Scientific Research Publishing

Theoretical Economics Letters, 2024, 14, 1165-1192
https://www.scirp.org/journal/tel
ISSN Online: 2162-2086
ISSN Print: 2162-2078

On the Validity of Probabilities in Uncertainty Assessment: The Role of Learning

Jean-Paul Chavas

Department of Agricultural and Applied Economics, University of Wisconsin, Madison, USA
Email: jchavas@wisc.edu

How to cite this paper: Chavas, J.-P. (2024). On the Validity of Probabilities in Uncertainty Assessment: The Role of Learning. *Theoretical Economics Letters, 14,* 1165-1192.
https://doi.org/10.4236/tel.2024.143060

Received: February 28, 2024
Accepted: June 25, 2024
Published: June 28, 2024

Abstract

Probabilities have been accepted as providing a general representation of uncertainty. We investigate the validity of probability assessments using a state-contingent model under uncertainty, with a focus on the role of learning. We use the model to define shadow prices of state-contingent goods under general conditions. Interpreting probabilities as normalized shadow prices of state-contingent goods, we obtain new insights into why and when probability theory fails to provide an adequate representation of uncertainty. Our analysis focuses on the validity of Kolmogorov's additivity axiom in probability theory. We identify three sets of factors contributing to non-additive probabilities: 1) uncertainty-loving behavior, 2) active learning under costly information, and 3) the presence of an infinite number of states.

Keywords

Uncertainty, Probability, Kolmogorov Axioms, Countable Additivity, Learning, Shadow Prices, Nonconvexity

1. Introduction

We live in an uncertain world facing many events that are not fully known ahead of time. But assessing such events has proved challenging. Probability theory has provided a general framework to represent uncertainty. This paper examines the validity of probabilities commonly used in risk assessment, with a focus on the role of learning. Special attention is given to Kolmogorov's additivity axiom in probability theory. We investigate the validity of Kolmogorov's axioms using a state-contingent model under uncertainty, allowing for asymmetric information and learning. We use the model to define shadow prices of state-contingent goods under general conditions. Assuming that probabilities represent normalized shadow prices of state-contingent goods, we examine when and how probability theory can fail to provide an adequate representation of uncertainty.

Note that we are not the first ones to question the relevance of probabilities in uncertainty assessments. de Finetti (1974) started his seminal contribution to probability theory by stating: "probability does not exist"[1]. He made this argument by stressing that there is no "objective probability", only "subjective probability" reflecting each individual's degree of belief in the occurrence of a given event. de Finetti's subjective interpretation of probabilities has been widely adopted. It has contributed to the development of probability theory (Savage, 1954; Kolmogorov, 1956; Billingsley, 1995) and its widespread use in the analysis of uncertain events. Yet, it remains unclear whether the use of probabilities is always appropriate. As argued by Knight (1921), there are situations where evaluating probabilities is difficult. To the extent that information is obtained as the outcome of a learning process, this suggests the need to examine the linkages between learning and uncertainty assessment and to revisit the behavioral foundations of probabilities. Addressing these issues is the main objective of this paper.

Learning processes are complex: they include individual learning as well as social learning. Individual learning involves multisensory integration of signals by the brain, with different signals often taking different amounts of time to be processed and with results that often interact across sensory systems. Building on developments in cognitive neuroscience, neural responses to information processing typically involve thresholds, with responses occurring only beyond some minimum stimulus levels (e.g. Gazzaniga et al., 2002; Quiroga et al., 2008). In addition, neural responses to multiple signals are often non-additive (e.g. Holmes & Spence, 2005; Stanford & Stein, 2007; Sadaf et al., 2023): they can be subadditive (when different signals inhibit each other) or superadditive (when different signals strengthen each other). In social learning, the signals are generated through social interactions and allow individuals to learn from each other (Hunt et al., 2012). In general, nonlinearities in learning processes are not captured in Bayesian analysis, indicating that probabilities and their Bayesian updating can fail to provide a realistic representation of learning (Bowers & Davis, 2012). Our paper examines these issues and their implications for uncertainty assessment.

Our analysis relies on a state-contingent approach to uncertainty. As noted by Debreu (1959: p. 98), the state-contingent approach is "free from any probability concept", making it suitable to investigate the behavioral validity of probabilities. Our analysis examines the allocation of production, consumption as well as information-gathering activities. We start with a representation of efficiency based on the maximization of aggregate willingness to pay. This representation holds under general conditions, allowing for asymmetric information and active learning under costly information. Following Yaari (1969), probabilities can be interpreted as normalized prices of state-contingent goods. But this raises an important is-

---

[1] Similar arguments were presented in Nau (2001).

sue: risk markets are typically incomplete (e.g. Radner, 1968), meaning that state-contingent market prices are often unobserved. As a result, this paper investigates the validity of probabilities based on shadow prices of state-contingent goods. Shadow prices are defined as marginal willingness-to-pay for state-contingent goods, which apply under general conditions (i.e. with or without risk markets). Our investigation has four important features: 1) using a state-contingent approach, we investigate the efficiency of resource allocation under uncertainty and learning; 2) we allow for individual as well as social learning; 3) we allow for externalities and nonconvexity; and 4) we do not require that risk markets necessarily exist. Externalities arise under social learning. Nonconvexity[2] arises in the presence of externalities (Starrett, 1972; Chavas, 2015), fixed cost (Radner, 1968) and/or increasing returns (Arthur, 1989; Chavas, 2017). Going beyond the Arrow-Debreu model, nonconvexity can require nonlinear pricing. Our analysis applies to markets as well as nonmarket allocations (which is important when risk markets fail to develop under asymmetric information). We define the shadow prices of state-contingent goods under general conditions (including situations where risk markets are incomplete). Following Yaari (1969), we interpret subjective probabilities as normalized shadow prices of state-contingent goods. This provides the framework for us to investigate the validity of Kolmogorov's axioms.

The paper obtains new insights into why and when probability theory fails to provide an adequate representation of uncertainty. Examining the properties of shadow prices for state-contingent goods, we study their implications for the validity of probabilities. Our arguments focus on questioning the validity of Kolmogorov's additivity axiom. While we discuss conditions under which Kolmogorov's axioms would hold, our main contributions are to find situations where non-additive probabilities are required. We identify three sets of factors contributing to non-additive probabilities: 1) uncertainty-loving behavior; 2) active learning under costly information; and 3) the presence of an infinite number of states. By building linkages between probabilities, learning and cognitive neuroscience, our findings indicate how probabilities and their Bayesian updating can fail to provide a realistic representation of learning. Besides identifying the inadequacies of probability theory, our analysis stresses the need to refine our understanding of the behavioral aspects of human learning.

The paper is organized as follows. Section 2 provides an overview of the history of probability theory. Section 3 develops a state-contingent model of production, consumption and information-gathering activities. In Section 4, the model is used to characterize efficiency as well as the shadow prices of state-contingent

---

[2] In a vector space, a set Z is said to be convex if $z_1 \in Z$ and $z_2 \in Z$ implies that $\left(\theta z_1 + \theta z_2\right) \in Z$ for any $\theta \in [0,1]$. Applied to a technology, assuming that the feasible set Z is convex is equivalent to assuming diminishing marginal productivity. Convexity is often assumed in economic analysis (e.g. Debreu, 1959). But as noted, nonconvexity arises in a number of situations that will prove relevant in our discussion. As a result, our analysis is presented under more general conditions (i.e. it does not assume convexity).

goods under general conditions. Section 5 explores the properties of these shadow prices and their implications for probability theory. Finally, Section 6 concludes.

2. Historical Perspectives

This section presents a brief overview of the history of probability theory. Starting in the sixteenth and seventeenth centuries, probabilities were first proposed in the evaluation of games of chance. Girolamo Cardano (1501-1576) calculated the odds of facing certain dice outcomes; and Blaise Pascal (1623-1662) and Pierre de Ferma (1601-1665) started the theory of probability assessing the likelihood of uncertain outcomes in gambling games (Lightner, 1991). Over the last three centuries, the mathematical concept has been refined and developed, leading to modern probability theory applied to general uncertain events (Kolmogorov, 1956; Billingsley, 1995). In this context, letting S be a non-empty sample space and $\mathbf{F}$ be a σ-field on S[3], a probability measure is defined as a function $P:\mathbf{F}\to\mathbb{R}$ satisfying the following axioms:

Axiom Ax1 (non-negativity): $P(A)\geq 0$ for all $A\in\mathbf{F}$ (1a)

Axiom Ax2 (normalization): $P(S)=1$ (1b)

Axiom Ax3 (countable additivity): If $A_1, A_2, \cdots$ is a countable sequence of subsets of $\mathbf{F}$ where $A_i \cap A_j = \varnothing$ for all $i \neq j$ and $(A_1 \cup A_2 \cup \cdots) \in \mathbf{F}$, then:

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \tag{1c}$$

Building on Kolmogorov (1956), axioms Ax1 - Ax3 have become a standard basis of probability and statistical analyses (Billingsley, 1995). Axioms Ax1 and Ax2 are not contentious. But Axiom Ax3 has been more controversial. Axiom Ax3 states that probabilities are countably additive, i.e. that the probability of a union of a finite or countably infinite collection of disjoint events is the sum of the corresponding probabilities. In Axiom Ax3, one issue is to decide between finite additivity and countable additivity in (1c). While Kolmogorov (1956) proposed to rely on countable additivity, Savage (1954) and de Finetti (1974) argued in favor of finite additivity (with Ax3 being defined under a finite sequence $A_1, A_2, \cdots, A_n$). Savage (1954) and de Finetti (1974) argued in favor of finite additivity in part because a uniform probability distribution over the integers exists under finite additivity but not under countable additivity (Bingham, 2010). As discussed in Bingham (2010) and Liu (2019), the arguments against countable additivity have been largely inconclusive. We will revisit this issue in Section 5. But the assumption of additivity in Equation (1c) is potentially more problematic. The main arguments we present below focus on the relevance of the additivity Assumption (1c) in probability and uncertainty assessment.

Probabilities were originally developed to assess outcomes in games of chance.

---

[3] A class $\mathbf{F}$ of subsets of S is a σ-field if it is closed under complement, countable unions and countable intersections.

In this context, a probability can be interpreted as a measure of the relative frequency of a specific repeatable event (e.g. the odd of obtaining a 6 when rolling a 6-sided die). The probability of a repeated event can then be assessed objectively through repeated sample observations of the event. The analysis was generalized by extending the interpretation of probabilities from a relative frequency to a relative subjective likelihood (Savage, 1954; de Finetti, 1974). This greatly expanded the realm of applications of probabilities in three ways: 1) it covered the uncertainty assessment of non-repeatable events; 2) it allowed subjective probabilities to be personal and to vary across individuals; and 3) it stimulated the use of Bayesian statistics (with Bayes theorem providing a formal framework to analyze the updating of probabilities as new information becomes available). Uncertainty assessment can be done using probabilities as long as each individual can evaluate the relative likelihood of specific events in a consistent way (DeGroot, 1970). Over the last few decades, these developments contributed to the widespread use of probability assessment in analyzing and managing uncertain situations.

