Random Evolving Lotteries and Intrinsic Preference for Information†

Faruk Gul

Paulo Natenzon

and

Wolfgang Pesendorfer‡

October 2016

Abstract

We introduce random evolving lotteries to study preference for non-instrumental information and history-dependent attitudes to risk-consumption. We provide representation theorems for separable and for non-separable risk-consumption preferences and analyze the trade off between smooth consumption paths and hedging path risk. We characterize information seeking and its opposite, information aversion. We show how our rich set of choice objects allows nuanced attitudes to information, including a preference for savoring the prospect of positive surprises, or a dreading news that will arrive soon.

† This research was supported by grant SES-1426252 from the National Science Foundation.

‡ Faruk Gul, Department of Economics, Princeton University, fgul@princeton.edu; Paulo Natenzon, Department of Economics, Washington University in St. Louis, pnatenzon@wustl.edu; Wolfgang Pesendorfer, Department of Economics, Princeton University; pesendor@princeton.edu.
1. Introduction

Consider a decision maker holding a risky prospect. At each moment, she identifies her current situation with a pair of lotteries, one describing her risky current consumption and the other a probability distribution over the (terminal) prize she will receive at some future date. Examples of terminal prizes are the decision maker’s retirement assets at a certain age, a future promotion, her children’s education, or her health status. At each time, the decision maker faces two distinct types of risk, one regarding her current consumption, the other regarding her current assessment of the probability of a future success. A decision maker may care not only about what prize she ultimately receives but also about what risk she “consumes” along the way. If so, the relevant outcomes are evolving lotteries; that is, functions that specify a lottery for each time period and the relevant choice objects are random evolving lotteries; that is, lotteries defined on such functions.

In this paper, we formulate such a model of risk consumption. We use it to study preference for (non-instrumental) information and the trade off between smooth consumption paths and path risk. We provide two representation theorems; in the first, utility is separable across time and the agent’s attitude to information is categorical: if a particular information is desirable at any time and after any history, then the agent always prefers to bring this information forward in time. Conversely, if the information is undesirable, then the agent always prefers to delay it. The separable model, like any standard separable consumption model, is indifferent to the trade off between smooth paths and path risk: whether a lottery between high and low consumption is drawn independently each period or once and for all at time 0 has no impact on the agent’s utility.

The separable model has four parameters; a utility index \( u^o \) that determines the decision maker’s attitude to current consumption risk, a utility index \( u^\bullet \) that determines the decision maker’s instantaneous risk attitude towards the terminal prize, a real valued function \( v \) that transforms instantaneous utilities, and hence, determines the decision maker’s attitude toward timing of resolution of uncertainty, and finally, a cumulative distribution function \( \lambda \) which aggregates the flow of transformed instantaneous utilities and hence determines time preference, in particular, the relative importance of receiving any given piece of information at time \( t \) or \( t - \varepsilon \) versus \( s \) or \( s - \varepsilon \).
Our main theorem generalizes the separable model and characterizes a class of such utility functions, which we call non-separable risk-consumption utilities (NRUs). The NRU replaces the cumulative distribution function $\lambda$ with a capacity $\eta$ and aggregates trajectories of transformed instantaneous utilities by identifying each such trajectory with its Choquet integral. Because capacities are more flexible, the NRU can accommodate a more nuanced attitude to information and allows us to analyze the trade-off between smooth paths and path risk.

One example of a nuanced attitude to information is the “Ostrich effect.”\textsuperscript{1} Karlsson, Loewenstein and Seppi (2009) provide evidence that shows investors check their stock portfolios (without transacting) more frequently after rises in the stock index than after falls. Such behavior suggests a greater desire for information after recently receiving some good news.\textsuperscript{2} NRU utility is consistent with the Ostrich effect and, in Theorem 7, we give conditions under which it occurs.

A second example of a nuanced attitude to information is savoring or dreading news that will arrive in the near future.\textsuperscript{3} In the former case, the prospect of a good draw causes the decision maker to prefer the information with some delay rather than immediately; in the latter case, the agent may be fearful of a bad draw and, as a result, prefer information immediately or not at all. NRU utility is consistent with savoring or dread and Theorem 6 provides conditions on the capacity that lead to one or the other.

The NRU’s capacity measures time non-separability. In section 4, we relate properties of the capacity to the trade-off between path smoothness and path risk. Consider two distinct random evolving lotteries, each leading to an equal chance of high or low consumption in every period. The first has two paths, one with high consumption the other with low consumption; the second has two paths that both alternate between high and low consumption. The first random evolving lottery offers smooth paths but exposes

\textsuperscript{1} The term “ostrich effect” was coined by Galai and Sade (2006) to describe investors who choose illiquid assets in an attempt to avoid information. Our use of the term follows Karlsson, Loewenstein and Seppi (2009) to describe investors who avoid information after bad news but may seek it after good news.

\textsuperscript{2} Of course, the same behavior might also be explained by an “income effect” in the decision maker’s desire for information; it may be that she becomes more information loving as her prospects improve.

\textsuperscript{3} Loewenstein (1987) introduces the concepts of savoring and dread in the context of consumption. A person may delay consuming ice cream because she wants to savor it. Kahnemann and Lovallo (2000) extend this idea to lotteries and identify savoring with a desire to delay information about the outcome of a lottery.
the agent to path risk; the second hedges path risk at the expense of path smoothness. We define what it means for an agent to have a “preference for hedging” or “smoothing” and relate this definition to features of the capacity. Specifically, we show that agents with a totally monotone capacity choose smooth paths at the expense of hedging. Our approach is related to Gilboa (1989) who was the first to use capacities to model time non-separability. Gilboa’s variation averse preferences satisfy our definition of a preference for smoothing adapted to his setting. Our definition represents a generalization of his that can be applied to a broader class of preferences. Models of habit formation (Pollak (1970)) are designed to capture related phenomena; to relate to this literature, we give conditions under which NRU utility can be interpreted as a model with a history dependent flow utility. The advantage of the NRU representation is that its parameters can be related to the agent’s preference for smoothing (or hedging) in a straightforward way.

1.1 Related Literature

Kreps and Porteus (1978) (henceforth KP) formulate the first model of preference for temporal of resolution of uncertainty. Their work offers a blue-print for a particular approach to behavioral economics, one in which novel psychological phenomena are modeled by enriching the set of choice objects, carriers of utility. The new choice objects in KP are temporal lotteries. Our choice objects, random evolving lotteries, are stochastic processes that take on values in \( IR^k \). In KP, each path is also a sequence of probability distribution but each of these distributions is over a more complicated space of probability distributions. Since the consequences over which our random evolving lotteries are defined are simpler, they are easier to relate to observables than temporal lotteries.\(^4\)

Our model and the KP model are not nested. Random evolving lotteries rule out the possibility that a decision maker may value information about what information she will have in the future even if this information has no effect on her beliefs about final outcomes at any point in time. The KP model does not. On the other hand, our axioms permit a decision maker to have a preference for resolving uncertainty in period 1 rather than in period 2 despite the fact that she does not value period-1 information about whether or not she will receive information in period 2. The KP model rules out this possibility.

\(^4\) Alternatively, they require fewer assumptions when relating to data.
To understand this comparison between the two models, consider the following concrete example: a patient undergoes genetic screening on October 1 \((t = 1)\). The results will be available on the afternoon of October 15 \((t = 3)\). The doctor explains to the patient that the test, when effective, determines whether or not a person has a particular genetic marker that renders him susceptible to a particular cancer. But, the test is only effective in patients that have a particular blood enzyme. In patients without the enzyme, the test is uninformative. The doctor assures the patient that checking for the blood enzyme is simple, painless and can be carried out either on the morning of October 8 \((t = 1)\) or on the morning of October 15, just before the test results become available. Note that the enzyme test conveys no information about the patient's health status without the results of the genetic screening; it only provides information about whether or not information will be available on the afternoon of October 15. Therefore, the decision to have the enzyme test on October 8 versus October 15 has no effect on the decision maker’s beliefs about her health status on October 8 or October 15.

In our model, the decision maker cares only about what he knows regarding his health status on each day and therefore, she is, by definition, indifferent between having the enzyme test on October 8 versus October 15. The KP model allows decision makers to prefer having the enzyme test on October 8 to having it on the 15th. Moreover, it requires that any decision maker who is indifferent between the two dates must also be indifferent between having the entire uncertainty (i.e., both the enzyme test and the genetic screening) resolve on the 8th or the 15th. Our model does not. In particular, in our model a decision maker who prefers early resolution will strictly prefer having both results on October 8 to having both results on the 15th despite being indifferent between situations that differ only in the date of the enzyme test.

The idea of using Choquet integration to express time non-separability is due to Gilboa (1989). Gilboa axiomatizes “variation averse” agents; that is, agents whose utility depends on a weighted average of the period utilities and on the variation in utility between consecutive periods.

Loewenstein (1987) introduces the terms savoring and dread to describe the anticipatory feelings regarding future consumption. Lovallo and Kahnemann (2000) interpret
anticipated feelings regarding the resolution of uncertainty as a form of consumption and extend Loewenstein’s notions to this domain. Both of these paper provide experimental evidence that relates the specifics of the anticipated consumption to the decision maker’s preference identifying conditions that lead the individual to savor or dread the future consumption.

Caplin and Leahy (2001) offer a theoretical model **anticipatory feelings**. They develop a two-period KP-style model which they call **psychological expected utility theory** (PEU). In PEU, a pair consisting of the decision maker’s consumption in period 1 and uncertain consumption in period 2 is mapped into a mental state. Caplin and Leahy relate properties of this mapping to various psychological phenomena, including dynamic uncertainty. The two-period version of our model is equivalent to the corresponding two-period KP model. Moreover, our model is stated entirely in terms of uncertain distributions over consequences without any reference to mental states. Nevertheless, our model is similar to Caplin and Leahy’s since we follow their lead in postulating that only the decision maker’s sequence of beliefs (in each period) over physical consequences is relevant for her payoffs and not the entire path describing the resolution of uncertainty.

Random evolving lotteries are similar to the choice objects studied by Ely, Frankel and Kamenica (2015). In their model, agents derive utility from **changes** in the lottery over terminal prizes. This is motivated by a setting in which agents seek surprise and suspense.

