for the possibility that $\hat{\lambda} \in \left\{0, \frac{1}{2}\right\}$ in case one occupation is strictly preferable for all types). If $\hat{\lambda} < \frac{1}{2}$ and not less than the median λ value, there is a unique $\tilde{\lambda}$ that satisfies

$$1 - H\left(\hat{\lambda}\right) \; = \; H\left(\hat{\lambda}\right) - H\left(\tilde{\lambda}\right)$$

and it will be an equilibrium for all types $\lambda \leq \tilde{\lambda}$ to participate in $T1$ firms, types $\lambda \in \left(\tilde{\lambda}, \hat{\lambda}\right)$ to work in $T2$ firms for fixed wage $w = K$, and the remaining types to own $T2$ firms. If $\hat{\lambda} \geq \frac{1}{2}$ it is an equilibrium for all agents to take occupations in the $T1$ sector. If $\hat{\lambda}$ falls below the median, there will exist an equilibrium in which all agents take occupations in $T2$ firms. In the latter case, we have not proved that the equilibrium sharing rules take the form of a fixed wage (subject to default), but this is a peripheral matter to our present concerns.

If $K > m_1$, the structure of equilibrium is less clear in the general case. When an enterprise is sufficiently uncertain that employees cannot be completely shielded from remuneration insecurity, we do not have a clear picture of the Pareto efficient rules for sharing that uncertainty. Indeed, in very new sectors employing very novel technologies, one often observes more exotic remuneration contracts: witness the proliferation of profit sharing arrangements, stock options and the like during the dot.com boom. Since our equilibrium is based on a standard competitive markets paradigm, it can only be expected to provide partial insights into such cases. In addition, when dealing with very uncertain technologies, the $\lambda \leq \frac{1}{2}$ assumption may be restrictive, as only agents from the extreme ranges of the optimism spectrum would be expected to take roles in such firms, and the efficient way to share uncertainty between such agents is even less well understood. Further progress on these issues is therefore beyond the scope of the present paper, but an important avenue for further research.

References

Aghion, P and P. Bolton (1997). "A Theory of Trickle-Down Growth and Develeopment", *Review of Economic Studies* **64:**151 -172**.**

Aghion, P and P. Howitt (1992): "A Model of Growth through Creative Destruction," *Econometrica*, **60**, 323-51.

Anscombe, F.J. and R.J. Aumann (1963). "A Definition of Subjective Probability", *Annals of Mathematical Statistics.* **34**, 199-205.

Arrow, K. J. and L. Hurwicz (1972). "An Optimality Criterion for Decision-Making Under Ignorance." In *Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle*, C. F. Carter and J. L. Ford (eds). Oxford: Basil Blackwell.

Aubin, J.-P. (1998). *Optima and Equilibria: An Introduction to Nonlinear Analysis* (2nd edition). Berlin: Springer-Verlag.

Balcer, Y. and S. A. Lippman (1984) "Technological Expectations and Adoption of Improved Technology" *Journal of Economic Theory* **34**, 292-318.

Banerjee, A. V and A.F. Newman (1993). "Occupational Choice and the Process of Development", *Journal of Political Economy.***101**(2): 274 - 298.

Barbarino, A. and B.Jovanovic (2005). *Shakeouts and Market Crashes*, NBER Working Paper No. W10556.

Berger, J.O. (1996): *Bayesian Robustness*, Institute of Mathematical Statistics Hayward, California.

Bernardo, A.E. and I. Welch (2001). "On the Evolution of Overconfidence and Entrepreneurs" *Journal of Economics and Management Strategy.***10,** 301–330**.**

Blanchflower, D. and A. Oswald (1998): "What Makes an Entrepreneur?", *Journal of Labor Economics* **16**, 26-60.

Border, K. (1985). *Fixed Point Theorems with Application to Economics and Game Theory.* Cambridge: CUP.

Bose, S., Ozdenoren, E. and Pape (2006): "Optimal auctions with ambiguity", Theoretical Economics **1**, 411-438

Calvo, G.A. and S. Wellisz (1980). "Technology, Entrepreneurs, and Firm Size," *Quarterly Journal of Economics* **95**, 663–677.

Camerer, C. (1995). "Individual Decision Making," in *The Handbook of Experimental Economics*, J.H. Kagel and A.E. Roth (eds.). Princeton: Princeton University Press.

Chateauneuf, A., Dana R.A., and J.M. Tallon (2000) "Optimal risk-sharing rules and equilibria with Choquet-expected-utility", Journal of Mathematical Economics **34**, .191–214.

Dempster, A.P. (1967). "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics* **38**, 325–339.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics* **75**, 643-669.

Evans, D.S. and B. Jovanovic (1989). "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *Journal of Political Economy* **97**, 808–827.

Evans, D.S., Jovanovic, B. and L.S. Leighton (1989) "Some Empirical Aspects of Entrepreneurship" *American Economic Review* **79**, 519–535.

Fox, C.R., and A. Tversky (1995): "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics* **110**, 585-603.

Geanakplos, J (2002). *Liquidity, Default and Crashes: Endogenous contracts in general equilibrium.* Cowles Foundation Working Paper No. 1316RR.

Ghatak, M., Morelli, M. and T. Sjoestroem (2001): "Occupational Choice and Dynamic Incentives," *Review of Economic Studies* **68**, 781-810.

Ghirardato, P. (2001). "Coping With Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty," *Economic Theory* **17**, 247-276.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility With Non-Unique Prior," *Journal of Mathematical Economics* 18, 141-153.

Gort, M. and S. Klepper (1982): "Time Paths in the Diffusion of Product Innovation", *Economic Journal* **92**, 630-653.

Griliches, Z. (1957). "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* **25**, 501-522.