In the analysis of uncertainty, a common approach has been to proceed in two steps: 1) rely on probability assessments (under Axioms Ax1 - Ax3) to provide a formal representation of uncertainty; and 2) use these probabilities to evaluate how uncertainty affects human behavior. This is a context where risk is equivalent to uncertainty: risk exposure is measured using probabilities in step 1; and risk preferences are used in step 2 to evaluate how risk affects welfare and behavior. Savage (1954) proposed to analyze decision making under uncertainty based on the expected utility (EU) model. In the EU model, risk preferences of the decision maker are represented by a von Neumann-Morgenstern utility function (defined up to a positive linear transformation), where risk aversion corresponds to a concave utility function (Savage, 1954; Arrow, 1971). Does the EU model provide an accurate representation of behavior under risk? It does not. As first pointed out by Allais (1953), most decision makers treat probabilities differently from the way postulated in the EU model. Compared to the EU model, Allais's evidence indicates that individuals tend to "overweight" the probability of rare events. This shows how and why the EU model fails: derived under Savage's "independence axiom", the EU model exhibits the property of being "linear in the probabilities" which is overly restrictive (Machina, 1982). Such arguments stimulated Kahneman and Tversky (1979) to develop prospect theory[4], specifying a utility function that is nonlinear in the probabilities with probability weights reflecting how decision makers depart from the EU model by "overweighing" low probability events.

Under either the EU model or prospect theory, the evaluation of risk relies on probabilities (under Axioms Ax1 - Ax3). A more fundamental question addressed in this paper is: Could there be situations where probability assessments

---

[4] Prospect theory also stresses that loss aversion is a common characteristic of individual risk preferences (Kahneman & Tversky, 1979).

are inappropriate? On the one hand, when applied to repeatable events, the relative frequency interpretation of probabilities appears intuitive. On the other hand, imposing Axiom Ax3 on all uncertain events may not always be appropriate. Knight (1921) argued that we should not impose a probability structure on all uncertain events. He proposed to distinguish between risk (where probability assessments are possible) and uncertainty (where probability assessments are not). This line of inquiry has been explored by Ellsberg (1961) and Halevy (2007) who presented evidence that individual behavior distinguishes between probability-based risk and ambiguity (corresponding to uncertain events with unknown probabilities)[5].

These arguments have stimulated a debate on how to evaluate uncertainty in the assessment of human behavior. One option is to relax Axiom Ax3 by allowing probabilities to be non-additive (e.g. Schmeidler, 1989; Tversky & Kaheman, 1992). Under an axiom of comonotonic independence (which is weaker than Savage's independence axiom), Schmeidler (1989) showed that decision makers behave in a way consistent with expected utility given by a Choquet integral where belief functions are measured by non-additive Choquet capacities. This approach was further explored in Tversky and Kahneman (1992) who proposed cumulative prospect theory (CPT) as a more general model of behavior under uncertainty. Tversky and Kahneman (1992) argued that their model has two interpretations: 1) for probabilistic prospects, CPT is an extension of prospect theory (as proposed in Kahneman & Tversky, 1979), which allows for nonlinearity in the probabilities; and 2) for uncertain prospects, CPT includes Schmeidler's (1989) model as a special case, which allows for non-additive probabilities. These two interpretations raise an important issue: in the analysis of behavior under uncertainty, CPT cannot distinguish between "risk preferences that are nonlinear in probabilities" versus "non-additive probabilities". This is a key identification issue, which means that neither Schmeidler's (1989) model nor CPT can be used to evaluate the validity of Axiom Ax3.

Another way to explore the validity of probability is to rely on a state-contingent representation of uncertainty. A state-contingent approach can support a general equilibrium analysis of economic decisions under uncertainty without using probabilities (Debreu, 1959: p. 98). This approach has four attractive features. First, it provides a flexible representation of preferences under uncertainty (e.g. it does not require Savage's "independence axiom" or Schmeidler's "comonotonic independence axiom"). Second, it can be used to evaluate each individual's willingness-to-pay to make a bet on uncertain events. As argued by Yaari (1969), this means using normalized prices of uncertain events as measures of the associated subjective probabilities. In this context, while the state-contingent approach

---

[5] The presence of ambiguity has stimulated the development of probability-based models representing ambiguity aversion. This includes the maxmin expected utility model under multiple prior probability distributions, the Choquet expected utility model (Schmeidler, 1989) and the models of ambiguity aversion proposed by Klibanoff et al. (2005) and Cerreia-Vioglio et al. (2011).

does not rely explicitly on probabilities, it can generate subjective probabilities as a byproduct of evaluating willingness-to-pay or the associated shadow prices of state-contingent good, thus providing a basis to evaluate the validity of Axioms Ax3. Third, the state-contingent approach has the advantage of avoiding identification issues that can arise if we try to separate uncertainty assessments from preference assessments (e.g. as done in the EU model or CPT). Fourth, there are situations where assessing probabilities can be difficult (Knight, 1921). This can occur in complex situations under limited cognition (Simon, 1955; Bowers & Davis, 2012). Since cognition is the outcome of a learning process, this argument stresses the need to investigate the linkages between learning and the assessment of uncertain events.

As noted in the introduction, learning processes are complex. Cognitive neuroscience has documented that neural and behavioral responses to signals involve thresholds, with responses occurring only beyond some minimum stimulus levels (e.g. Gazzaniga et al., 2002; Quiroga et al., 2008). In addition, our brain typically receives signals from multiple sources (including signals from others in the case of social learning) and integrates them to generate cognition. Often, this integration involves non-additive neural responses (e.g. Holmes & Spence, 2005; Stanford & Stein, 2007; Sadaf et al., 2023). Such nonlinearities suggest that the additivity postulate in Ax3 and Bayesian analysis can fail to provide a realistic representation of learning (Bowers & Davis, 2012).

Relying on a state-contingent approach to decision making under uncertainty, our analysis provides a framework to investigate the role of learning. As access to information and the ability to process information typically vary across individuals, we allow for information and learning to be heterogenous across members of society. The state-contingent approach presented below captures such features, thus supporting the analysis of linkages between learning, the assessment of uncertainty and the validity of Axiom Ax3. Following Yaari (1969), our investigation of the validity of Axiom Ax3 is based on the shadow prices of state-contingent goods. Evaluating these shadow prices is the topic of Sections 3 and 4 below.

3. A Model of Decisions under Uncertainty

This section presents a model of resource allocation under uncertainty. It also sets up the notation for the rest of the paper. Consider an economy involving the allocation of goods among n individuals under uncertainty, where $N = \{1, \cdots, n\}$ is the set of individuals. The uncertainty is represented by a state of nature $s \in S$, where S is the non-empty set of all possible states of nature (e.g. Debreu, 1959; Radner, 1968; Chavas & Briec, 2019). Our analysis expands on previous research in three ways: 1) we study the role of learning in uncertainty assessment; 2) we allow for nonconvexity; and 3) we consider the case where S is a countable set (thus allowing us to examine the difference between countable versus finite state space). As discussed below, these three features will play a role in evaluating the validity of probability assessments.

We consider several types of activities: production denoted by $y = (y_1, \cdots, y_m) \in \mathbb{R}^m$, individual consumption denoted by $x_i = (x_{i1}, \cdots, x_{im}) \in \mathbb{R}^m, i \in N$, and information gathering activities denoted by z. We assume the learning takes place as individuals invest some time in K learning processes, each learning process representing alternative ways to acquire information. We let $z = (z_1, \cdots, z_n) \in Z \subset \mathbb{R}^{nK}$ where $z_i = (z_{i1}, \cdots, z_{iK}) \in \mathbb{R}^K$, z_{ik} being the amount of time allocated by the i-th individual to the k-th learning process, $k \in \{1, \cdots, K\}$. The presence of multiple learning processes allows for heterogeneity in learning across individuals. And it allows for individual learning (when a learning process is individual-specific) as well as social learning (when a learning process involvesmultiple individuals).

First, we consider production activities in the economy. With some abuse of notation, we let $y(s) \equiv (y_1(s), \cdots, y_m(s))$ where $y_i(s)$ $y_j(s)$ denotes the state-contingent decision rule stating the quantity of the j-th netput chosen under state $s \in S$, $j \in M \equiv \{1, \cdots, m\}$. We use the netput notation where outputs in y are positive and inputs are negative. Similarly, considering consumption activities in the economy, we let $x_i(s) \equiv (x_{i1}(s), \cdots, x_{im}(s))$ where $x_{ij}(s)$ denotes the decision rule stating the state-contingent quantity of the j-th good chosen by the i-th individual under state s, $j \in M$, $s \in S$ and $i \in N$. Finally, we assume that the information gathering activities $z \in Z$ are chosen *ex ante*, i.e. that they are the same across all states $s \in S$ [6].

Let $\tilde{y} \equiv [y(s) : s \in S]$ and $\tilde{x} = (\tilde{x}_1, \cdots, \tilde{x}_n)$ where $\tilde{x}_i \equiv [x_i(s) : s \in S]$, $i \in N$. Denote by $F \subset \prod_{s \in S} \mathbb{R}^m \times \mathbb{R}^{nK}$ the feasible set for $(\tilde{y}, z)$, where $(\tilde{y}, z) \in F$ means that the decisions $(\tilde{y}, z)$ are feasible. Conditional on z, the set $Y(z) \equiv \{\tilde{y} : (\tilde{y}, z) \in F\}$ provides a representation of the technology supporting production decisions. Allowing $Y(z)$ to depend on the information gathering activities z reflects that, under active learning, scarce resources can be allocated to both the learning process and the production process (e.g. the case of learning-by-doing). Also denote by $X_i \subset \prod_{s \in S} \mathbb{R}^n$ the set of possible alternatives for $\tilde{x}_i$, with $X \equiv \prod_{i \in N} X_i$. Below, we assume that each X_i is equipped with the product topology τ_{prod}, in which case (X_i, τ_{prod}) are each metrizable when S is either finite or countably infinite (Munkres, 1975: p. 123). We also assume that F is equipped with the product topology τ_{prod}, implying that the space (F, τ_{prod}) is metrizable.