Our formal analysis is related to the literature on ambiguity, in particular, to Schmeidler’s (1989) Choquet expected utility theory. Our setting has no ambiguity but we use the Choquet integral to describe preferences that are not separable across time. Non-separable time preference models include Kreps and Porteus (1978) and Epstein and Zin (1989). Finally, our proofs use a characterization of integration with a total monotone (or dual-totally monotone) capacity similar to the one provided by Gilboa and Schmeidler (1994).
2. Model

A **probability** (on some $\Omega$) is a function $\theta : \Omega \to [0,1]$ such that $\{\omega \in \Omega \mid p(\omega) > 0\}$ is finite and $\sum \theta(\omega) = 1$. Given any probability and $A \subset \Omega$, we let $\theta A = \sum_{\omega \in A} \theta(\omega)$ and define a sum over the null set as 0. We call any probability *degenerate* if it has a single element in its support. For any real-valued function $f$ on some set $\Omega$ and probability $\theta$ on $\Omega$, we let $E_\theta[f]$ denote the expectation of $f$; that is, $E_\theta[f] = \sum_{\omega \in \Omega} f(\omega) \theta(\omega)$. If $f$ takes values in $\mathbb{R}^k$, then $E_\theta[f] = (E_\theta[f_1], \ldots, E_\theta[f_k])$. When $f$ is the identity function, we sometimes write $E_\theta[\omega]$ instead of $E_\theta[f]$.

Let $K_1 = \{1, \ldots, k_1\}$ be the set of (flow) consumption levels and let $K_2 = \{k_1 + 1, \ldots, k_1 + k_2\}$ be the set of terminal prizes. A probability on $K_1$ is a consumption _lottery_ and a probability on $K_2$ is a terminal prize lottery. Below, we refer to a terminal prize lottery simply as a _prize lottery_. Let $\Delta_i$ be the set of probabilities on $K_i$ for $i = 1, 2$ and let $\Delta = \Delta_1 \times \Delta_2$ be the set of lotteries. We write $\alpha, \beta \in \Delta$ for a generic lottery and write $\alpha^\circ$ for the consumption lottery and $\alpha^\bullet$ for the prize lottery of $\alpha$. When convenient, we identify $\Delta$ with the $k_1 + k_2 - 2$ dimensional simplex.

A **second-order lottery**, is a probability on $\Delta$. We let $M$ denote the set of all second-order lotteries and write $p, q \in M$ for its generic elements. A function from the unit interval to $\mathbb{R}^n$ is a _step-function_ if it is right-continuous, continuous at 1 and takes on finitely many distinct values. A step-function $x : [0,1] \to \Delta$ is an _evolving lottery_ if for each terminal prize $i$, $x(t)(i) = 0$ implies $x(s)(i) = 0$ for all $s > t$. We sometimes write $x_t$ rather than $x(t)$ so that we can write $x_t(i)$ instead of the more cumbersome $x(t)(i)$. Let $D$ be the set of all evolving lotteries. We endow $D$ with the topology induced by the $L^1$ metric $d(x,y) = \int_0^1 |x_t - y_t| \, dt$.

Let $\Pi$ denote the set of all probabilities on $D$. For $P \in \Pi$, define $P_t \in M$ as follows:

$$P_t(\alpha) = P\{x \mid x_t = \alpha\}$$

Hence, $P_t$ is the $t$-th coordinate distribution of $P$. For any set $A \subset D$, such that $PA > 0$, let $P_A$ be the conditional probability of $P$ given $A$, that is:

$$P_A(y) = \begin{cases} \frac{P(y)}{PA} & \text{if } y \in A \\ 0 & \text{otherwise} \end{cases}$$
We say that $P \in \bar{\Pi}$ is a random evolving lottery (REL) if it satisfies the following martingale property: for any $\alpha_1, \ldots, \alpha_n, s_1 < \ldots < s_n < t$, let $A = \{x \in D \mid x_{s_i} = \alpha_i\}$. Then, $PA > 0$ implies

$$E_{P_A}[x_t^\bullet] = \alpha_n^\bullet$$

Let $\Pi$ be the set of RELs. It follows from the martingale property (and the law of iterated expectation) that $E_P[x_t^\bullet] = E_P[x_0^\bullet]$.

For $\alpha \in \Delta$, let $x_\alpha$ denote the constant evolving lottery such that $x_t = \alpha$ for all $t \in [0, 1]$. By the martingale property, if $P(x) = 1$ for some $x$, then $x^\bullet = x_\alpha^\bullet$ for some $\alpha \in \Delta$. Let $R^\alpha$ denote the degenerate REL such that $R^\alpha(x_\alpha) = 1$ for some $\alpha$; thus, the REL $R^\alpha$ reveals no information along the way and the decision-maker consumes $\alpha$ throughout. For $p \in M$, let $R^p$ be the REL such that $R^p(x_\alpha) = p(\alpha)$. If $p$ is non-degenerate, then the REL $R^p$ reveals some information at time 0 but reveals no further information thereafter.

Our first representation of preferences over RELs yields the simplest model of risk-consumption. This model serves as our benchmark for the history-dependent model that we consider in the next section. Four parameters characterize the benchmark model; a linear utility function defined on consumption lotteries; a linear utility function defined on prize lotteries; a function that describes the decision maker’s preference for information; and a distribution function that describes the relative importance of each time interval.

Let $\succeq$ be a binary relation on $\Pi$; that is, a subset of $\Pi \times \Pi$. We say that $\succeq$ is degenerate if $R^\alpha \sim R^\beta$ whenever $\alpha^\bullet = \beta^\bullet$ or if $R^\alpha \sim R^\beta$ whenever $\alpha^\circ = \beta^\circ$. We require $\succeq$ to be a non-degenerate binary relation that satisfies the following axioms:

Axiom 1: $\succeq$ is a complete and transitive.

We let $\succ$ denote the strict part of $\succeq$; that is, $P \succ Q$ if and only if $[P \succeq Q$ and $Q \not\preceq P]$. For any $P, Q \in \Pi$ and $a \in [0, 1]$, let $aP + (1 - a)Q$ denote the usual mixture of probabilities. Clearly, with this operation $\Pi$ is a mixture space. We impose the independence axiom on this mixture space:

Axiom 2: $P \succ Q$ and $a \in (0, 1)$ implies $aP + (1 - a)R \succ aQ + (1 - a)R$. 

7
We endow \( \Pi \) with the Prohorov metric. Our next axiom is continuity:

**Axiom 3:** The sets \( \{P \in \Pi \mid P \succeq Q\} \) and \( \{P \in \Pi \mid Q \succeq P\} \) are closed for every \( Q \in \Pi \).

The restriction of \( \succeq \) to \( \{R^\alpha \in \Pi \mid \alpha \in \Delta\} \) induces a preference on \( \Delta \). The next Axiom guarantees that this induced preference satisfies independence.

**Axiom 4:** If \( R^\alpha \succeq R^\beta \) and \( a \in (0, 1) \) then \( R^{a\alpha+(1-a)\gamma} \succeq R^{a\beta+(1-a)\gamma} \).

Define the binary relation \( \succeq_0 \) on \( M \) as follows: for \( p, q \in M \), \( p \succeq_0 q \) if (and only if) \( R^p \succeq R^q \). We call \( \succeq_0 \) the induced preference over second-order lotteries. Axiom 2 implies that \( \succeq_0 \) satisfies independence. We say that \( P \) dominates \( Q \) if \( P_t \succeq_0 Q_t \) for all \( t \); \( P \) strictly dominates \( Q \) if \( P \) dominates \( Q \) and \( Q \) does not dominate \( P \). The next axiom is necessary to ensure separability across time intervals:

**Axiom 5:** \( P \) strictly dominates \( Q \) implies \( P \succ Q \).

A continuous function \( u : \Delta \to [0, 1] \) is a utility if it is onto and separable; that is, if there exists \( u^o : \Delta_1 \to [0, 1] \) and \( u^* : \Delta_2 \to [0, 1] \) such that \( u(\alpha) = u^o(\alpha^o) + u^*(\alpha^*) \) for all \( \alpha \in \Delta \). A utility is linear if \( u(a\alpha + (1-a)\beta) = au(\alpha) + (1-a)u(\beta) \). Let \( \Phi \) denote the set of all step-functions \( \phi : [0, 1] \to [0, 1] \). A function \( w : \Phi \to [0, 1] \) is a path utility if it is onto, continuous, and strictly increasing; that is \( \phi_t \geq \hat{\phi}_t \) for all \( t \) and \( \phi \neq \hat{\phi} \) implies \( w(\phi) > w(\hat{\phi}) \). Let \( \Lambda \) be the set of all continuous, strictly increasing functions from \( [0, 1] \) onto itself. A path utility \( w \) is symmetric-separable if there exists \( v, \lambda \in \Lambda \) such that \( w(\phi) = \int v(\phi_t) d\lambda \). We identify the corresponding \((v, \lambda)\) with the path utility.

A function \( V \) represents \( \succeq \) if \( P \succeq Q \) if and only if \( V(P) \succeq V(Q) \). Such a function is a separable risk-consumption utility (SRU) if there is a linear utility \( u \) and a symmetric-separable path utility \((v, \lambda)\) such that

\[
V(P) = E_P[w] = \sum_x \int v(u(x_t)) d\lambda P(x)
\]

5 More precisely, for \( A \subset D \) and \( \epsilon > 0 \), let \( A^\epsilon = \{x \in D \mid \inf_{y \in A} d(x, y) < \epsilon\} \). Then, let

\[
d_p(P, Q) = \inf\{\epsilon \geq 0 \mid PA \leq QA^\epsilon + \epsilon \text{ and } QA \leq PA^\epsilon + \epsilon \text{ for all } A \subset D\}
\]

The function \( d_p : \mathcal{D} \times \mathcal{D} \to \mathbb{R}_+ \), where \( \mathcal{D} \) is the set of all finite nonempty subsets of \( D \), is the Prohorov metric.
for all $P$. If $\succeq$ can be represented by an SRU, we call it a *separable risk-consumption preference* (SRP). If the $V$ above represents $\succeq$, we call it an SRU and identify it with both $(u, v, \lambda)$ and $\succeq$.

**Theorem 1:** A non-degenerate $\succeq$ satisfies Axioms (1)–(5) if and only if it is an SRP. Moreover, its SRU representation is unique.