Grüner, H.P (2003). "Redistribution as a Selection Device", *Journal of Economic Theory* **108**, 194–216.

Guiso, L., Sapienza, P. and L. Zingales (2004): "The Role of Social Capital in Financial Development," *American Economic Review* **94**, 526-556.

Heath, C. and A. Tversky (1991). "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty", *Journal of Risk and Uncertainty* **4**, 5–29.

Holtz-Eakin, D., Joulfaian, D., and H. Rosen (1994a): "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints", *Journal of Political Economy* **102**, 53-75.

Holtz-Eakin, D., Joulfaian, D., and H. Rosen (1994b): "Entrepreneurial Decisions and Liquidity Constraints," *RAND Journal of Economics* **25**, 334-347.

Jaffray, J.-Y. (1989). "Linear Utility Theory for Belief Functions," *Operations Research Letters* **8**, 107-112.

Jaffray, J.-Y. (1991). "Linear Utility Theory and Belief Functions: A Discussion," in A. Chikán (ed.) *Progress in Decision, Utility and Risk Theory.* Dordrecht: Kluwer Academic.

Jaffray, J.-Y. (1994): "Dynamic decision making with belief functions", in R. R. Yager, J. Kacprzyk, & M. Fedrizzi (Eds.), *Advances in the Dempster-Shafer theory of evidence* (pp. 331–352). Wiley, New York.

Jaffray, J.-Y. and F. Phillipe. (1997). "On the Existence of Subjective Upper and Lower Probabilities," *Mathematics of Operations Research* **22**, 165-185.

Jaffray, J.-Y. and P. Wakker. (1994). "Decision Making with Belief Functions: Compatibility and Incompatibility with the Sure-Thing Principle," *Journal of Risk and Uncertainty* **8**, 255–271.

Jensen, R. (1982). "Adoption and Diffusion of an Innovation of Uncertain Profitability," *Journal of Economic Theory* **27**, 182–193.

Jensen, R. (1983). "Innovation Adoption and Diffusion When There Are Competing Innovations," *Journal of Economic Theory* **29**, 161-171.

Jovanovich, B. and S. Lach (1989): "Entry, Exit, and Diffusion with Learning by Doing," *American Economic Review*, **79**, 690-699

Jovanovich, B. and G. M. MacDonald (1994): "The Life Cycle of a Competitive Industry," *Journal of Political Economy* **102**, 322-347.

Kelsey, D. and W. Spanjers. (1997). "Uncertainty in Partnerships," Department of Economics Discussion Paper 97-16, University of Birmingham.

Kihlstrom, R.E. and J.-J. Laffont. (1979). "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy* **87**, 719–748.

King, R. G. and R. Levine (1993): "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics* **108**, 717-37.

Knight, F. (1921). *Risk, Uncertainty and Profit.* Cambridge (MA): Houghton Mifflin.

Lazear, E. (2002). "Entrepreneurship", NBER Working Paper 9109.

Lloyd-Ellis, H. and D. Bernhardt (2000): "Enterprise, Inequality and Economic Development", *Review of Economic Studies* **67**, 147-168.

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

Mansfield, E. (1968). *Economics of Technological Change.* Norton.

Moskowitz, T.J. and A. Vissing-Jørgensen (2002). "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?", *American Economic Review* **92**, 745–778.

Mukerji, S. (1997). "Understanding the Non-Additive Probability Model," *Economic Theory* **9**, 23–46.

Mukerji, S. (1998). "Ambiguity Aversion and Incompleteness of Contractual Form," *American Economic Review* **88**, 1207–1231.

Mukerji, S. and J.-M. Tallon (2001). "Ambiguity Aversion and Incompleteness of Financial Markets," *Review of Economic Studies* **68**, 883–904.

Philippe, F. (2002). "Tools for Decision Making Under Imprecise Risk," *Statistical Papers* **43**, 95–110.

Philippe, F., G. Debs and J.-Y. Jaffray (1999). "Decision Making with Monotone Lower Probabilities of Infinite Order," *Mathematics of Operations Research* **24**, 767–784.

Pikkety, T. (1997): "The Dynamics of the Wealth Distribution and the Interest Rate with Credit-Rationing," *Review of Economic Studies* **64**, 173-189.

Schmeidler D. (1989): "Subjective Probability and Expected Utility without Additivity", *Econometrica*, 57: 571–587.

Schumpeter, J.A. (1947): "The Creative Response in Economic History," *Journal of Economic History* **7**, 149-159.

Shafer, G. (1976): *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press.

Shapley, L.S. (1971): "Cores of convex games", *International Journal of Game Theory*, **1**: 11–26.

Srivastava, R.P. and T.J. Mock (2002). *Belief Functions in Business Decisions.* New York: Physica-Verlag.

Vettas, N. (1998): "Demand and Supply in New Markets: Diffusion and Bilateral Learning," *RAND Journal of Economics* **29**, 215–233.

Viero, M.-L. (2006): "Exactly What Happens After the Anscombe-Aumann Race? Representing Preferences in Vague Environments", QED Working Paper No. 1094

Wasserman, L.A. (1990a): "Prior Envelopes Based on Belief Functions", *Annals of Statistics* **18**, 454-464.

WASSERMAN, L.A. (1990b): "Belief Functions and Statistical Inference", *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* **18**, 183-196.

WASSERMAN, L.A. AND KADANE, J.B. (1990): "Prior Envelopes Based on Belief Functions", *The Annals of Statistics*, **18**: 454–464.

WASSERMAN, L.A. AND KADANE, J.B. (1992): "Symmetric Upper Probabilities", *The Annals of Statistics*, **20**: 1720–1736.

WASSERMAN, L.A. AND KADANE, J.B. (1996): "Symmetric, Coherent, Choquet Capacities", *The Annals of Statistics*, **24**: 1250–1264.