For the i-th individual, consumer preferences are represented by an ordinal utility function $v_i(\tilde{l}_i, \tilde{x}_{i2}, \cdots, \tilde{x}_{im})$ where $\tilde{l}_i \equiv [l_i(s) : s \in S]$ denotes state-contingent leisure time and $(\tilde{x}_{i2}, \cdots, \tilde{x}_{im})$ denotes the state-contingent consumption of other consumer goods (besides leisure). The decisions $(\tilde{l}_i, \tilde{x}_{i2}, \cdots, \tilde{x}_{im})$ being state-

---

[6] This simplifying assumption could be relaxed. It is motivated by the evolution of information over time: 1) the choice of the information gathering activities z cannot be based on information that has not yet been obtained; and 2) if people forget, then information acquired at one time may not be available later. These issues do not arise when z is chosen *ex ante*.

contingent, the utility function $v_i\left(\tilde{l}_i, \tilde{x}_{i2}, \cdots, \tilde{x}_{im}\right)$ provides a general *ex ante* evaluation of decisions under uncertainty for the *i*-th individual. Time allocation is subject to the time constraint: $T = \tilde{l}_i + \left|\tilde{x}_{i1}\right| + \sum_{k=1}^{K} z_{ik}$, where T is total time available, $\tilde{x}_{i1}$ is labor supply (defined to be negative by convention) and z_{ik} is the time spent by the *i*-th individual in the *k*-th learning process. Assuming that $0 < \tilde{l}_i < T$ and using the time constraint, the utility function for the *i*-th individual can be written as: $u_i\left(\tilde{x}_i, z_i\right) \equiv v_i\left(T - \left|\tilde{x}_{i1}\right| - \sum_r z_{ir}; \tilde{x}_{i2}, \cdots, \tilde{x}_{im}\right)$, where $\tilde{l}_i = T - \left|\tilde{x}_{i1}\right| - \sum_r z_{ir}$. For the *i*-th individual, the utility function $u_i\left(\tilde{x}_i, z_i\right)$ reflects consumer preferences, including time allocation between leisure, labor and information gathering activities, making it clear that spending time learning has an opportunity cost.

Under uncertainty, production and consumption decisions are made depending on the information available to the decision makers. This information can vary depending on who makes each decision and its timing. In general, there are two sources of heterogeneity in information: temporal heterogeneity as each decision maker can learn over time[7]; and spatial heterogeneity reflecting that information typically varies among decision makers due to differences in access to information and/or the ability to process it. As a result, each decision can be made based on different information. The role of information is captured using information partitions of the state space *S* (Radner, 1968). $\mathbf{F}$ being a *σ*-field on *S*, for a given decision rule $w(\cdot)$, an information partition P_w is a collection of disjoint subsets of $\mathbf{F}$, where an information partition P_w is finer (more informative) than information partition P'_w if every element of P_w is a subset of some element of P'_w. For a given information partition $P_w = \left\{P_{w1}, P_{w2}, \cdots\right\}$, the decision maker can distinguish among states that are in different elements of the partition but not among states within each partition. This imposes the following restrictions on the decision rule $w(s)$ (Radner, 1968):

$$w(s) = w(s') \text{ when states } s \text{ and } s' \text{ are both in the same element } P_{wk} \in P_w. \quad (2)$$

At one extreme, if a decision $w(s)$ is made *ex ante* (i.e. without any information), the associated partition would have only one element, yielding the information constraints $w(s) = w(s')$ for all $s \neq s' \in S$, the decision $w(s)$ being restricted to be the same across all states. At the other extreme, if the decision $w(s)$ is made *ex post* (i.e. under perfect information), there would be a single state in each element of P_w, yielding no information constraint. In between, information constraints would be less (more) restrictive as P_w becomes finer (coarser).

We assume that information can vary for each of the production and consumption decision. Let $P_y = \left(P_{y1}, \cdots, P_{ym}\right)$ where P_{yj} is the information partition associated with $y_j : j = 1, \cdots, m$. And let $P_x = \left(P_{x11}, \cdots, P_{x1m}, P_{xn1}, \cdots, P_{xnm}\right)$ where P_{xij} is the information partition associated with $x_{ij}, i \in N, j \in M$. It follows that

---

[7] Note that our analysis also allows for information loss over time (when a decision maker forgets information obtained during earlier periods).

the information structure supporting the choice of production and consumption decisions in the economy is $P=\left(P_y,P_x\right)$. Let $P\left(z\right)=\left(P_y\left(z\right),P_x\left(z\right)\right)$ denote the information partitions obtained under information gathering activities $z=\left(z_1,\cdots,z_n\right)\in Z$ using K learning processes. The mapping $P\left(z\right)$ represent show information is obtained and processed by all decision makers in society. The K learning processes represent alternative ways of acquiring information, including both individual learning and social learning. Thus, $P\left(z\right)$ reflects how information is managed in society as it supports the functioning of the economy, allowing for asymmetric information as well as different learning processes across individuals.

The information constraints (2) impose restrictions on both production and consumption decisions. Let:

$$Y_c\left(z\right)\equiv\left\{\tilde{y}:\tilde{y}\in Y\left(z\right):\tilde{y}\ \text{ satisfies (2) given }\ P\left(z\right)\right\} \tag{3a}$$

$$X_c\left(z\right)=\left(X_{c1}\left(z\right),\cdots,X_{cn}\left(z\right)\right)\equiv\left\{\tilde{x}:\tilde{x}\in X:\tilde{x}\ \text{ satisfies (2) given }\ P\left(z\right)\right\}. \tag{3b}$$

Note that the information gathering activities $z\in Z$ play three roles in (3): 1) they involve individual time allocated to learning (which has labor and leisure as opportunity cost); 2) they can be combined with other goods y to produce information about the state of nature; and 3) they contribute to generating finer information partitions supporting both production and consumption decisions in the economy.

We want to investigate the choice of allocation $\left(\tilde{x},\tilde{y},z\right)$.

Definition 1: An allocation $\left(\tilde{x},\tilde{y},z\right)$ is said to be feasible if it satisfies $\tilde{x}\in X_c\left(z\right)$, $\tilde{y}\in Y_c\left(z\right)$, $z\in Z$ and[8]

$$\sum\nolimits_{i\in N}\tilde{x}_i\le\tilde{y}\,. \tag{4}$$

Our analysis starts with an evaluation of economic efficiency under uncertainty and learning. For that purpose, we will make use of a reference bundle of goods $g=\left(0,g_2,\cdots,g_m\right)\in\mathbb{R}_+^m$ satisfying $g\ne 0$. The bundle g is defined to be a non-negative bundle of consumer goods. Let $\overline{g}=\left[g\left(s\right):g\left(s\right)=g,s\in S\right]\in\prod\nolimits_{s\in S}\mathbb{R}^m$. The bundle $\overline{g}$ is a sure bundle as it is the same across all states $s\in S$.

We make the following assumptions.

Assumption As1: There exists a feasible point $\left(\tilde{x}_0,\tilde{y}_0,z_0\right)$ such that the set $Y_c\left(z_0\right)-\sum\nolimits_{i\in N}X_{ci}\left(z_0\right)$ has a non-empty interior.

---

[8] Equation (4) states the supply-demand balance of goods at the aggregate level. Note that it implicitly treats individual labor $\tilde{x}_{i1}$ as an undifferentiated good. Treating labor as a differentiated product could be done easily by considering the case where there are J types of labor, each typedefined according to labor productivity. Then, partitioning the set of individuals N according to labor productivity $N=\left(N_1,\cdots,N_J\right)$ and letting $\tilde{y}_1=\left(\tilde{y}_{1k}:k=1,\cdots,J\right)$, Equation (4) would take the form:

$$\sum\nolimits_{i\in N_k}\tilde{x}_{i1}\le\tilde{y}_{1k},\ k=1,\cdots,J, \tag{4a'}$$

$$\sum\nolimits_{i\in N}\tilde{x}_{ij}\le\tilde{y}_j,\ j>1. \tag{4b'}$$

Assumption As2: As a representation of preferences for the *i*-th individual, $i \in N$, the ordinal utility function $u_i : X_i \times \mathbb{R}^K \to \mathbb{R}$ is continuous and satisfies $u_i(\tilde{x}_i + k\overline{g}, z_i) > u_i(\tilde{x}_i, z_i)$ for all $(\tilde{x}_i, z_i) \in X_i \times \mathbb{R}^K$ and all $k > 0$.

Assumption As1 guarantees the existence of a feasible allocation. It also assumes that the feasible set $Y_c(z_0) - \sum_{i \in N} X_{ci}(z_0)$ has an interior point[9]. Assumption As2 is a form of non-satiation in consumer preferences: each consumer always prefers to have more of the bundle of goods $\overline{g}$. Note that we do not assume that the utility function is quasi-concave, or that the set $\{(\tilde{x}_i', z_i') : u_i(\tilde{x}_i', z_i') \geq u_i(\tilde{x}_i, z_i)\}$ is convex. This means that we do not require consumers to be uncertainty averse. (Note that this differs from Debreu (1959: p. 101) or Yaari (1969: p. 315) who assumes that all consumers are uncertainty averse). We will explore below the linkages between this assumption and the validity of Equation (1c).

Note the generality of the approach. We allow for flexible information structures: information can vary across individuals; and it can vary over time when the analysis is applied to temporal allocations. And our analysis allows for a general representation of the technology used in both production and learning activities. Our investigation allows for active learning, including both individual learning and social learning (with information externalities across individuals). Going beyond Debreu (1959), we allow for externalities as well as nonconvexity. Externalities occur under social learning. And nonconvexity arises in the presence of externalities (Starrett, 1972; Chavas & Hall, 2017), fixed cost (Radner, 1968) and/or increasing returns (Arthur, 1989). In addition, learning and production are part of a joint technology. The jointness arises under learning-by-doing (when learning includes experimentation among production activities). And we allow the productivity of any netput to depend on all inputs used and all outputs produced in the economy. In a multi-firm economy, this allows for production externalities which would arise when production decisions made by a firm affect the productivity of other firms[10]. As discussed below, conducting our study without convexity assumptions has significant implications for welfare analysis and the evaluation of uncertainty.

4. Valuation under Uncertainty: The Shadow Pricing of State-Contingent Goods

Building on Section 3, this section characterizes efficiency as well as the shadow prices of state-contingent goods under general conditions. We use the classical concept of Pareto efficiency: a feasible allocation is Pareto efficient if there does

---

[9] This is a constraint qualification imposed to support a Lagrangian approach in constrained optimization (Slater, 1950).