An SRU is linear in three distinct ways; it is a linear function on the mixture space $\Pi$, the instantaneous utility is a linear function on the mixture space $\Delta$ and, finally, the utility of each path is a linear function of the flow of (instantaneous) utilities.

Recall that we have normalized $u$ so that it takes on the value zero at the worst lottery and the value 1 at the best lottery. Similarly, we have normalized $v$ so that $v(0) = 0$ and $v(1) = 1$. Given these normalizations, Theorem 1 shows that the parameters $(u, v, \lambda)$ are unique. The parameter $u$ describes the agent’s ranking of lotteries, whereas $v$ captures the agent’s attitude to receiving information along the path. The parameter $\lambda$ measures how much weight is given to a time interval; that is, discounting.

### 2.1 Preference for Information

In this section, we consider three notions of preference for information. We show that for SRU representations all three concepts are equivalent to the curvature condition on the function $v$. To focus on just the information revelation, we assume that consumption lottery is constant on all paths in every REL we consider in this section. Thus, we assume that every REL is an element of the set

$$\Pi_c := \{ P \in \Pi \mid P(x) \cdot P(y) > 0 \text{ implies } x_i^s = y_i^s \text{ for all } s, t \}$$

The REL $P$ resolves earlier than REL $Q$ if (i) $P, Q$ have the same constant current consumption path and (ii) for some $\varepsilon > 0$, the decision maker knows at time $t$ under $P$ exactly what she would know at time $t + \varepsilon$ under $Q$ regarding the probability of obtaining each terminal prize. More precisely, for any $x \in D$ and $\varepsilon \in [0, 1]$, define $\varepsilon(x)$ as follows

$$\varepsilon(x)(t) = \begin{cases} (x^o(t), x^s(t + \varepsilon)) & \text{for } t \in [0, 1 - \varepsilon] \\ (x^o(t), x^s(1)) & \text{otherwise} \end{cases}$$
Then, define \( \varepsilon[P] \in \Pi \) as follows:

\[
\varepsilon[P]A = P\varepsilon(A)
\]

for all \( A \subset D \), where \( \varepsilon(A) = \{ \varepsilon(x) \mid x \in A \} \).

We say that \( \succeq \) is information seeking (information averse) if \( \varepsilon[P] \succeq P (P \succeq \varepsilon[P]) \) for all \( P \). If \( \varepsilon = 1 \), then \( \varepsilon[P] \) reveals all the information of \( P \) at date 0. Thus, a weaker notion of preference for information is a preference for immediate disclosure: \( \succeq \) prefers immediate disclosure if \( 1[P] \succeq P \) for all \( P \). Finally, define \( \bar{x}(P) \) to be the expected path of the REL \( P \). That is, \( \bar{x}_t(P) = \sum_{x \in D} x_t P(x) \) for all \( t \). The REL \( R\bar{x}(P) \) reveals no information about the prize lottery and, at each time \( t \), yields the expected consumption lottery of \( P \). Thus, \( R\bar{x}(P) \) conceals all information. We say that the preference is averse to no disclosure if \( P \succeq R\bar{x}(P) \) for all \( P \). The following proposition characterizes preference for information for SRU preferences:

**Theorem 2:** Let \( \succeq = (u, v, \lambda) \) be an SRU. Then, the following four statements are equivalent: (i) \( \succeq \) prefers immediate disclosure (ii) \( \succeq \) is averse to no disclosure (iii) \( \succeq \) is information seeking (iv) \( v \) is convex.

A symmetric counterpart of Theorem 2 also holds: the counterpart can be derived from Theorem 2 by replacing prefers with averse to, averse to with prefers, information seeking with information averse and convex with concave in the above statement. As we noted above, one of the virtues of RELs is their simplicity. Each path describes how the lottery over outcomes evolves over time. Theorem 2 exploits this simplicity to show that several notions of preference for information are all equivalent and relates these notions to the curvature of \( v \). Thus, there are no SRUs that prefer information either very quickly or not at all; similarly, there is no SRU that prefers a gradual resolution of uncertainty over both no disclosure and immediate resolution. In other words, SRU preferences are not rich enough to analyze applications that go beyond a categorical preference for information. In the following section, we provide a generalization of SRU that allows more nuanced attitudes towards information.
3. History-Dependent Attitudes to Consumption

In this section, we weaken the separability requirement (Axiom 5) of the SRU model to facilitate the analysis of history dependent attitudes in risk consumption. To simplify the construction and discussion of examples, we will adopt the following conventions and notation: unless otherwise indicated, all RELs will have equiprobable paths with a constant and identical prize lottery. Thus, paths differ only with respect to the consumption lottery. Furthermore, we will assume the consumption lottery yields one of two prizes; prize 2, the better prize, and prize 1, the worse prize. Let $\Pi^*$ be the set of all RELs with these properties for some fixed prizes terminal prize lottery. We can write every REL in $\Pi^*$ as a matrix:

$$R = \begin{pmatrix}
S_1 & S_2 & \ldots & S_n \\
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{1n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{12} & \cdots & x_{mn}
\end{pmatrix}$$

Thus, the REL $R$ has $m$ paths and assigns probability $\frac{1}{m}$ to each of them. Path $i$ yields prize 2 with probability $x_{ij}$ at every $t \in S_j$.

We typically suppress the sets of states, $S_1, \ldots, S_n$ and sometimes write the REL matrices in block form:

$$R = \begin{pmatrix} X & Y \\ X^* & Z \end{pmatrix}$$

Here $X$ is a $m \times k$ matrix, $Y$ is $m \times k'$, $X^*$ is $m' \times k$ and $Z$ is $m' \times k'$. The REL $R$ has $m + m'$ paths. Each column represents the consumption lotteries of the $m + m'$—paths for a set of states in which all paths are constant. We refer to the matrices $X, X^*, Y, Z$ as fragments and let $O$ denote a fragment in which every entry is a 0.

Consider the following two RELs:

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & \frac{1}{8} \end{pmatrix}$$

$$Q = \begin{pmatrix} 1 & 0 & \frac{3}{4} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}$$

Clearly, $Q_t \succeq_0 P_t$ for all $t$ and, therefore, any SRU would assign a greater utility to $Q$ than to $P$. However, $Q$ exposes the agent to greater variability across time while $P$ exposes the agent to greater path risk; that is, the utility difference between the two paths of $P$ is
larger than the corresponding difference for \( Q \). If a subject values smoothing the flow of utility over time more than hedging path risk, then she might prefer \( P \) to \( Q \). A subject who is less concerned with smoothing utility over time than with hedging path risk would prefer \( Q \) to \( P \).

Next, consider the comparison between \( R \) and \( R' \) below:

\[
R = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & \frac{1}{8} \end{pmatrix} \quad R' = \begin{pmatrix} 0 & 0 & \frac{3}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}
\]

Again, there is a clear trade-off between hedging and smoothing. Nevertheless, our weaker alternative to Axiom 5 still implies that \( R' \) is preferred to \( R \). To see the difference between the comparison of \( P \) and \( Q \) versus \( R' \) and \( R \), note that \( R \) and \( R' \) differ only in the third column; that is, during the time period in which all paths of both RELs yield their highest utility. Thus, compared to switching from \( Q \) to \( P \), switching from \( R' \) to \( R \) offers relatively less smoothing: it increases the utility of the worse path during the best time period at the expense of a corresponding reduction during the best period of the better path. The difference between \( Q \) and \( P \) is much more significant: \( P \) offers nearly constant paths while \( Q \) offers nearly perfect hedging of path risk. To provide a formal statement that differentiates between these comparisons, we borrow a concept from ambiguity theory, comonotonicity. Note that episodes of improvements and deteriorations over past consumption coincide along every path of the two RELs \( R \) and \( R' \); that is, the two RELs are comonotone. In that case, we assume that the agent trades off utility across distinct paths linearly; that is, we assume that the agent still satisfies monotonicity and prefers \( R' \) to \( R \) if it offers a better coordinate distribution at each time \( t \). Axiom 5*, below, expresses this requirement.

Call \( \iota = (S_1, \ldots, S_n) \) an ordered partition of \([0, 1]\) if the sets \( S_i \in \iota \) are pairwise disjoint and \( \bigcup_i S_i = [0, 1] \). Given any ordered partition \( \iota = (S_1, \ldots, S_n) \), let

\[
A_\iota = \{ x \in D \mid R^{x_t} \succ R^{x_s} \text{ if and only if } t \in S_i, s \in S_j \text{ for some } i < j \}
\]

**Definition:** \( P \) rank-dominates \( Q \) if \( PA_\iota = QA_\iota = 1 \) for some \( \iota \) and \( P \) dominates \( Q \). \( P \) strictly rank-dominates \( Q \) if \( P \) rank-dominates \( Q \) but \( Q \) does not rank-dominate \( P \).

**Axiom 5\*:** \( P \) strictly rank-dominates \( Q \) implies \( P \succ Q \).
Let \( S \) be the set of subsets of the unit interval that can be expressed as the finite union intervals and let \( l \) be the Lebesgue measure on \( S \). A function \( \eta : S \to [0,1] \) is a continuous capacity if (i) \( \eta \emptyset = 0, \eta[0,1] = 1 \), (ii) \( \eta S \leq \eta T \) whenever \( S \subset T \), (iii) \( S_{n+1} \subset S_n \) for all \( n \) and \( \bigcap S_n = \emptyset \) implies \( \lim \eta S_n = 0 \) and (iv) \( S \cap T = \emptyset \) implies \( [\eta(S \cup T) - \eta S > 0 \) if and only if \( lT > 0 \).

We say that \( f : [0,1] \to \mathbb{R} \) is \( S \)-measurable if \( \{ t \mid f(t) \geq \zeta \} \in S \) for all \( \zeta \in \mathbb{R} \). For any bounded, \( S \)-measurable function \( f : [0,1] \to \mathbb{R} \), the Choquet integral of \( f \) with respect to the capacity \( \eta \) is

\[
\int f d\eta := \int \eta\{ t \mid f(t) \geq \zeta \} d\zeta
\]

Recall that \( \Phi \) is the set of all utility paths; that is, the set of all step-functions \( \phi : [0,1] \to [0,1] \). Given any \( v \in \Lambda \) and continuous capacity \( \eta \), define the path utility as follows:

\[
w(\phi) = \int v(\phi_t) d\eta
\]

We identify the function \( w \) with the corresponding \((v, \eta)\) and call it a Choquet path utility. The function \( U : \Pi \to \mathbb{R} \) is an non-separable risk-consumption utility (NRU) if there exists a utility \( u \) and a Choquet path-utility \((v, \eta)\) such that

\[
U(P) = E_P[w] = \sum_x \int v(u(x_t)) d\eta P(x)
\]

If \( \succeq \) can be represented by an NRU, we call it a non-separable risk-consumption preference (NRP) and identify it with both its representation \( U \) and the corresponding \((u, v, \eta)\). For our main theorem, we replace Axiom 5 of Theorem 1 with Axiom 5*.