[10] In an economy involving *T* firms, each firm producing $y_t \in \mathbb{R}^m$, $t = 1, \cdots, T$, total production is given by $y = \sum_{t=1}^{T} y_t \in Y(z)$. In general, this allows the decisions made by each firm to affect the productivity of others. The absence of externalities across firms is obtained as a special case where $y_t \in Y_t(z)$ and $y = \sum_{t=1}^{T} y_t \in \sum_{t=1}^{T} Y_t(z)$ where $Y_t(z)$ is the feasible set for the *t*-th firm, $t = 1, \cdots, T$.

not exist another feasible allocation that can make one individual better off without making others worse off. As discussed below, our analysis applies to market economies as well as nonmarket economies (e.g. in situations where contingent markets fail to develop due to asymmetric information). To investigate when an allocation is efficient, our analysis makes use of the benefit function as a measure of willingness-to-pay. For the *i*-th individual and a given $z \in Z$, following Luenberger (1992a), the benefit function is defined as:

$$b_i(\tilde{x}_i, z, U_i) = \max_\beta \left\{ \beta : u_i(\tilde{x}_i - \beta \overline{g}, z_i) \geq U_i, (\tilde{x}_i - \beta \overline{g}) \in X_{ic}(z) \right\}$$

if a maximum exists.

$= -\infty$ otherwise.

The benefit function $b_i(\tilde{x}_i, z, U_i)$ measures the number of units of the sure reference bundle $\overline{g}$ the *i*-th consumer is willing to give up in order to reach utility level $U_i, i \in N$. In general, $b_i(\tilde{x}_i, z, U_i)$ is a welfare measure for the *i*th individual, $i \in N$. Indeed, in the case where the bundle $\overline{g}$ is chosen such that one unit of $\overline{g}$ is worth \$1, $b_i(\tilde{x}_i, z, U_i)$ becomes a monetary evaluation of the *i*-th individual's willingness-to-pay starting at point $(\tilde{x}_i, z)$ to obtain utility level $U_i, i \in N$. The properties of the benefit function have been investigated by Luenberger (1992a). One of these is the translation property: for $\tilde{x}_i \in X_i$ and $\tilde{x}_i + \alpha \overline{g} \in X_i$, it satisfies $b_i(\tilde{x}_i + \alpha \overline{g}, z, U_i) = \alpha + b_i(\tilde{x}_i, z, U_i)$.

Consider the optimization problem:

$$\begin{aligned} W(U) \equiv \max_{\tilde{x}, \tilde{y}, z} \Big\{ & \textstyle\sum_{i \in N} b_i(\tilde{x}_i, z, U_i) : \sum_{i \in N} \tilde{x}_i \leq \tilde{y}, \tilde{x}_i \in X_{ic}(z), i \in N; \\ & \tilde{y} \in Y_c(z), z \in Z \Big\} \end{aligned} \quad (5)$$

where $U \equiv \{U_i : i \in N\}$ are reference utility levels for the *n* individuals. Conditional on *U*, Equation (5) maximizes aggregate willingness-to-pay subject to feasibility. As discussed below, Equation (5) will play an important role in our evaluation of economic efficiency under uncertainty. Conditional on *U*, let $(\tilde{x}^*(U), \tilde{y}^*(U), z^*(U))$ be a solution to (5), where $\tilde{x}^*(U) = \{x_i^*(s, U) : s \in S, i \in N\}$, $x_i^*(s, U)$ denoting the optimal $x_i(s)$ for the *i*-th individual under state *s*, $i \in N$, $s \in S$.

$W(U)$ being conditional on U in (5) raises the question: how to choose *U*? Next, we consider the following feasible set $\mathbb{U}$ for *U*:

$$\mathbb{U} \equiv \left\{ U : W(U) \geq 0; \tilde{x}_i^*(U_i) - \epsilon \overline{g} \in X_i(z) \text{ for all } i \in N, z \in Z \right\} \quad (6)$$

where $\epsilon \in \mathbb{R}_+$ is a small positive number.

Below, we consider the case where $U \in \mathbb{U}$. We make the following additional assumption:

<u>Assumption As3</u>: $\mathbb{U} \neq \varnothing$.

Assumption As3 implies the existence of utilities $U = \{U_i : i \in N\}$ satisfying two conditions: 1) fiscal feasibility represented by anon-negative aggregate willingness-to-pay: $W(U) \geq 0$; and 2) as stated in (6), it is possible for each individual to giveup some goods $\overline{g}$, thus ruling out situations of completed estitution. Note that $U \in \mathbb{U}$ under As3 still allows for a wide distribution of purchasing power across individuals (including situations of poverty where some individuals have limited purchasing power).

The strong linkages between Equations (5) and (7) and Pareto efficiency are presented next. (All proofs are presented in **Appendix**.)

Proposition 1: Under Assumptions As1 - As3, a feasible allocation $(\tilde{x}, \tilde{y}, z)$ is Pareto efficient if and only if it satisfies (5) and:

$$W(U) = 0, U \in \mathbb{U}. \tag{7}$$

Proposition 1 states that a Pareto efficient allocation if and only it maximizes aggerate willingness-to-pay in (5) and $U \in \mathbb{U}$ satisfies $W(U) = 0$. First, any allocation not maximizing aggregate willingness-to-pay cannot be Pareto efficient: it would imply the existence of a Pareto improving move that can make an individual better without making anyone else worse off. Second, $W(U) \geq 0$ can be interpreted as an aggregate budget constraint reflecting fiscal feasibility. In this context, Equation (7) states that, under efficiency, the aggregate budget constraint must be binding.

Proposition 1 associates Pareto efficiency under uncertainty with the constrained maximization problem (5). This maximization problem has a dual formulation that can be used to evaluate state-contingent prices under uncertainty. To define this dual formulation, let $H : \prod_{s \in S} \mathbb{R}^m \to \mathbb{R}$ be the set of continuous and non-decreasing functions with the properties that $h \in H$ satisfies $h(0) = 0$ along with the translation property: $h(\tilde{y} + \alpha \bar{g}) = \alpha + h(\tilde{y})$ for all finite α. When $\bar{g}$ is chosen such that one unit of $\bar{g}$ is worth \$1, the function $h \in H$ provides a monetary measure that we will use to evaluate state-contingent prices.

Consider the functional $L : X \times F \times H \times \mathbb{U}$:

$$L(\tilde{x}, \tilde{y}, z, h, U) = \sum_{i \in N} b_i(\tilde{x}_i, z, U_i) + h(\tilde{y}) - h\left(\sum_{i \in N} \tilde{x}_i\right) \tag{8}$$

where $\tilde{x} \in X_c(z), \tilde{y} \in Y_c(z), z \in Z, h \in H, U \in \mathbb{U}$. Equation (8) is a generalized Lagrangian as the function $h \in H$ is allowed to be nonlinear (Gould, 1969; Chavas & Briec, 2012). Indeed, Equation (8) would reduce to the standard Lagrangian when h is taken to be linear, with $h(\tilde{y}) = \lambda \tilde{y}$, λ being standard Lagrange multipliers. The standard Lagrangian would apply under convexity, where the separating hyperplane theorem implies the existence of a hyperplane with slopes given by the Lagrange multipliers (e.g. Rockafellar, 1974a: p. 45; Luenberger, 1997: pp. 217-221). But the separating hyperplane theorem does not apply under nonconvexity. Since we do not assume convexity, this motivatesus to consider the generalized Lagrangian (8) where we allow h to be nonlinear, in which case h representsa separating hypersurface (Chavas & Briec, 2012).

In the general case, the dual formulation associated with (5) is given by a saddle-point of the generalized Lagrangian in (8). For given $U \in \mathbb{U}$, let

$\left(\tilde{x}^* \in X_c(z), \tilde{y}^* \in Y_c(z), z^* \in Z, h^* \in H\right)$ be a saddle point of $L(\tilde{x}, \tilde{y}, z, h, U)$ in (8) satisfying:

$$L(\tilde{x}, \tilde{y}, z, h, U) \leq L\left(\tilde{x}^*, \tilde{y}^*, z^*, h^*, U\right) \leq L\left(\tilde{x}^*, \tilde{y}^*, z^*, h, U\right) \tag{9}$$

for all $\tilde{x} \in X_c(z)$, $\tilde{y}_c \in Y_c(z)$, $z \in Z$ and $h \in H$. There are strong linkages between the primal problem (5) and the dual problem (9). Letting $L^*(U) \equiv L\left(\tilde{x}^*, \tilde{y}^*, z^*, h^*, U\right)$, this is formally stated next.

Proposition 2: For a given $U \in \mathbb{U}$, let $\left(\tilde{x}^*, \tilde{y}^*, z^*, h^*\right)$ be a saddle point of the generalized Lagrangian in (8)-(9). Then, $\left(\tilde{x}^*, \tilde{y}^*, z^*\right)$ is a solution to (5) and satisfies $W(U) = L^*(U)$.

Proposition 2 states that Equation (9) is a dual formulation of the primal problem in (5), and that the strong duality result holds: $W(U) = L^*(U)$. Importantly, Proposition 2 holds under convexity as well as non-convexity but with one important difference. As noted above, under convexity, the separating hyperplane theorem holds, meaning that the function h can be taken to be linear and the saddle-point problem in (9) reduces to the standard Lagrangian approach However, for non-convex problems, the separating hyperplane theorem no longer applies, and there is need to allow h to be nonlinear in the dual formulation (9). As discussed below, the nonlinearity of h will have important implications for economic assessment under uncertainty.

While Proposition 2 states that the dual problem (8)-(9) is sufficient to solve the primal problem (5), it is not necessary. Indeed, to guarantee that the dual problem in (8)-(9) solves the primal problem (5), some regularity conditions are needed. As explored in previous literature (e.g. Rubinov et al., 2002; Chavas & Briec, 2012), such conditions involve the perturbation function defined as:

$$\begin{aligned} W_0(U,d) \equiv \max_{\tilde{x},\tilde{y},z} \Big\{ & \textstyle\sum_{i\in N} b_i\left(\tilde{x}_i, z, U_i\right) : \sum_{i\in N} \tilde{x}_i \leq \tilde{y} + d, \tilde{x}_i \in X_{ic}(z), i \in N; \\ & \tilde{y} \in Y_c(z), z \in Z \Big\} \end{aligned} \tag{10a}$$

with associated generalized Lagrangian

$$L_0(\tilde{x}, \tilde{y}, z, h, U, d) = \textstyle\sum_{i\in N} b_i\left(\tilde{x}_i, z, U_i\right) + h(\tilde{y} + d) - h\left(\sum_{i\in N} \tilde{x}_i\right) \tag{10b}$$

where changing $d \in \prod_{s\in S} \mathbb{R}^m$ reflects a change in resource availability. As showed by Chavas and Briec (2012), obtaining strong duality (with $W(U) = L^*(U)$) requires the perturbation function $W_0(U,d)$ to be upper-semicontinuous in d at $d = 0$. Noting that this condition is not overly restrictive, we assume that it is satisfied in the rest of our analysis. In this context, Equations (7) and (9) provide a dual representation of a Pareto efficient allocation.

The linkages between the dual Problem (9)-(10) and the valuation of state contingent goods are presented next. Applied to the generalized Lagrangian $L(\tilde{x}, \tilde{y}, z, h, U)$ in (8)-(9), Chavas and Pagani (2020) obtained the following result.

Lemma 1: For given $U \in \mathbb{U}$, let $(\tilde{x}_d, \tilde{y}_d, z_d, h_d)$ be asaddle point of the generalized Lagrangian in (10b) evaluated at *d*. Consider a change from *d* to *d'*. Then, the following inequalities hold:

$$\begin{aligned} h_d(\tilde{y}_{d'} + d') - h_d(\tilde{y}_{d'} + d) &\geq W_0(U, d') - W_0(U, d) \\ &\geq h_{d'}(\tilde{y}_d + d') - h_{d'}(\tilde{y}_d + d) \end{aligned} \tag{11}$$

where $\left[h_{d'}(\tilde{y}_{d'} + d') - h_{d'}(\tilde{y}_d + d)\right] \geq 0$ when $d' \geq d$.

Lemma 1 involves $\left[W_0(U, d') - W_0(U, d)\right]$ which can be interpreted as marginal willingness-to-payassociated with a change in *d*. It follows from Equation (11) that this marginal willingness-to-pay can be measured by the slope of *h*. Importantly, this result holds under general conditions: it holds without Fréchet differentiability, and it holds under nonconvexity. When the function $h_d(\tilde{y}_{d'} + d)$ is Gateaux differentiable at *d*, its Gateaux derivative at *d* in the direction *v* is $Dh_d(\tilde{y}_{d'} + d, v) = \lim_{\delta \to 0}\left\{\frac{h_d(\tilde{y}_{d'} + d + \delta v) - h_d(\tilde{y}_{d'} + d)}{\delta}\right\}$ where $v \in \prod_{s \in S} \mathbb{R}^m$. Then, Equation (11) implies that:

$$DW_0(U, d, v) = Dh_d(\tilde{y}_{d'} + d, v) \geq 0. \tag{12}$$

Equations (11)-(12) make it clear that the hypersurface defined by $h_d(\cdot)$ has an important economic interpretation: its slope measures the marginal willingness-to-pay (or shadow prices) of state-contingent goods. In the special case where $h_d(\tilde{y}_{d'} + d)$ is Fréchet differentiable in *d*, then (12) becomes $Dh_d(\tilde{y}_{d'} + d, v) = \frac{\partial h_d(\tilde{y}_{d'} + d)}{\partial d}$ for all *v*. If in addition h_d is linear, then $\frac{\partial h_d(\tilde{y}_{d'} + d)}{\partial d}$ would reduce to the standard Lagrange multipliers with their standard interpretation of measuring the (global) shadow prices of the constraints (e.g. Rockafellar, 1974a: p. 45; Luenberger, 1997: pp. 217-221). From (11)-(12), this interpretation continues to hold under nonconvexity, but with an important difference: when *h* is nonlinear, its slopes are no longer constant, meaning that the shadow price interpretation is valid only locally. As we will see below, this argument will play an important role in examining the validity of probabilities in uncertainty assessment.

Note that Equation (9) provides useful insights into the choices of $(\tilde{x}, \tilde{y}, z, h)$. Indeed, following Chavas and Briec (2012), Equation (9) can be decomposed as follows:

$$\begin{aligned} E(h, z, U) &= \inf_{\tilde{x}}\left\{h\left(\sum_{i \in N} \tilde{x}_i\right) - \sum_{i \in N} b_i(\tilde{x}_i, z, U_i) : \tilde{x} \in X_c(z)\right\} \\ &= \inf_{\tilde{x}}\left\{h\left(\sum_{i \in N} \tilde{x}_i\right) : u_i(\tilde{x}_i, z) \geq U_i, i \in N; \tilde{x} \in X_c(z)\right\} \end{aligned} \tag{13a}$$

$$\pi(h, z, U) = \sup_{\tilde{y}}\left\{h(\tilde{y}) : \tilde{y} \in Y_c(z)\right\} \tag{13b}$$

$$L^*(U) = \inf_{h \in H}\left\{\sup_{z \in Z}\left\{\pi(h, z, U) - E(h, z, U)\right\}\right\}, \tag{13c}$$

where $E(h, z, U)$ in (13a) is an aggregate consumer expenditure function and

$\pi(h,z,U)$ in (13b) is an aggregate profit function. From Corollary 1, choosing $U \in \mathbb{U}$ to satisfy $L^*(U)=0$ means that the decisions described by Equations (13a)-(13c) support an efficient allocation.

Equations (13a)-(13b) show that maximizing aggregate profit and minimizing aggregate consumer expenditure are integral parts of efficiency evaluation. This result holds under general conditions, includes situations of uncertainty with heterogenous information across decision makers, heterogenous consumer preferences, a general technology, active learning under costly information, and the presence of information externalities.

Finally, the analysis applies to markets as well as nonmarket allocations. Indeed, Equation (13c) shows how the pricing scheme $h \in H$ is chosen. In general, from (11)-(12), the slopes of h can be interpreted as shadow prices of the state-contingent goods $\tilde{y}$. These prices would become market prices if markets for the state-contingent goods $\tilde{y}$ are present (e.g. Debreu, 1959). But, as showed by Radner (1968), the presence of asymmetric information among traders can severely limit the development of state-contingent markets (as state-contingent trade can only develop among traders who have access to the same information). This is the main reason why state-contingent markets (e.g. insurance markets) are not very common. Butfrom Equations (11)-(12), interpreting the slopes of $h \in H$ as shadow prices of state-contingent goodsremains valid under general conditions. This argument applies to situations where there are (implicit or explicit) contracts among individuals in households, in firms, in social networks, or in society. The implications of these results for uncertainty assessments are discussed next.

5. Implications for the Validity of Probabilities

Yaari (1969) proposed to measure probabilities as normalized prices of state-contingent goods. As noted in the introduction, this is problematic since risk markets are typically incomplete. But we have just derived a way to assess the shadow prices of state-contingent goods under general conditions, including absent risk markets. In this section, following Yaari (1969), we propose to evaluate probabilities as (normalized) shadow prices. We explore the properties of these shadow prices and their implications for probability theory. A key focus is to investigate the validity of the additivity property of probabilities stated in Equation (1c) (Axiom Ax3).

Our analysis examines two broad situations: 1) the case where uncertainty is subject to passive learning; and 2) the case where uncertainty is subject to active learning. We define passive learning as situations where individual access to information does not depend on information gathering activities *z*. In contrast, we define active learning as situations where individual access to information requires the use of information gathering activities *z*. In this context, we assume the absence of learning corresponds to $z=0$. We start with the case of passive learning which is simpler to analyze.

Under passive learning, the learning process exhibits two properties: 1) information is provided by processes that are beyond the control of each individual; and 2) individuals would choose $z = 0$ (as they have no incentive to invest their time in the learning process). This means that the information structure of society is still given by P, which allows information to vary across individuals and over time, but P is determined exogenously. Situations of passive learning are illustrated in **Figure 1** (under convexity) and **Figure 2** (under non-convexity).

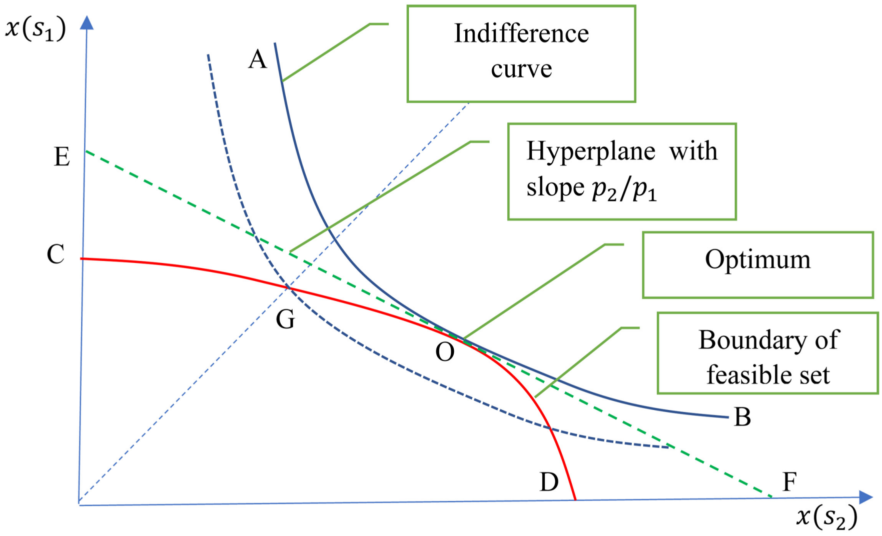

Figure 1. Convex case under passive learning.

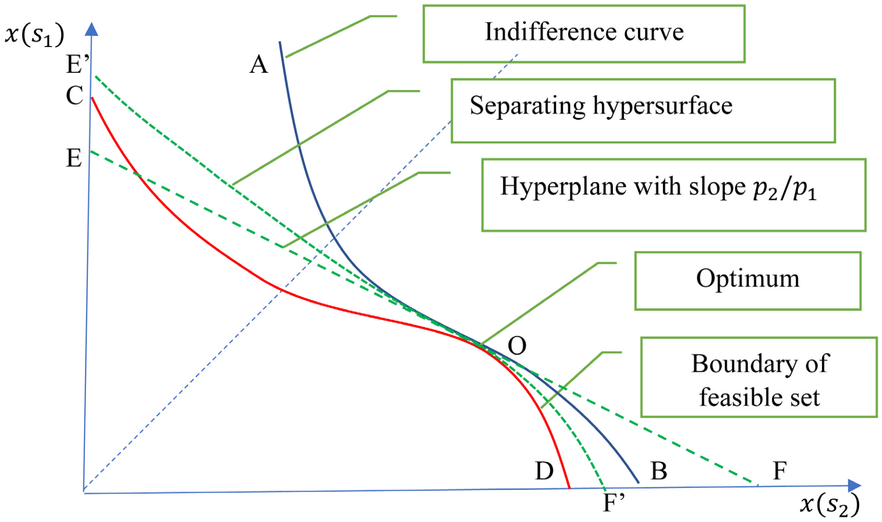

Figure 2. Non-convex case under passive learning.