**Theorem 3:** A non-degenerate \( \succeq \) satisfies Axioms (1)-(4) and (5*) if and only if it is a NRU preference. Moreover, its NRU representation is unique.

Like an SRU, a NRU is a linear function on \( \Pi \) and the instantaneous utility is a linear function on \( \Delta \) but, unlike an SRU, a NRU’s path utility is not separable across time periods. This lack of separability will enable us to model the trade-off between hedging and smoothing, savoring, dread and the ostrich effect.

13
4. Hedging versus Smoothing

In this section, we propose two criteria: the first captures the agent’s attitude to path risk or hedging; the second captures the agent’s attitude to variation in utility across time or smoothing. Our definition of a preference for hedging (and smoothing) is in terms of RELs in $\Pi^*$ with the additional property that each path yields zero or one at every $t \in [0, 1]$. Thus, each path yields one of two possible deterministic consumptions at each time. Recall that 1 denotes the more desirable consumption. Let

\[
X = \begin{pmatrix} S_1 & S_2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} S_1 & S_2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.
\]

Clearly, the fragment $Y$ provides more hedging than the fragment $X$ while $X$ is smoother than $Y$. Next, let $S_1^* = S_1 \cup S_2$ and consider the following two fragments:

\[
Y^* = \begin{pmatrix} S_1^* & S_2^* \\ Y & 1_2 \\ X & 0_2 \end{pmatrix} \quad \text{and} \quad X^* = \begin{pmatrix} S_1^* & S_2^* \\ Y & 0_2 \\ X & 1_2 \end{pmatrix}.
\]

Consider an agent who prefers hedging and, therefore, prefers the fragment $Y$ to the fragment $X$. For this agent, $X^*$ matches the worse fragment $X$ with the better fragment $1_2$ and matches the better fragment $Y$ with the worse fragment $0_2$. Since $Y^*$ does the reverse, it follows that $Y^*$ provides more hedging than $X^*$.

Conversely, consider an agent who prefers smoothing and, therefore, prefers the fragment $X$ to the fragment $Y$. For this agent, the REL $X^*$ matches the preferred fragment $X$ with the better continuation $1_2$ and matches the worse fragment $Y$ with the worse continuation $0_2$. Thus $X^*$ matches good fragments with good fragments and, therefore, is smoother than $Y^*$ which does the reverse. Notice that $X^*$ is both smoother and less risky than $Y^*$. Thus, while smoothness and risk are in conflict when we compare $X$ and $Y$, both concepts agree on the ranking of $X^*$ and $Y^*$.

Below, we generalize this idea to provide an inductive definition of “less risky” and “smoother.” We say that $Y$ (as defined above) is $1$-less risky than $X$ and write $Y \succeq_h X$. 

14
Similarly, we say that \( X \) is 1-smoother than \( Y \) and write \( X \succeq_1 Y \). For \( k > 1 \), we say \( W \) is \( k \)-less risky than \( Z \), \((W \succeq_h^k Z)\), if

\[
W = \begin{pmatrix} S'_1 & S'_2 \\ W' & 1_m \\ Z' & 0_m \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} S'_1 & S'_2' \\ W' & 0_m \\ Z' & 1_m \end{pmatrix}
\]

where and \( W', Z' \) are \( n \times m \) matrices such that \( Z' \succeq_h^{k-1} W' \). Replacing \( Z' \succeq_h^{k-1} W' \) with \( W' \succeq_s^{k-1} Z' \), we obtain an analogous definition of “\( k \)-smoother.” These two constructions facilitate the definition below. Recall that \( O \) denotes a matrix in which every entry is 0. We write \((Z, O)\) for an REL that has fragment \( Z \) defined on some \( S \in S \) and \( O \) on \([0,1]\) \( \setminus S \).

**Definition:** An NRU decision maker has a preference for hedging (smoothing) if she prefers \( P = (X', O) \) to \( Q = (Y', O) \) whenever \( X' \succeq_h^k Y' \) (\( X' \succeq_s^k Y' \)) for some \( k \).

The comparison between \( X^* \) and \( Y^* \) reveals that it is possible for one REL to be both smoother and less risky than another. In fact, a more precise statement is possible: for \( k \) odd, we have \( P \succeq_h^k Q \) if and only if \( P \succeq_s^k Q \) and for \( k \) even \( P \succeq_h^k Q \) if and only if \( Q \succeq_h^k P \).

The definitions of preference for hedging and preference for smoothing above consider only a small subset of RELs and hence are weak. Below, we provide alternative, stronger conditions. Theorem 4 establishes that for NRUs, the two characterizations are equivalent. The alternative conditions are easier to state and to apply to all RELs.

For any set \( S \in \mathcal{S} \), let

\[
A^{S\alpha} = \{ x \mid R^{x_s} \succeq R^\alpha \text{ for all } s \in S \}
\]

be the paths that are no worse than \( \alpha \) at each time \( t \in S \).

**Definition:** REL \( P \) upper-dominates REL \( Q \) if and only if \( PA^{S\alpha} \succeq QA^{S\alpha} \) for all \( S \in \mathcal{S} \) and all \( \alpha \in \Delta \).

Lower domination, defined below, is the mirror image of upper domination. Let

\[
A_{S\alpha} = \{ x \mid R^\alpha \succeq R^{x_s} \text{ for all } s \in S \}
\]

be the paths that are no better than \( \alpha \) at each time \( t \in S \).
**Definition:** The REL $P$ lower-dominates $Q$ if and only if $PA_{S\alpha} \le QA_{S\alpha}$ for all $S \in S$ and all $\alpha \in \Delta$.

To illustrate these definitions, consider the following examples:

$$P = \begin{pmatrix} 0 & 1 & 1/2 \\ 1/2 & 0 & 1 \\ 1 & 1/2 & 0 \end{pmatrix} \quad Q = \begin{pmatrix} 1 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}$$

The REL $Q$ upper dominates $P$ since for all $\alpha \in [0, 1]$ the probability of a path uniformly greater than $\alpha$ is higher under $P$ than under $Q$. For example, if $S = [0, 1]$ and $\alpha = 1/2$, we have $QA_{S\alpha} = 2/3$ and $PA_{S\alpha} = 0$. Conversely, $P$ lower dominates $Q$ because the probability of a path that uniformly below $\alpha \in [0, 1]$ is smaller under $P$ than under $Q$. For example, if $S = [0, 1]$ and $\alpha = 1/2$, we have $QA_{S\alpha} = 2/3$ while $PA_{S\alpha} = 0$.

Note that $Q$ is smoother than $P$ while $P$ has less path risk. Theorem 4, below, establishes that this relationship between preference for smoothing and upper domination is general and shows that, in an NRU, a preference for upper domination corresponds to the **total monotonicity** of the capacity while a preference for lower-domination corresponds to its **dual total monotonicity**. The capacity $\eta$ is totally monotone if for all $k \geq 1$ and all families of sets $\{S_1, \ldots, S_k\}$ such that $S_i \in S$,

$$\eta \left( \bigcup_{i=1}^{n} S_i \right) \geq \sum_{L \subseteq \{1, \ldots, k\}, L \neq \emptyset} (-1)^{|L|-1} \eta \left( \bigcap_{i \in L} S_i \right)$$

The capacity $\eta^\#$ is the dual of $\eta$ if $\eta^\# S = 1 - \eta([0, 1]\setminus S)$ for all $S$ and $\eta$ is dual-totally monotone if $\eta^\#$ is totally monotone. If $\eta$ were a probability measure, the inequalities above would be equalities. To see this, note that the inclusion-exclusion principle applied to the family $(S_1, \ldots, S_k)$ implies that the right-hand side of the above inequality is simply the probability of $S = \bigcup_{i=1}^{k} S_i$. Total monotonicity requires that the capacity of any set $S = \bigcup_{i=1}^{k} S_i$ is at least what the capacity of $S$ would be were the inclusion-exclusion principle to hold. Dempster (1967) and Nguyen (1978) analyze the properties of totally monotone capacities.

**Theorem 4:** Let $\succeq = (u, v, \eta)$ be a NRU preference. Then, the following three statements are equivalent:
(i) \( \geq \) has a preference for smoothing (hedging)

(ii) \( P \geq Q \) if \( P \) upper-dominates (lower-dominates) \( Q \)

(iii) \( \eta \) is totally monotone (dual-totally monotone).

Next, we provide some intuition regarding the equivalence of (i) and (iii) and the equivalence of (ii) and (iii) in Theorem 4. Let \( \mathcal{T} \) denote a subalgebra of \( \mathcal{S} \). A totally monotone capacity \( \eta \) can be identified with a collection of probabilities \( h_\tau : \mathcal{T} \rightarrow [0,1] \), one for each finite subalgebra \( \mathcal{T} \) of \( \mathcal{S} \), such that for all \( S \in \mathcal{T} \),

\[
\eta_S = \sum_{\mathcal{T} \subset S} h_\tau(T) \tag{1}
\]

The probabilities \( h_\tau \) yield the following characterization of an NRU path utility: let \( \mathcal{T} \) be the (set inclusion) algebra generated by the partition \( \{S_1, \ldots, S_n\} \) and let \( x = (\alpha_1, \ldots, \alpha_n) \) be a path such that \( x_t = \alpha_i \) whenever \( t \in S_i \). Hence, the corresponding utility path is \( \phi = (u(\alpha_1), \ldots, u(\alpha_n)) \). Then, the utility of this path for an agent with the totally monotone capacity \( \eta \) is

\[
w(\phi) = \sum_{S \in \mathcal{T}} h_\tau(S) \min_{\{i : S_i \subset S\}} \phi_i \tag{2}
\]

where \( h_\tau \) is the probability that satisfies (1) above. Thus, with each \( S \), the agent associates the worst utility realization in \( S \). With this formulation, it is easy to see that \( P \) will yield a greater utility than \( Q \) whenever \( P \) upper-dominates \( Q \). For a dual-totally monotone \( \eta \), the same characterization obtains except that a maximization replaces the minimization in (2). In this case, the agent associates the best utility realization in \( S \) with each \( S \subset \mathcal{T} \). Then, it follows quite easily that for NRUs that have a dual-totally monotone capacity, \( P \) lower-dominates \( Q \) implies \( P \) has a greater utility than \( Q \).