Figure 1 presents a situation of passive learning in the simple case where $n = 1$, $m = 1$, $S = \{s_1, s_2\}$, the production technology is convex (i.e. reflecting diminishing marginal productivity), and the set $\{(\tilde{x}_i', z') : u_i(\tilde{x}_i', z') \geq u_i(\tilde{x}_i, z)\}$ is convex (which corresponds to uncertainty aversion, as assumed in Debreu, 1959 and Yaari, 1969). Under convexity, in a way consistent with the separating hyperplane theorem, **Figure 1** shows the existence of a separating hyperplane given by the line EOF (corresponding to h being linear in our dual formulation), where point O is the efficient point. In this case, point O is off the 45-degree line, meaning that it is efficient to have state-contingent decisions $x(s)$ that vary with s. Indeed, choosing the sure point G (located on the 45-degree line where $x(s_1) = x(s_2)$) would be inefficient: it would generate lower consumer utility. And the slopes of the hyperplane are the Lagrange multipliers measuring the shadow price of the state-contingent good: (p_1, p_2). Following Yaari (1969), normalizing these shadow prices provides a measure of the associated probabilities: $Prob(s_1) = \frac{p_1}{p_1 + p_2}$ and $Prob(s_2) = \frac{p_2}{p_1 + p_2}$. These probabilities satisfy Kolmogorov's Axioms Ax1 - Ax3. Thus, under passive learning and convexity (including uncertainty aversion), **Figure 1** presents a situation where there is no inconsistency with Kolmogorov's probability axioms. This is stated formally next:

Result 1: Under passive learning, uncertainty aversion and a convex technology, there are no inconsistencies with Kolmogorov's probability Axioms Ax1 - Ax3.

Note that this positive result needs to be interpreted with caution. First, while passive learning allows information to vary across individuals and over time, it treats the information structure P in society as exogenous. This cannot be fully satisfactory in the sense that passive learning provides no insight into the learning process. This is problematic to the extent that Axioms Ax1 - Ax3 have also been used as foundations to motivate the Bayesianupdating of probabilities (e.g. DeGroot, 1970). In addition, passive learning and convexity (assumed in **Figure 1**) will prove to be the only situation where there is no inconsistency with Kolmogorov's probability axioms, as discussed next.

Figure 2 presents a situation of passive learning similar to **Figure 1**, but with two important differences: it does not assume that the production technology is convex (i.e. it no longer assumes that the technology necessarily exhibits diminishing marginal productivity); and it does not assume that the set $\{(\tilde{x}_i', z') : u_i(\tilde{x}_i', z') \geq u_i(\tilde{x}_i, z)\}$ is convex (i.e. it allows for uncertainty-loving behavior). Without these convexity assumptions, the separating hyperplane theorem no longer applies. Indeed, **Figure 2** shows that point O is efficient, but the hyperplane given by the line EOF is no longer a separating hyperplane: the line EOF enters the feasible set when $x(s_1)$ is high (a region where the technology exhibits increasing marginal productivity); and the line EOF crosses the indifference curve AOB when $x(s_2)$ is high (a region where the individual exhibits uncertainty-loving behavior). Yet, in **Figure 2**, there is a separating nonlinear hypersurface (corresponding to h in our dual formulation) given by the line E'OF'. At point O, this line is tangent to both the boundary of the feasible set and the indifference curve. But its slopes (measuring shadow prices) change with the evaluation point: it is steeper when $x(s_2)$ is small but less steep when $x(s_2)$ is large. From (11)-(12), this means that we have nonlinear pricing of the state-contingent goods. Again, the subjective probabilities can be evaluated as normalized shadow prices, with one important difference: the shadow prices being nonlinear imply that the associated probabilities would also be nonlinear. In this case, our evaluation of uncertainty motivated by willingness-to-pay measures would fail to satisfy Kolmogorov's Axiom Ax3. What would happen if we were to force $h$to be linear (so that its slopes would remain constant as implicitly assumed in Axiom Ax3)? In **Figure 2**, this would amount to forcing a switch from the hypersurface E'OF' to the hyperplane EOF. Note that the two lines have the same slope at point O. But using (13b), switching to the hyperplane EOF would induce profit-maximizing producers to switch from the efficient point O to the inefficient point C. In other words, imposing Axiom Ax3 would generate an inefficient allocation. The reason for this inefficiency arises from the need to consider nonlinear shadow pricing under nonconvexity. In **Figure 2**, the nonconvexity comes from two sources: 1) from nonconvex technology where diminishing marginal productivity does not hold (e.g. under increasing returns); and 2) from consumer preferences when we allow for the presence of uncertainty-loving behavior. These results are summarized next.

Result 2: A nonconvex technology and/or uncertainty-loving behavior contribute to generate inconsistencies with Kolmogorov's probability Axiom Ax3.

The inconsistencies noted in Result 2 reflect that the behavioral motivations for Axiom Ax3 have not been fully explored in previous literature. Note that result 2 identifies the role of uncertainty-loving behavior. This may not be serious issue to the extent that risk aversion is a rather common characteristic of human behavior (Gollier, 2004). But Result 2 indicates that the presence of a nonconvex technology canraise questions about the general validity of Axiom Ax3. Indeed, nonconvexity provides the motivation to switch from a separating hyperplane in **Figure 1** to a nonlinear separating hypersurface in **Figure 2**.

Next, we investigate how situations of active learning uncover additional reasons why Axiom Ax3 may be inappropriate. Situations of active learning (where individual access to information requires the use of information gathering activities) are illustrated in **Figure 3**. Like **Figure 1**, **Figure 3** presents a simple case where $n = 1$, $m = 1$, $S = \{s_1, s_2\}$, but with a special focus on the role of active learning. Under active learning, information is costly which affects the learning process. In **Figure 3**, there are "fixed costs" in learning corresponding to thresholds indicating that no learning can take place below some minimum amount of resources committed to the learning process[11]. **Figure 3(a)** reports a situation where information cost is low and it is efficient to learn, while **Figure 3(b)** shows a situation where information cost is high and it is optimal not to learn.

---

[11] This is consistent with the evidence of thresholds in neuroscience, where neural and cognitive responses take place only beyond some minimal stimulus levels (e.g. Gazzaniga et al., 2002; Quiroga et al., 2008).

Figure 3(a) and Figure 3(b) illustrate how information cost affects decisions under uncertainty. In Figure 3(a), the boundary of the feasible set is given by the line CO'OD. Note that point O' is on the 45-degree line where $x(s_1) = x(s_2)$: at this point, $x(s)$ is chosen *ex ante* and no information is used in decision making. It means that point O' can be obtained with no information gathering activities: $z = 0$. Except for points along the 45-degree line, the boundary of the feasible set along the lime COD are below point O', reflecting that active learning requires the use of information gathering activities z that divert resources away from the production of consumer goods. In this case, fixed costs in learning make production technology non-convex. Indeed, the role of costly information in creating nonconvexity in production technology is the main difference between Figure 1 and Figure 3(a) and Figure 3(b). Figure 3(a) shows that point O (where state-contingent decisions are made under learning) is efficient: point O generates a higher level of utility than point O' (where there is no learning). In this case, the information produced under active learning helps improve the decision-making process so that the benefit of new information is greater than its cost. In other words, learning is optimal in Figure 3(a). Figure 3(a) shows the hyperplane EOF that goes point O and is tangent to the indifference curve AOB.

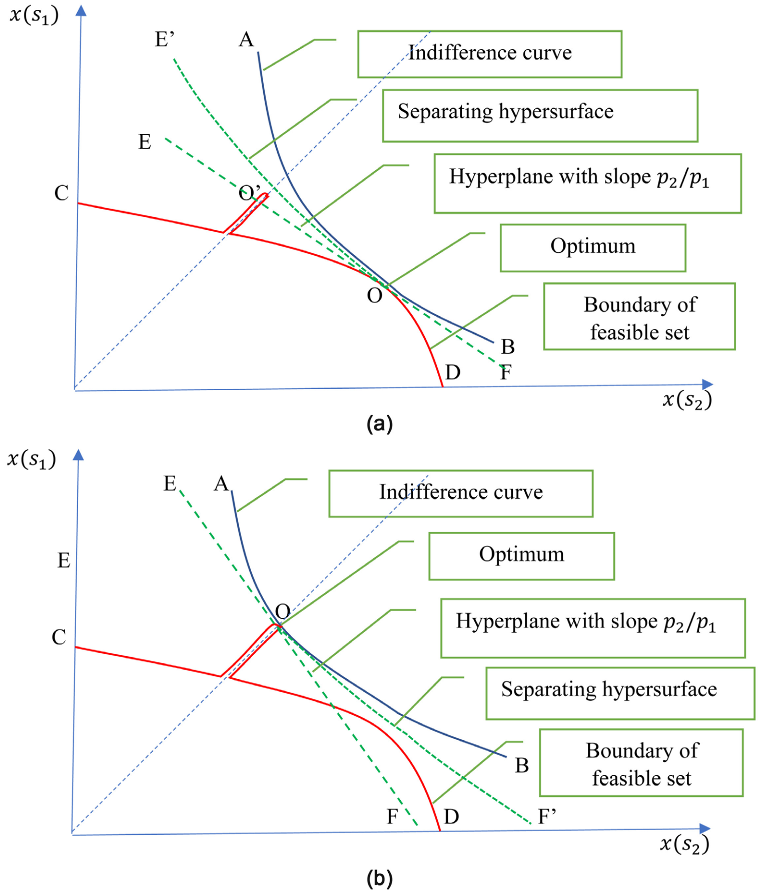

Figure 3. (a): Active learning: The case where learning is optimal. (b): Active learning: The case where no learning is optimal.

Importantly, this hyperplane is not a separating hyperplane due to a nonconvex feasible set: the line AOB enters the interior of the feasible set as it cuts the 45-degree line below point O'. **Figure 3(a)** also reports a separating nonlinear hypersurface E'OF (representing *h* in our dual formulation) that is also tangent to the indifference curve AOB. Using (13b), note that E'OF provides the proper incentives to produce at point O; this contrasts with EOF which would provide producers the incentive to produce at the inefficient point O'. It means that forcing *h* to be linear would be inefficient. In other words, supporting an efficient allocation requires that state-contingent shadow prices be nonlinear. This means that probabilities (measured as normalized shadow prices) must be nonlinear, which is inconsistent with Ax3.

Figure 3(b) is similar to **Figure 3(a)**: they both illustrate the role of information cost under active learning. But there is one important difference: in **Figure 3(b)**, the cost of information is higher, and it is nolonger optimal to learn. Indeed, the efficient point O in **Figure 3(b)** is now located along the 45-degree line where $x(s_1)=x(s_2)$, i.e. where $x(s)$ is chosen *ex ante* without learning. **Figure 3(b)** that there is a hyperplane EOF going through point O and tangent to the indifference curve AOB, but this is not a separating hyperplane (as the line EOF enters the feasible set for large values of $x(s_2)$). Again, there is a nonlinear separating hypersurface EOF' (representing *h* in our dual formulation), which provides valid measures of state-contingent shadow prices. Thus, under active learning, we obtain results qualitatively similar to those obtained in **Figure 3(a)**: supporting an efficient allocation requires that state-contingent shadow prices be nonlinear. Measuring probabilities as normalized shadow prices, it follows that probabilities must be nonlinear, which is inconsistent with Ax3. Note that this result holds whether a decision maker decides to learn (in **Figure 3(a)**) or not (in **Figure 3(b)**). This is summarized next.