To see the relationship between total monotonicity and preferences for smoothing consider the RELs \( P = X^*O \) and \( Q = Y^*O \) with \( X^*, Y^* \) defined as above. Let \( \mathcal{T} \) be the finite subalgebra of \( \mathcal{S} \) generated by the two RELs and let \( h_\tau \) be as defined in (1). Then, using the characterization in (2) and canceling common terms, it can be shown that \( P \) has a greater utility greater than \( Q \) if and only if \( h_\tau(T) \geq 0 \) where \( T \) is the event corresponding to the fragment \( X^* \) (and \( Y^* \)). Similar constructions, together with an appeal to the implications of preference smoothing for every \( k \), establish the equivalence

17
between preference for smoothing and $h_T(S) \geq 0$ for all $S$ and all finite subalgebras $\tau$. Put differently: preference for smoothing turns out to be the translation of the characterization of total monotonicity in (1) to preferences on $\Pi^*$.

Theorem 4 shows that an NRU agent with a totally monotone capacity prefers smooth utility over time. This NRU agent shares some similarities with agents whose utility index depends on the consumption history, as in models of habit formation (Pollak (1970)). To illustrate the connection between the models, we show that Choquet path utilities with a totally monotone (or a dual totally monotone) capacity can be represented as a history dependent path utility.

Recall that $\Phi$ represents the set of utility paths. Next, we define what it means for a path utility $w : \Phi \to [0, 1]$ to have a habit representation. In a habit model, the flow utility at time $t$ depends on consumption at time $t$ and on the consumption history. The function $V : [0, 1] \times \Phi \to [0, 1]$ is a history dependent utility if $V(\cdot, \phi)$ is measurable and if, for all $t \in [0, 1]$,

$$
V(t, \phi) \geq V(t, \phi') \text{ if } \phi \geq \phi' ;
$$

$$
V(t, \phi) = V(t, \phi') \text{ if } \phi_s = \phi'_s \text{ for } s \leq t.
$$

The first part of (3) requires that dominating paths yield greater flow utilities while the second condition requires that $V(t, \cdot)$ depends only on the history prior to time $t$. We say that the path utility $w$ has a habit representation if there exists a history dependent utility $V$ and an index $\lambda$ such that

$$
w(\phi) = \int_0^1 V(t, \phi) d\lambda(t)
$$

In Appendix E, we show that all Choquet path utilities with a totally monotone or a dual totally monotone capacity have a habit representation.\textsuperscript{6} Here, we illustrate this fact with a simple example.

Consider the Choquet path utility with parameters $(v, \eta)$ such that $\eta(S) = [l(S)]^2$ and let $w$ be the corresponding path utility, that is, $w(\phi) = \int v(\phi_t) d\eta$. Note that $\eta$ is

\textsuperscript{6} More generally, any NRU utility with a capacity $\eta = a\mu^1 + (1-a)\mu^2$ where $a \in [0, 1]$, $\mu^1$ totally monotone and $\mu^2$ dual totally monotone has a habit representation.
totally monotone. Define the path \( \phi^t \) such that \( \phi^t_s = \min\{\phi_t, \phi_s\} \) and define the history dependent utility \( V^\eta \) such that

\[
V^\eta(t, \phi) = 2 \int_0^t v(\phi^t_s) ds
\]

It is straightforward to verify that \( V^\eta \) satisfies the condition above for a history dependent utility. Then, after some manipulations of the Choquet integral, we obtain that

\[
w(\phi) = \int_0^1 V^\eta(t, \phi) dt
\]

for all \( \phi \in \Phi \). The history dependent utility function \( V^\eta \) has a straightforward interpretation: the utility flow at time \( t \) is an integral of the utility of past consumptions censored from above by the utility of time \( t \)'s consumption. As is typical in habit models, consumption utility at time \( t \) and consumption utility at time \( s < t \) are complements for the function \( V^\eta_t \).

Next, consider the NRU path utility with parameters \((v, \eta^+)\) such that \( \eta^+(S) = 2l(S) - [l(S)]^2 \) and let \( w^+ \) be the corresponding path utility, that is, \( w^+(\phi) = \int v(\phi_t) d\eta^+ \). Note that \( \eta^+ \) is the dual of \( \eta \) and is, therefore, dual totally monotone. Define the path \( \bar{\phi}^t \) such that \( \bar{\phi}^t_s = \max\{\phi_t, \phi_s\} \) and define the history dependent utility function \( V^{\eta^+} \) such that

\[
V^{\eta^+}(t, \phi) = 2 \int_0^t v(\bar{\phi}^t_s) ds
\]

Again, we obtain that

\[
w^+(\phi) = \int_0^1 V^\eta^+ t(\phi) dt
\]

for all \( \phi \in \Phi \). In this case, the utility flow at time \( t \) is an integral of the utility of past consumptions censored from below by the utility at time \( t \). For the function \( V^\eta^+ t \), consumption utility at time \( t \) and consumption utility at time \( s < t \) are substitutes.

The above discussion illustrates that NRU utility can be interpreted as a special case of a habit model as long as the capacity \( \eta \) is a convex combination of a totally monotone and a dual totally monotone capacity. Theorem 4 highlights a key advantage of the NRU representation: we can interpret the capacity as a parameter that measures the agent’s
attitude to hedging path risk and to smoothing. Moreover, as we show in the next section, the index $v$ can be interpreted in terms of the agent’s attitude to information. Thus, not only are the parameters of the NRU representation uniquely identified (Theorem 2), they also measure how the agent resolves key trade offs. The general habit model (as outlined above) does not allow a similarly straightforward interpretation of its parameters.

5. Preference for Information, Savoring and Dread

In Theorem 2, above, we showed that for SRU preferences, three definitions of preference for coincide: if the agent has a convex $v$, she is information seeking, she prefers immediate disclosure and she is averse to no disclosure. As the following example illustrates, NRU allows for more nuanced attitudes to information.

Recall that $\Pi_c$ is the set of all RELs with a constant consumption lottery along every path. Let $\alpha, \beta$ be two lotteries that yield the same immediate consumption but differ in the prize lotteries and let $Q^t \in \Pi_c$ be the REL that has two equiprobable paths and reveals the uncertainty about the prize lottery ($\alpha$ or $\beta$) at time $t$:

$$Q^t = \begin{pmatrix}
[0, t] & [t, 1] \\
\gamma & \alpha \\
\gamma & \beta
\end{pmatrix}$$

Let $\preceq = (u, v, \eta)$ be an NRU such that $u(\alpha) > u(\beta)$, let $\gamma = \alpha/2 + \beta/2$, and define $r := 2v(\gamma)/(v(\alpha) + v(\beta))$.

Define the capacity $\nu$ as follows: $\nu(S) = 2l(S) - l(S)^2$, where $l$ is Lebesgue measure and let $\mu^\nu$ be its dual. A straightforward calculation shows that if $\eta = \nu$, then $U(Q^t) < U(Q^r)$ for $t \neq r$. Thus, if $\eta = \nu$, the agent’s ideal time for learning the information is $t = r$. This implies that irrespective of the specification of $v$, the agent is not information seeking since disclosure at time $t$ is worse than disclosure at time $r$ for $t < r$. However, if $v$ is convex and, therefore, $r \leq 1/2$, the agent is averse to no disclosure; that is, disclosure at any time $t$ is preferred to no disclosure at all. Conversely, if $v$ is concave and, therefore, $r \geq 1/2$, the agent is averse to immediate disclosure, that is, disclosure at any time $t$ is preferred to disclosure at time $0$.

Theorem 5, below, gives necessary and sufficient conditions for an NRU agent to prefer (or be averse to) no disclosure and sufficient conditions for a preference for immediate
disclosure. The example above shows that, unlike in the SRU case, this characterization does not extend to information seeking.

**Definition:** The capacity $\eta$ is **supermodular** if $\eta(S \cup T) + \eta(S \cap T) \geq \eta(S) + \eta(T)$; $\eta$ is **submodular** if $\eta^\#$ is supermodular.

Note that every totally monotone (dual totally monotone) capacity is supermodular (submodular). We say that REL $P$ has constant immediate consumption if $x^\circ$ is constant for all $x$ such that $P(x) > 0$.

**Theorem 5:** Let $\succeq = (u, v, \eta)$ be an NRU. Then,

(i) if $\succeq$ prefers immediate disclosure, then $v$ is convex; if $v$ is convex and $\eta$ is supermodular, then $\succeq$ prefers immediate disclosure.

(ii) if $\succeq$ is averse to no disclosure, then $v$ is convex; if $v$ is convex and $\eta$ is submodular, then $\succeq$ is averse to no disclosure.

Theorem 5 is analogous to Theorem 2, above and, like Theorem 2, has a symmetric counterpart that can be obtained from Theorem 5 by switching the places of averse to with prefers, convex with concave and submodular with supermodular throughout the statement above.

SRU agents tend to have a categorical attitude towards information; if $v$ is convex, then the SRU is information seeking, prefers immediate disclosure and is averse to no disclosure. In contrast, an NRU with a convex $v$ is averse to no disclosure if its capacity is submodular and prefers immediate disclosure if its capacity is supermodular. In either case, information seeking cannot be guaranteed. Hence, SRUs are the subset of NRUs for which there is no conflict between preference for immediate disclosure and aversion to no disclosure.

Below, we take advantage of this conflict between preference for immediate disclosure and aversion to no disclosure to develop a model of savoring and dread. Loewenstein (1987) introduced these notions to facilitate his analysis of anticipatory utility in dynamic consumption. Lovallo and Kahnemann (2000) extend these notions to the analysis of resolution of uncertainty. The latter authors examine subjects’ willingness to live with uncertainty as a function of the attractiveness of the gamble confronting them. Lovallo
and Kahnemann find that subjects are willing to delay the resolution of uncertainty for sufficiently attractive gambles and interpret this behavior as a savoring positive future outcomes. Below, we take advantage of NRUs more nuanced attitude towards information to formulate a definition of savoring distinct from aversion to information.

For $t \in (0, 1)$, let $\Pi^t_c$ be the set of all RELs with constant immediate consumption that provide significant information at time $t$ and only at time $t$. Thus, $P \in \Pi^t_c$ has paths of the form $x_\alpha tx_\beta_1, \ldots, x_\alpha tx_\beta_n$ for $t \in (0, 1)$ and $\alpha = \sum_{i=1}^n \beta_i P(\beta_i)$. Also, the information must be significant; that is, $R_{\beta_1} \succ R_{\beta_2}$ and $P(x_\alpha tx_\beta_1) \cdot P(x_\alpha tx_\beta_2) > 0$.

**Definition:** The preference $\succeq$ savors $P \in \Pi^t_c$ if $P \succ 1[P] \succeq R^x(P)$; it dreads $P \in \Pi^t_c$ if $R^x(P) \succeq 1[P] \succ P$.

Thus, a person savors information if she prefers immediate disclosure to no disclosure but enjoys some, possibly short, delay even more than immediate disclosure. Similarly, a person dreads information if she prefers no disclosure but finds certain intermediate levels of delay even more onerous than immediate disclosure.

Theorem 6 below shows that savoring results when an NRU decision maker has a sufficiently submodular capacity. To make this statement precise, we offer the following definition of “more submodular than.” The symmetric counterpart of this definition; that is, “more supermodular than,” is derived by reversing the inequality below.

**Definition:** The capacity $\eta$ is more submodular than $\eta^*$ if there is a concave function $f$ such that $\eta = f \circ \eta^*$.

Theorem 6 below shows that a convex $v$ together with a sufficiently submodular $\eta$ can make the decision maker savor any $P \in \Pi^t_c$. The requirement that $\eta$ is sufficiently submodular is needed to overcome the effect of the convexity of $v$ which tends to make immediate disclosure more attractive. Similarly, a concave $v$ and sufficiently supermodular $\eta$ yield dread.

**Theorem 6:** Assume $P \in \Pi^t_c$ and $v$ is convex. Then, there is $\eta^*$ such that $(u, v, \eta)$ savors $P$ whenever $\eta$ is more submodular than $\eta^*$. Similarly, if $v$ is concave, then there is $\eta^*$ such that $(u, v, \eta)$ dreads $P$ whenever $\eta$ is more supermodular than $\eta^*$. 

The two theorems of this section together provide a comprehensive description of NRU attitudes toward information and reveal how NRUs differ from SRUs. Theorem 5 shows that a convex $v$ and a supermodular $\eta$ imply preference for immediate disclosure while Theorem 6 establishes that a convex $v$ and a sufficiently submodular $\eta$ tend to make the decision maker savor uncertainty. Thus, replacing a supermodular $\eta$ with a sufficiently submodular $\eta$ does not affect how the decision maker ranks immediate versus no disclosure but it does effect her utility of disclosure at intermediate time periods. With a sufficiently submodular $\eta$, disclosure at time $t \in (0,1)$ will be even better for the decision maker than immediate disclosure.

6. The Ostrich Effect

In this section, we define what it means for one REL to have more “news after good news” than another REL and use this definition to relate the ostrich effect to the NRU parameters. To understand our formulation of news after good news, consider the REL $Q$ below. This REL has equiprobable paths with identical and constant immediate consumption along each path. In particular, $Q$ is an element of $\Pi_c$:

$$Q = \begin{pmatrix} .6 & .8 & .8 \\ .6 & .4 & .4 \\ .2 & .4 & .4 \\ .2 & 0 & 0 \end{pmatrix}$$

The first and second paths start with a probability of .6. At some time $t$, the agent learns whether the probability (of getting the good prize) is .4 or .8. The third and fourth paths start with a probability of .2 and the agent learns whether this probability is 0 or .4. In particular, there are two distinct paths that lead to a probability of .4: along the second path, the agent arrives at .4 following bad news at time $t$ while along the third path, the agent arrives at .4 following good news at time $t$.

Now, consider two modifications of $Q$ that yield additional information at time $\tau > t$. In the first, $Q^g$, the news arrives after history (.2,.4), thus after earlier good news. In modification $Q^b$, the additional information arrives after the history (.6,.4), thus after
earlier bad news.\textsuperscript{7} 

\[
Q^g = \begin{pmatrix}
.6 & .8 & .8 \\
.6 & .8 & .8 \\
.6 & .4 & .4 \\
.6 & .4 & .4 \\
.2 & .4 & .6 \\
.2 & .4 & .6 \\
.2 & 0 & 0 \\
.2 & 0 & 0 \\
\end{pmatrix}
\quad \quad \quad 
Q^b = \begin{pmatrix}
.6 & .8 & .8 \\
.6 & .8 & .8 \\
.6 & .4 & .6 \\
.6 & .4 & .6 \\
.2 & .4 & .2 \\
.2 & .4 & .2 \\
.2 & 0 & 0 \\
.2 & 0 & 0 \\
\end{pmatrix}
\]

Experiments by Karlsson, Loewenstein and Seppi (2009) suggest that some decision makers prefer \(Q^g\) to \(Q^b\). Our objective is to show that this ranking is compatible with NRU utility and relate it to the parameters of NRU.

Recall that for any two paths \(x, y\), we let \(x_s y\) denote the path that is equal to \(x\) prior to time \(t\) and equal to \(y\) after time \(t\). Let \(Q \in \Pi(x, y)\) if there exist \(\alpha > 0, \beta > 0\) and \(0 < t, a < 1\) such that:

(i) \(Q(x) = Q(y) > 0\);
(ii) \(x_s = y_s = a\alpha + (1 - a)\beta\) if \(s \in [t, 1]\).
(iii) \(x_s \succeq_0 \alpha \succ_0 \beta \succeq_0 y_s\) for \(s < t\)

Thus, \(Q \in \Pi(x, y)\) contains two equally likely paths\textsuperscript{8} that are constant \((\gamma = a\alpha + (1 - a)\beta)\) on the interval \([t, 1]\). Along the path \(x\), the lottery \(\gamma\) is the worst lottery, whereas along \(y\), \(\gamma\) is the best lottery. Hence, along the path \(y\), at time \(t\), the agent received good news whereas along the path \(x\) the agent received bad news.\textsuperscript{9}

Let \(Q \in \Pi(\alpha, \beta, a, t)\) and suppose the agent receives additional information at time \(\tau \in (t, 1)\) that reveals either \(\alpha\) or \(\beta\). The REL \(Q^g\) reveals this information along the path \(y\). That is,

\[
Q^g(z) = \begin{cases} 
  aQ(y) & \text{if } z = y\tau x_\alpha \\
  (1 - a)Q(y) & \text{if } z = y\tau x_\beta \\
  Q(z) & \text{if } z \neq y, y\tau x_\alpha, y\tau x_\beta
\end{cases}
\]

\textsuperscript{7} The duplicate rows in the matrix representations of \(Q^g\) and \(Q^b\) accommodate the fact that some paths of these RELs are twice as likely as others. The duplication ensures that each row is equiprobable.

\textsuperscript{8} The assumption that the two paths are equally likely is made only for simplicity. A slightly more cumbersome definition would only require that both paths have strictly positive probability.

\textsuperscript{9} We assume that \(x\) is uniformly above \(\alpha\) and \(y\) is uniformly below \(\beta\). It would be sufficient to require, for all \(s < t\), that \(x_s\) is uniformly above \(y_s\) and that neither \(x_s\) nor \(y_s\) are “between” \(\alpha\) and \(\beta\).
The REL $Q^b$ reveals the same information along the path $x$. That is,

$$Q^b(z) = \begin{cases} 
aQ(y) & \text{if } z = x\tau x_\alpha \\
(1 - a)Q(y) & \text{if } z = x\tau x_\beta \\
Q(z) & \text{if } z \neq x, x\tau x_\alpha, x\tau x_\beta \end{cases}$$

Thus, $Q^b$ reveals the information after previous bad news while $Q^g$ reveals this information after previous good news. Notice that $Q^g$ and $Q^b$ reveal the same information at time $\tau$; they differ only in the history that precedes the information revelation. We say that $\succeq$ prefers news after good news, or equivalently, displays the ostrich effect if $Q^g \succeq Q^b$ for all $Q^g, Q^b$ that fit the above description.

**Theorem 7:** An NRU $(u, v, \eta)$ displays the ostrich effect if $v$ is convex and $\eta$ is dual totally monotone or if $v$ is concave and $\eta$ is totally monotone.

Theorem 7 provides conditions under which NRU agents will exhibit the ostrich effect noted in Karlsson, Loewenstein and Seppi (2009). Theorem 4 provides a tighter control of the circumstances under which the effect is observed than Karlsson et al. In particular, $Q^g$ and $Q^b$ provide exactly the same information; they only differ in the history preceding the information. Also, the information is small relative to the good news or bad news that precedes it. This last constraint follows from the requirement that $x$ is above $\alpha$ and $y$ is below $\beta$ prior to $t$.

If $v$ is concave and $\eta$ is totally monotone, the agent prefers no disclosure (Theorem 5) and, therefore, is averse to information. In that case, additional information at time $\tau$ reduces the agent’s utility, but it does so less if it follows good news. If $v$ is convex and $\eta$ is dual totally monotone, additional information at time $s$ increases the agent’s utility and this increase is enhanced if it follows previous good news. Thus, the two cases describe polar opposite attitudes to information but both lead to a preference for news after good news. Theorem 7 also implies that the agent displays the ostrich effect when $v$ is linear and $\eta = \eta^1 + \eta^2$ for some totally monotone $\eta^1$ and dual totally monotone $\eta^2$. 

25
7. Appendix A: Proof of Theorem 1

First, we prove the only if part of the representation theorem. That is, we assume that $\succeq$ is non-degenerate and satisfies Axioms 1–5 and establish the representation.