Result 3: Under active learning, costly information contributes to generate inconsistencies with Kolmogorov's probability Axiom Ax3. This result holds whether or not a decision maker decides to learn.

Results 2 and 3 show that uncertainty-loving behavior and costly information lead to inconsistencies with Kolmogorov's Axiom Ax3. Importantly, these inconsistencies are not due to ambiguity aversion (Chavas & Briec, 2019). Rather, they reflect more fundamental issues related to the validity of probabilities as representations of uncertainty.

Result 3 raises significant concerns about the validity of Axiom Ax3. Indeed, obtaining information always requires some effort and resources (including time commitment), i.e. learning is never costless. While improvements in information technology have greatly reduced information costover the last few decades, learning is typically a slow process for three reasons: 1) limited time availability restricts how much each individual can spend learning; 2) cognitive ability and imperfect memory constrain each individual's ability to process information (Simon, 1955; Bowers & Davis, 2012); and 3) in a complex world, uncertainty comes from many sources, possibly leading to "information overload".

This last factor raises a related question: does our analysis provide insights into defining Axiom Ax3 in the context of a finite state space versus a countably infinite state space? As discussed in Section 2, this issue arose between Kolmogorov (who argued in favor of a countable state space) and Savage (who argued in favor of a finite state space). Next, we examine this issue by considering the following constrained optimization problem.

Problem 1: Following Karney (1981) and Martin et al. (2016), consider the optimization problem:

$$V = \min_x \left\{ x_1 : -1 - x_1 \le 0; x_2 \le 0; \frac{x_2}{j} - x_1 \le 0,\ j = 3, \cdots, m; x \in \mathbb{R}^2 \right\}. \tag{14}$$

When m is finite, (14) has for solution $x^* = (-1, -m)$ with $V = -1$. Then, (14) is a simple linear programming problem. Its feasible set has a non-empty interior (e.g. $x = (1, -1)$, implying that Problem (14) satisfies Assumption As1. Its associated dual formulation is:

$$L(x, h) = x_1 + h_1(-1 - x_1) + h_2(x_2) + \sum_{j=3}^{m} h_j \left(\frac{x_2}{j} - x_1 \right) \tag{15}$$

where $h \in H$. Taking h to be linear (with λ as associated Lagrange multipliers), strong duality holds and a saddle-point of the Lagrangian (15) is given by $x^* = (-1, -m)$, $\lambda = (1, 0, \cdots, 0)$ and $L^* = -1$.

Now, consider Equations (14)-(15) when $m \to \infty$. In this infinite dimensional case, Problem (14) has for solution $x^* = (0, 0)$ with $V = 0$. But its dual formulation becomes more complex. First, taking h to be linear (with λ as associated Lagrange multipliers), strong duality no longer holds: a saddle-point of the Lagrangian (15) would occur at $x^* = (0, 0)$, $\lambda = (1, 0, 0, \cdots)$ and $L^* = -1 < V = 0$. As discussed in Martin et al. (2016), this failure of duality occurs because the dual space is not countably additive. But this issue is resolved if we consider the case where h is nonlinear and satisfies $h(c) = \lambda_1 c_1 + \lambda_2 c_2 + \sum_{j \ge 3} \lambda_{aj} c_j I(c_j < 0) + \sum_{j \ge 3} \lambda_{bj} c_j I(c_j > 0)$ where I is an indicator function[12]. In this case, strong duality is restored: a saddle-point of the Lagrangian (15) occurs at $x^* = (0, 0)$, $(\lambda_1, \lambda_2) = (0, 0)$, $\lambda_{aj} = 0$, $\lambda_{bj} \ge 0$, and $\sum_{j \ge 3} \lambda_{bj} = 1$, with $L^* = 0 = V$.

This example illustrates the effects of introducing an infinite state space in a constrained optimization problem. It shows two results: 1) if we insist on keeping h linear, strong duality can fail as the dual space is no longer countably additive; 2) restoring strong duality requires nonlinear h. Interpreting the slopes of h as shadow prices, this means that analyzing infinite space can require nonlinear shadow prices. Applying these results to situations under uncertainty, it follows that probabilities (defined as normalized shadow prices of state-contingent goods)

---

[12] In this case, $h(c)$ is Gateaux differentiable, but not Fréchet differentiable. This indicates that taking $h(c)$ to be a quadratic function (e.g. Rockafellar, 1974b) would be inappropriate.

can fail to satisfy countable additivity (as stated in Axiom Ax3) when the state space becomes infinite. This result is summarized next.

Result 4: Having an infinitely countable state space contributes to generate inconsistencies with Kolmogorov's probability Axiom Ax3.

Result 4 indicates that the additivity property of probabilities (stated in Axiom Ax3) can become problematic when the number of states becomes infinite. This result could be used in favor of assuming finite additivity in Axiom Ax3. Indeed, it would boost the arguments presented by Savage (1954) and de Finetti (1974) in favor of finite additivity. More generally, working with a finite number of states helps simplify the empirical analysis of uncertain outcomes. Yet, in a complex world, limiting the analysis to a finite number of states seems restrictive. In this context, Result 4 creates additional challenges to Axiom Ax3, its validity and its use in the assessment of uncertainty.

6. Conclusion

Our analysis has examined the validity of probabilities in the assessment of uncertainty, with a focus on the role of learning. Using a general state-contingent approach, our arguments involved uncertainty assessments based on willingness-to-pay measures across states of nature. By not requiring probabilities, our approach allowed us to investigate the validity of probabilities. More specifically, we questioned the validity of a key assumption in probability theory: Kolmogorov's additivity axiom (stated as Axiom Ax3 in Section 2). We identified three sets of factors contributing to non-additive probabilities: 1) uncertainty-loving behavior, 2) active learning under costly information, and 3) the presence of an infinite number of states. These are important results: relaxing additivity in Axiom Ax3 would undermine Bayes theorem and its use in statistical analysis.

One of our key results states that the additivity Axiom Ax3 is problematic under costly learning. This result seems particularly important: it underlines the complexity of human learning. But it also makes the analysis of uncertainty assessment challenging. Our findings reflect that Bayesian learning has one key drawback: it fails to address the role of perceptual and cognitive limitations (Bowers & Davis, 2012). Our analysis indicates the need to establish closer linkages between uncertainty assessments, cognition and the way our brain processes information. As noted in the introduction, the evidence from neuroscience indicates that neurons exhibit threshold responses to signals (Gazzaniga et al., 2002; Quiroga et al., 2008) and that responses from multiple signals are often non-additive (e.g. Holmes & Spence, 2005; Stanford & Stein, 2007; Sadaf et al., 2023). This is consistent with our argument that assessing uncertainty under costly learning must involve nonlinear evaluations. More research is needed to explore further how individuals process information under bounded rationality and to investigate the linkages between uncertainty assessment and cognitive processes in human learning.

Conflicts of Interest

The author has no relevant financial or material conflict of interest that relates to the research described in this paper.

References

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21,* 503-546. https://doi.org/10.2307/1907921

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing.* Markham.

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal, 99,* 116-131. https://doi.org/10.2307/2234208

Billingsley, P. (1995). *Probability and Measure.* John Wiley & Sons.

Bingham, N. H. (2010). Finite Additivity versus Countable Additivity. *Electronic Journal for History of Probability and Statistics, 6,* 1-35.

Bowers, J. S., & Davis, C. J. (2012). Bayesian Just-So Stories in Psychology and Neuroscience. *Psychological Bulletin, 138,* 389-414. https://doi.org/10.1037/a0026450

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., & Montrucchio, L. (2011). Uncertainty Averse Preferences. *Journal of Economic Theory, 146,* 1275-1330. https://doi.org/10.1016/j.jet.2011.05.006

Chavas, J. P. (2015). Coase Revisited: Economic Efficiency under Externalities, Transaction Costs and Non-Convexity. *Journal of Institutional and Theoretical Economics, 171,* 709-734.

Chavas, J., & Briec, W. (2012). On Economic Efficiency under Non-Convexity. *Economic Theory, 50,* 671-701. https://doi.org/10.1007/s00199-010-0587-1

Chavas, J., & Briec, W. (2019). Efficiency under Uncertainty and Non-Convexity: Evaluating the Role of Probabilities. *Economica, 86,* 832-853. https://doi.org/10.1111/ecca.12293

Chavas, J., & Hall, T. (2017). Ricardo Revisited: Benefits from Trade and the Role of Non-Convex Technologies. *Theoretical Economics Letters, 7,* 263-293. https://doi.org/10.4236/tel.2017.72022

Chavas, J., & Pagani, E. (2020). On Nonlinear Pricing. *Theoretical Economics Letters, 10,* 1213-1226. https://doi.org/10.4236/tel.2020.106073

de Finetti, B. (1974). *Theory of Probability (Vol. 1).* Wiley.

Debreu, G. (1959). *The Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* Wiley.

DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill.

Ellsberg, D. (1961) Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics, 75,* 643-669. https://doi.org/10.2307/1884324

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind* (2nd ed.). W.W. Norton.

Gollier, C. (2004). *The Economics of Risk and Time.* The MIT Press.

Gould, F. J. (1969). Extensions of Lagrange Multipliers in Nonlinear Programming. *SIAM Journal on Applied Mathematics, 17,* 1280-1297. https://doi.org/10.1137/0117120

Halevy, Y. (2007). Ellsberg Revisited: An Experimental Study. *Econometrica, 75,* 503-536. https://doi.org/10.1111/j.1468-0262.2006.00755.x

Holmes, N. P., & Spence, C. (2005). Multisensory Integration: Space, Time and Superadditivity. *Current Biology, 15,* R762-R764. https://doi.org/10.1016/j.cub.2005.08.058

Hunt, C., Borgida, E., & Lavine, H. (2012). Social Cognition. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior* (pp. 456-462). Academic Press. https://doi.org/10.1016/B978-0-12-375000-6.00331-1

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47,* 263-291. https://doi.org/10.2307/1914185

Karney, D. F. (1981). Duality Gaps in Semi-Infinite Linear Programming—An Approximation Problem. *Mathematical Programming, 20,* 129-143. https://doi.org/10.1007/bf01589340

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A Smooth Model of Decision Making under Ambiguity. *Econometrica, 73,* 1849-1892. https://doi.org/10.1111/j.1468-0262.2005.00640.x

Knight, F. H. (1921). *Risk, Uncertainty, and Profit: Hart, Schaffner & Marx.* Houghton Mifflin.