**Lemma 1:** There are continuous, linear functions $u : \Delta \rightarrow [0, 1]$, $u^\circ : \Delta_1 \rightarrow [0, 1]$ and $u^\bullet : \Delta_2 \rightarrow [0, 1]$ such that (i) $R^\alpha \succeq R^\beta$ if and only if $u(\alpha) \geq u(\beta)$, (ii) $u(\alpha) = u^\circ (\alpha^\circ) + u^\bullet (\alpha^\bullet)$, and (iii) $u$ is onto.

**Proof:** The restriction of $\succeq$ to $\{R^\alpha \in \Pi | \alpha \in \Delta\}$ induces a complete and transitive preference $\succeq_*$ on $\Delta$. Since $d_p(R^\alpha, R^\beta) = ||\alpha - \beta|| = d_s(x_\alpha, x_\beta)$, Axiom 3 implies that $\succeq_*$ is continuous. Axiom 4 states that $\succeq_*$ satisfies independence on the mixture space $\Delta$. Hence, there exists a linear function $u$ that represents $\succeq_*$.

Since $\Delta$ is finite dimensional and $\succeq$ is not degenerate, we can assume, without loss of generality, that there is $\sigma \in \arg\max_{\Delta} u(\cdot)$ and $\varsigma \in \arg\min_{\Delta} u(\cdot)$ such that $u(\sigma) = 1$ and $u(\varsigma) = 0$. For any $\alpha, \beta \in \Delta$, the linearity of $u$ implies

$$\frac{1}{2}u(\alpha) + \frac{1}{2}u(\beta) = u\left(\frac{1}{2}\alpha^\circ + \frac{1}{2}\beta^\circ, \frac{1}{2}\alpha^\bullet + \frac{1}{2}\beta^\bullet\right) = \frac{1}{2}u(\alpha^\circ, \beta^\bullet) + \frac{1}{2}u(\beta^\circ, \alpha^\bullet).$$

Hence,

$$u(\alpha) + u(\beta) = u(\alpha^\circ, \beta^\bullet) + u(\beta^\circ, \alpha^\bullet). \quad (A1)$$

Then, let $u^\circ (\alpha^\circ) = u(\alpha^\circ, \varsigma^\bullet)$ and let $u^\bullet (\alpha^\bullet) = u(\varsigma^\circ, \alpha^\bullet)$. Equation (A1) implies $u(\alpha) = u^\circ (\alpha^\circ) + u^\bullet (\alpha^\bullet)$ as desired. □

**Lemma 2:** There is a continuous, linear and onto function $V : \Pi \rightarrow [0, 1]$ that represents $\succeq$. Moreover, if $P_t \succeq_0 Q_t$ for all $t$, then $V(P) \geq V(Q)$.

**Proof:** The set $\Pi$ is a mixture space under the usual mixture operation and, therefore, Axioms 1–3 and the mixture space theorem guarantee the existence of a linear $\hat{V}$ that represents $\succeq$. Axiom 3 also ensures that $\hat{V}$ is continuous.

To prove the second assertion of the lemma, let $P_t \succeq_0 Q_t$ for all $t$ and assume, contrary to the assertion, that $Q \succ P$. Since $V$ is non-degenerate, we have $V(R^\alpha) > V(R^\beta)$ for some $\alpha, \beta$. Then, for $b$ sufficiently small, we have $\bar{Q} := (1-b)Q + bR^\beta \succ (1-b)P + bR^\alpha := \bar{P}$. But, $\bar{P}$ strictly dominates $\bar{Q}$ and hence, by Axiom 5, $\bar{P} \succ \bar{Q}$, a contradiction.
It follows from the second assertion proven above that $R^{\alpha_1} \succeq P \succeq R^{\alpha_2}$ for all $P$ for some $\alpha_1, \alpha_2 \in \Delta$. Hence, the range of $\hat{V}$ is a compact interval. Then, a suitable affine transformation of $\hat{V}$ yields the desired $V$. \hfill \Box

For $r \in [0, 1]$, define $v(r) = V(R^\alpha)$ for $\alpha$ such that $u(\alpha) = r$. Lemmas 1 and 2 ensure that $v$ is a well-defined element of $\Lambda$.

Let the lotteries $\sigma, \varsigma$ respectively maximize and minimize $u$ and let $\beta = (\sigma^*, \varsigma^*)$. By non-degeneracy and Lemma 1, $u(\beta) > u(\varsigma) = 0$. For $t \in [0, 1)$, let $\beta t \varsigma$ denote the evolving lottery such that yields $\beta$ at all $s < t$ and $\varsigma$ otherwise. Then, let $\lambda(1) = 1$ and, for $0 \leq t < 1$, define $\lambda(t) = V(R^{\beta t \varsigma})/V(R^\beta)$. By Axiom 5, $\lambda$ is a strictly increasing function. By Lemmas 1 and 2, $\lambda(0) = 0$, $\lim_{t \to 1} \lambda(t) = 1$. The continuity of $V$ ensures that $\lambda$ is also continuous and hence $\lambda \in \Lambda$.

Let $D_0$ be the set of all step-functions from the unit interval to the unit interval. Define $f : \Pi \to D_0$ as follows

$$f(P)(t) = E_P[v(u(x_t))].$$

Let $D_* = \{f(P) | P \in \Pi\}$. Clearly, all constant functions are in $D_*$. The linearity of $V$ ensures that $az + (1-a)z' \in D_*$ whenever $z, z' \in D_*$. Lemma 2 ensures that $f(P) = f(Q)$ implies $V(P) = V(Q)$. Therefore, we interpret $V$ as a function on $D_*$; that is, let $W(z) = V(f(P))$ whenever $f(P) = z$.

We will say that $z \in D_*$ is normal if

$$W(z) = \int zd\lambda$$

Hence, the definition of $\lambda$ ensures that all constant functions and functions $z$ of the form $z(s) = v(u(\beta))$ for $s < t$ and $z(s) = 0$ for $s \geq t$ are normal.

Lemma 3: If $z = \sum_{i=1}^n a_i z_i$, for $a_i \in \mathbb{R}$ such that $\sum_i a_i = 1$ and $z_i$ is normal for all $i$, then $z$ is also normal.

Proof: Rearrange, if needed, the equation $z = \sum_{i=1}^n a_i z_i$, by moving all terms with $a_i < 0$ to the left-hand side and divide the resulting equation by the sum of the coefficients on
the left-hand side. Then, apply the linearity of $V$ and rearrange the terms again to get

$$W(z) = \sum_i a_i \int z_i(t) \, d\lambda(t) = \int \sum_i a_i z_i(t) \, d\lambda = \int z \, d\lambda.$$  

\[\square\]

**Lemma 4:** Every $z \in D_*$ is normal.

**Proof:** For any $0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = 1$, choose $\beta$ such that $u(\beta) > 0$ and set $z_i(t) = v(u(\beta))$ for $s < s_i$ and 0 otherwise. Then, there exists some $k$ and $a_i \in \mathbb{R}$ for $i = 1, \ldots, k$ such that $z = \sum_{i=1}^{k} a_i z_i$. Let $a_{k+1} = 1 - \sum_{i=1}^{k} a_i$ and $z_{k+1}(t) = 0$ for all $t$. Note that $z = \sum_{i=1}^{k+1} a_i z_i$, each $z_i$ is normal and $\sum_{i=1}^{k+1} a_i = 1$. Then, appeal to Lemma 3 to conclude that $z$ is normal.  

To complete the if part of the proof, note that Lemma 4 implies

$$V(P) = W(f(P)) = \int f(P)(t) \, d\lambda(t) = \int E_P[v(u(x_t))] \, d\lambda(t) = E_P \left[ \int v(u(x_t)) \, d\lambda(t) \right].$$

Hence, $V$ is an SRU.  

\[\square\]

The proof of the ‘if’ part of the representation theorem is straightforward. To prove the uniqueness assertion, assume that $(u, v, \lambda)$ is a representation of some non-degenerate $\succeq$. Let $(\hat{u}, \hat{v}, \hat{\lambda})$ be a second representation. Pick $\alpha, \beta$ such that $u(\alpha) = 1$ and $u(\beta) = 0$. Then, we must have $\hat{u}(\alpha) = 1$ and $\hat{u}(\beta) = 0$. Since $u, \hat{u}$ represent the same preference relation, $\succeq_*$ on $\Delta$, agree at two distinct points $\alpha, \beta$, and are both linear, we must have $u = \hat{u}$. Similarly, the utility index $v \circ u = v \circ \hat{u}$ and $\hat{v} \circ u$ represent the same linear preference over $M^*$ and agree at points $p, q$ where $p(\alpha) = 1$ and $q(\beta) = 1$. Hence, $v \circ \hat{u} = \hat{v} \circ \hat{u}$ and since $\hat{u} \in \Lambda$, we conclude $v = \hat{v}$. The same argument ensures that $V = \hat{V}$.

Choose $\alpha, \beta$ such that $\alpha^* = \beta^*$ and $u(\alpha) < u(\beta)$. Since, $\succeq$ is non-degenerate, such $\alpha, \beta$ must exist. Clearly, $R^\alpha \beta \in \Pi$ for all $t < 1$. Then, the representation yields

$$\hat{\lambda}(t) = \frac{\hat{V}(R^\alpha \beta) - \hat{v}(\hat{u}(\alpha))}{\hat{v}(\hat{u}(\beta)) - \hat{v}(\hat{u}(\alpha))} = \frac{V(R^\alpha \beta) - v(u(\alpha))}{v(u(\beta)) - v(u(\alpha))} = \lambda(t)$$

as desired.  

\[\square\]
8. Appendix B: Proof of Theorem 2

The equivalence of (i) and (iv) is immediate as is the fact that (iii) implies (ii). Suppose $v$ is convex and fix any REL $P$ and $\varepsilon > 0$. Take $0 = s_0 < s_1 < s_2 < \cdots < s_n < 1$ such that every path $x \in D$ in the support of $P$ and every path $x \in D$ in the support of $\varepsilon[P]$ is constant in each time interval $(s_{i-1}, s_i)$ for $i = 1, \ldots, n$. Letting $\lambda_i = \lambda(s_i) - \lambda(s_{i-1})$ be the weight of each time interval, we have

$$V(\varepsilon[P]) = \sum_x \int_0^1 v(u(x_t)) d\lambda(t) \varepsilon[P](x)$$

$$= \sum_x \sum_i v(u(x_{s_{i-1}})) \lambda_i \varepsilon[P](x)$$

$$= \sum_i \lambda_i \sum_x v(u^o(x_{s_{i-1}}) + u^*(x_{s_{i-1}} + \varepsilon)) P(x)$$

$$\geq \sum_i \lambda_i \sum_x v(u^o(x_{s_{i-1}}) + u^*(x_{s_{i-1}})) P(x)$$

$$= V(P)$$

with the convention that $x_t^* = x_1^*$ for every $t > 1$. The inequality above follows from the martingale property, the linearity of $u^*$ and the convexity of $v$. Hence the SRU is information-seeking and therefore (iv) implies (iii).