Kolmogorov, A. (1956). *Foundations of the Theory of Probability.* Chelsea.

Lightner, J. E. (1991). A Brief Look at the History of Probability and Statistics. *The Mathematics Teacher, 84,* 623-630. https://doi.org/10.5951/mt.84.8.0623

Liu, Y. (2019). Countable Additivity, Idealization, and Conceptual Realism. *Economics and Philosophy, 36,* 127-147. https://doi.org/10.1017/s0266267118000536

Luenberger, D. (1997). *Optimization by Vector Space Methods.* John Wiley and Sons.

Luenberger, D. G. (1992a). Benefit Functions and Duality. *Journal of Mathematical Economics, 21,* 461-481. https://doi.org/10.1016/0304-4068(92)90035-6

Luenberger, D. G. (1992b). New Optimality Principles for Economic Efficiency and Equilibrium. *Journal of Optimization Theory and Applications, 75,* 221-264. https://doi.org/10.1007/bf00941466

Machina, M. J. (1982). “Expected Utility” Analysis without the Independence Axiom. *Econometrica, 50,* 277-323. https://doi.org/10.2307/1912631

Martin, K., Ryan, C. T., & Stern, M. (2016). The Slater Conundrum: Duality and Pricing in Infinite-Dimensional Optimization. *SIAM Journal on Optimization, 26,* 111-138. https://doi.org/10.1137/140974584

Munkres, J. R. (1975). *Topology: A First Course.* Prentice Hall.

Nau, R. F. (2001). De Finetti Was Right: Probability Does Not Exist. *Theory and Decision, 51,* 89-124. https://doi.org/10.1023/a:1015525808214

Quiroga, R. Q., Mukamel, R., Isham, E. A., Malach, R., & Fried, I. (2008). Human Single-Neuron Responses at the Threshold of Conscious Recognition. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 3599-3604. https://doi.org/10.1073/pnas.0707043105

Radner, R. (1968). Competitive Equilibrium under Uncertainty. *Econometrica, 36,* 31-56. https://doi.org/10.2307/1909602

Rockafellar, R. T. (1974a). *Conjugate Duality and Optimization.* SIAM. https://doi.org/10.1137/1.9781611970524

Rockafellar, R. T. (1974b). Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM Journal on Control, 12,* 268-285. https://doi.org/10.1137/0312021

Rubinov, A. M., Huang, X. X., & Yang, X. Q. (2002). The Zero Duality Gap Property and

Lower Semicontinuity of the Perturbation Function. *Mathematics of Operations Research, 27,* 775-791. https://doi.org/10.1287/moor.27.4.775.295

Sadaf, M. U. K., Sakib, N. U., Pannone, A., Ravichandran, H., & Das, S. (2023). A Bio-Inspired Visuotactile Neuron for Multisensory Integration. *Nature Communications, 14,* Article No. 5729. https://doi.org/10.1038/s41467-023-40686-z

Savage, L. J. (1954). *The Foundations of Statistics.* John Wiley & Sons.

Schmeidler, D. (1989). Subjective Probability and Expected Utility without Additivity. *Econometrica, 57,* 571-587. https://doi.org/10.2307/1911053

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics, 69,* 99-118. https://doi.org/10.2307/1884852

Slater, M. (1950). *Lagrange Multipliers Revisited.* Cowles Commission Discussion Paper No. 403.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in Multisensory Integration: Putting the Computation in Context. *NeuroReport, 18,* 787-792. https://doi.org/10.1097/wnr.0b013e3280c1e315

Starrett, D. A. (1972). Fundamental Nonconvexities in the Theory of Externalities. *Journal of Economic Theory, 4,* 180-199. https://doi.org/10.1016/0022-0531(72)90148-2

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323. https://doi.org/10.1007/bf00122574

Yaari, M. E. (1969). Some Remarks on Measures of Risk Aversion and on Their Uses. *Journal of Economic Theory, 1,* 315-329. https://doi.org/10.1016/0022-0531(69)90036-2

Appendix: Proofs

Proof of Proposition 1: The proof follows arguments presented in Luenberger (1992b), treating the objective function in (5) as a surplus measure that is first maximized and then redistributed to the n individuals. First, consider a feasible allocation $(\tilde{x}, \tilde{y}, z)$ where (5) does not hold. It follows that $W(U) > \sum_{i \in N} b_i(\tilde{x}_i, z_i, U_i)$. Let $\Delta W \equiv W(U) - \sum_{i \in N} b_i(\tilde{x}_i, z_i, U_i) > 0$ be a positive quantity of the sure reference bundle $\bar{g}$. This quantity can be redistributed to a specific individual. From Assumption As2 (non-satiation in $\bar{g}$), this would make this individual better off without making anyone else worse off, which is inconsistent with Pareto efficiency. Thus, Pareto efficiency implies that Equation (5) must hold.

Second, note that $W(U) \geq 0$ corresponds to fiscal feasibility. Assume that (7) does not hold for some feasible $(\tilde{x}, \tilde{y}, z)$. It follows that $W(u_1(\tilde{x}_1, z_1), \cdots, u_n(\tilde{x}_n, z_n)) > 0$. With $W(u_1(\tilde{x}_1, z_1), \cdots, u_n(\tilde{x}_n, z_n))$ being the positive quantity of the sure reference bundle $\bar{g}$, consider redistributing this quantity to a specific individual. From Assumption As2, this would make this individual better off without making anyone else worse off, which is inconsistent with Pareto efficiency. Thus, Pareto efficiency implies that (7) must hold.

Finally, consider an allocation $(\tilde{x}^*, \tilde{y}^*, z^*)$ which satisfies Equations (5) and (7), with $b_i(\tilde{x}_i^*, z_i^*, u_i(\tilde{x}_i^*, z_i^*)) = 0$, $\forall i \in N$. Assume that $(\tilde{x}^*, \tilde{y}^*, z^*)$ is not Pareto efficient. Then, there is a feasible allocation $(\tilde{x}^a, \tilde{y}^a, z^a)$ such that $u_i(\tilde{x}_i^a, z_i^a) \geq u_i(\tilde{x}_i^*, z_i^*)$ for all $i \in N$ and $u_j(\tilde{x}_j^a, z_j^a) > u_j(\tilde{x}_j^*, z_j^*)$ for some $j \in N$. Having $u_i(\tilde{x}_i^a, z_i^a) \geq u_i(\tilde{x}_i^*, z_i^*)$ implies that $b_i(\tilde{x}_i^a, z_i^a, u_i(\tilde{x}_i^*, z_i^*)) \geq 0$ for all $i \in N$. Under Assumptions As2 (non-satiation in $\bar{g}$) and As3 (no destitution), having $u_j(\tilde{x}_j^a, z_i^a) > u_j(\tilde{x}_j^*, z_j^*)$ implies that $b_j(\tilde{x}_j^a, z_j^a, u_j(\tilde{x}_j^*, z_j^*)) > 0$. But this would contradict Equations (5) and (7). Thus, under Assumptions As2 - As3, Equations (5) and (7) imply Pareto efficiency.

Proof of Proposition 2: The proof follows Gould (1969) and Chavas and Briec (2012). For a given $U \in \mathbb{U}$, assume that $(\tilde{x}^*, \tilde{y}^*, z^*, h^*)$ is a saddle-point of the generalized Lagrangian in (8)-(9). The second inequality in (9) implies:

$$h^*(\tilde{y}^*) - h^*\left(\sum_{i \in N} \tilde{x}_i^*\right) \leq h(\tilde{y}^*) - h\left(\sum_{i \in N} \tilde{x}_i^*\right) \; \forall h \in H. \tag{A1}$$

Assume that the inequalities $\tilde{y}^* \geq \sum_{i \in N} \tilde{x}_i^*$ do not hold. It means that there exist some $j \in M$ and $s \in S$ where $y_j^*(s) < \sum_{i \in N} x_{ij}^*(s)$. Consider $h(\tilde{y}) = 0$ for all $\tilde{y}$ and $h'(\tilde{y}) = h(\tilde{y}) + k y_j(s)$ where $k > 0$. Note that $h \in H$ and $h' \in H$. Letting $k \to \infty$, it would follow that $h'(\tilde{y}^*) - h'\left(\sum_{i \in N} \tilde{x}_i^*\right) \to -\infty$, i.e. that $h'(\tilde{y}^*) - h'\left(\sum_{i \in N} \tilde{x}_i^*\right)$ does not have a lower bound. But this contradicts (A1). Thus,

$$\tilde{y}^* \geq \sum_{i \in N} \tilde{x}_i^*. \tag{A2}$$

Noting that $h \in H$ is a non-decreasing function, Equation (A2) implies that $h^*(\tilde{y}^*) - h^*\left(\sum_{i \in N} \tilde{x}_i^*\right) \geq 0$. And choosing $h(\tilde{y}) = 0$ for all $\tilde{y}$ in Equation (A1) implies that $h^*(\tilde{y}^*) - h^*\left(\sum_{i \in N} \tilde{x}_i^*\right) \leq 0$. Combining these two inequalities yields:

$$h^*\left(\tilde{y}^*\right)-h^*\left(\sum_{i\in N}\tilde{x}_i^*\right)=0\,. \tag{A3}$$

Using (A3) and the first inequality in (9) gives:

$$\begin{aligned}\sum_{i\in N}b_i\left(\tilde{x}_i^*,z^*,U_i\right)&=\sum_{i\in N}b_i\left(\tilde{x}_i^*,z^*,U_i\right)+h^*\left(\tilde{y}^*\right)-h^*\left(\sum_{i\in N}\tilde{x}_i^*\right)\\&\geq\sum_{i\in N}b_i\left(\tilde{x}_i,z,U_i\right)+h^*\left(\tilde{y}\right)-h^*\left(\sum_{i\in N}\tilde{x}_i\right)\\&\geq\sum_{i\in N}b_i\left(\tilde{x}_i,z,U_i\right)\ \text{if}\ \tilde{y}\geq\sum_{i\in N}\tilde{x}_i\end{aligned} \tag{A4}$$

since $h^*\in H$ is a non-decreasing function. Result (A4) holds for all $\tilde{x}\in X_c\left(z\right),\tilde{y}_c\in Y_c\left(z\right)$ and z, thus proving that $\left(\tilde{x}^*,\tilde{y}^*,z^*\right)$ is a solution to problem (5). Combining this result with (A3) gives the strong duality property: $W\left(U\right)=L^*\left(U\right)$.