Conversely, suppose $v$ is not a convex function. Then,

$$v(au_1 + (1 - a)u_2) > av(u_1) + (1 - a)v(u_2)$$

for some $u_1 < u_2 \in [0, 1]$ and $a \in (0, 1)$. Without loss of generality, $u_2 - u_1 < \max\{u^*(\alpha^*) : \alpha^* \in \Delta_2\} - \min\{u^*(\alpha^*) : \alpha^* \in \Delta_2\}$. Then, we can take $\alpha^*, \beta^* \in \Delta_2$ and $\gamma^o \in \Delta_1$ such that $u^*(\alpha^*) + u^o(\gamma^o) = u_1$ and $u^*(\beta^*) + u^o(\gamma^o) = u_2$. Let $x$ be an evolving lottery such that $x^o(t) = \gamma^o$ for all $t$, $x^*(t) = a\alpha^* + (1 - a)\beta^*$ for $t < 1/2$ and $x^*(t) = \alpha^*$ for $t \geq 1/2$. Also, let $y$ be an evolving lottery with $y^o(t) = \gamma^o$ for all $t$, $y^*(t) = a\alpha^* + (1 - a)\beta^*$ for $t < 1/2$ and $y^*(t) = \beta^*$ for $t \geq 1/2$. Finally, let $P$ be the REL such that $P(x) = a = 1 - P(y)$. Hence, $P$ offers the constant consumption lottery $\gamma^o$ throughout, and the decision maker learns if she gets the prize lottery $\alpha^*$ or $\beta^*$ at time $1/2$. The display equation above yields $V(P) > V(1[P])$ and therefore (ii) implies (iv).
9. Appendix C: Proof of Theorem 3

First, we prove the only if part of the representation theorem. That is, we assume that \( \succeq \) is non-degenerate and satisfies Axioms 1–4, 5* and establish the representation.

Whenever \( \iota = (S_1, \ldots, S_k) \) is an ordered partition of \([0, 1]\), call \( P \) an \( \iota \) REL if \( PA_\iota = 1 \). Let \( \Pi_\iota \) be the set of all \( \iota \) RELs. Hence, all RELs that yield only constant evolving lotteries are \(([0,1])\) RELs. Since Lemma 1 only involves such RELs, it still holds. Hence, there exists \( \sigma \) such that \( u(\sigma) = 1 \), \( \varsigma \) such that \( u(\varsigma) = 0 \) and hence \( R^\sigma \succeq R^\alpha \succeq R^\varsigma \) for all \( \alpha \in \Delta \).

To prove a result analogous to Lemma 2, we need the following Lemma:

**Lemma 5:** For all \( r \in (0, 1) \), there is \( x^n \in A_\iota \) converging to \( x_\alpha \) such that \( u(\alpha) = r \).

**Proof:** Let \( \iota = (S_1, \ldots, S_k) \) and assume \( r \geq u^*(\sigma^*) \). Choose \( \alpha^o \in \Delta_1 \) such that \( u(o)(\alpha^o) = r - u^*(\sigma^*) \). Lemma 1 ensures that such an \( \alpha^o \) exists. Let \( \beta_i^n = ((1-n^{-i})\alpha^o+n^{-i}\varsigma^o, \sigma^o) \in \Delta \). Then, define \( x^n \) as follows: \( x^n_i = \beta_i^n \) if and only if \( t \in S_i \). Clearly, \( x^n \in A_\iota \) for all \( n \) and \( x^n \) converges to \( x_\beta \) where \( \beta = (\alpha^o, \sigma^o) \). Hence \( R^\alpha \in \Pi_\iota \) and \( R^\alpha \) converges to \( R^\beta \). Since \( u(\beta) = r \), we have completed the proof for this case. The proof for the \( r < u^*(\sigma^*) \) case is symmetric and omitted.

**Lemma 6:** There is a continuous, linear and onto function \( V : \Pi \to [0, 1] \) that represents \( \succeq \). Moreover, if \( P \) rank-dominates \( Q \), then \( V(P) \geq V(Q) \).

**Proof:** The proof of the existence of a continuous linear representation is identical to the corresponding proof in Lemma 2. Let \( \hat{V} \) be this representation. Then, suppose \( P \) rank-dominates \( Q \) but \( Q \not\succeq P \). Lemma 5 ensures the existence of \( P^n, Q^n \in \Pi_\iota \) converging respectively to \( R^{x^\alpha}, R^{x^\beta} \) for some \( \alpha, \beta \) such that \( u(\alpha) > u(\beta) \). Choose \( n \) so that \( P^n \) strictly dominates \( Q^n \). Then, continuity ensures that \( \hat{V}(aQ^n + (1-a)Q) > \hat{V}(aP^n + (1-a)P) \) for \( a \) close to zero. But, since \( P^n \) strictly dominates \( Q^n \) and \( P \) dominates \( Q \), \( aP^n + (1-a)P \) strictly rank-dominates \( aQ^n + (1-a)Q \), contradicting Axiom 5*.

Hence, \( R^\sigma \succeq P \) and by a symmetric argument \( P \succeq R^\varsigma \). It follows that the range of \( \hat{V} \) is a compact interval. Then, a suitable affine transformation of \( \hat{V} \) yields the desired \( V \).

For \( r \in [0, 1] \), define \( v(r) = V(R^\alpha) \) for \( \alpha \) such that \( u(\alpha) = r \). Lemmas 1 and 6 ensure that \( v \) is a well-defined element of \( \Lambda \). Recall that \( D_0 \) is the set of all step-functions from the unit interval to the unit interval. Define \( f : \Pi \to D_0 \) as in the proof of Theorem 1:

\[
f(P)(t) = E_P[v(u(x_t))]
\]

30
Let $D_t = \{ f(P) | P \in \Pi_t \}$. Lemma 6 ensures that $f(P) = f(Q)$ implies $V(P) = V(Q)$. The linearity of $V$ ensures that $az + (1-a)z' \in D_t$ whenever $z, z' \in D_t$.

Fix any ordered partition $\iota = (S_1, \ldots, S_k)$ and define the $k \times k$ matrix $A$ as follows: $a_{ij} = (k-i)\epsilon$. For $\epsilon > 0$ sufficiently small, it is possible to find $\alpha_{ij}$ such that $v(u(\alpha_{ij})) = a_{kj}$ and $\alpha_{ij}^* = \alpha^*$. Then, define the evolving lotteries $x^1, \ldots, x^k$ as follows: $x^i_t = \alpha_{ij}$ whenever $t \in S_j$. By invoking elementary properties of systems of linear equations, we can verify that $A$ has a non-zero determinant. Consider the following system of linear equations:

$$Ay = v \tag{A1}$$

where $v$ is a column vector such that $v_i = V(R^i)$. Let $\eta_i(S_j) = y_j$ where $y$ is the solution to the system of equations (A1).

Identify each $z \in D_t$ with the appropriate $k$-vector $(z_1, \ldots, z_k)$ and call it normal if

$$\sum_j z_j \eta_i(S_j) = V(Q) \tag{A2}$$

for $z = f(Q)$. Hence, the rows of $A$, viewed as elements of $D_t$, are normal. Then, the convex hull of these vectors is $R^k$ and therefore, contains $D_t$. Hence, for any $z \in D_t$. Then, arguing as in the proof of Lemma 3, we conclude that every $z \in D_t$ is normal.

Since every $z \in D_t$ is normal, Lemma 6 ensures that $\eta_i(S_j) \geq 0$ for all $j$. Suppose $\sum_j \eta_i(y_j) < 1$, then it is easy to find $\beta$ and construct a second order lottery $p$ such that $f(R^p) \in D_t$, $p \succ_0 \beta$ for all $t$ and $V(R^\beta) > V(R^p)$, contradicting Axiom 5*. A symmetric argument yields a contradiction if $\sum_j \eta_i(y_j) > 1$ and hence, $\sum_j \eta_i(y_j) = 1$.

For all $S \in S$, let $\eta(S) = \eta_i(S)$ for $\iota^* = (S, [0, 1]\setminus S)$ and set $\eta(\emptyset) = 0$. Next, we show that for all $\iota = (S_1, \ldots, S_n)$, $S = \bigcup_{j \leq i} S_j$ implies $\eta(S) = \sum_{j=1}^i \eta_i(S_j)$. To see this, for $\iota = (S_1, \ldots, S_n)$ and $\epsilon > 0$ choose a sequence $a^m = (a_1^m, \ldots, a_n^m)$ such that $a_j^m > a_{j+1}^m$ for all $j, m$, $\lim_{m \to \infty} a_j^m = 1$ for all $j \leq i$ and $\lim_{m \to \infty} a_j^m = 0$ for all $j > i$. Then, choose $\beta, \alpha$ such that $\alpha^* = \beta^*$ and $\beta \succ_0 \alpha$. Let $S = \bigcup_{j \leq i} S_j$ and $\iota^* = (S, [0, 1]\setminus S)$. Finally, define two RELs, $Q^m \in \Pi_t, Q \in \Pi_{t^*}$ as follows: $Q^m_t = a_j^m + (1-a_j)\alpha$ whenever $t \in S_j$; $Q_t = \beta$ if $t \in S$ and $Q_t = \alpha$ otherwise. Then, equation (A2) implies

$$\sum_{j=1}^i \eta_i(S_j)u(\beta) + (1 - \sum_{j=1}^i \eta_i(S_j))u(\alpha) = \lim V(Q^m)$$

$$\eta(S)u(\beta) + (1 - \eta(S))u(\alpha) = V(Q)$$

31
Since $V$ is continuous, we have $\lim V(Q^n) = V(Q)$ and hence

$$\eta(S) = \sum_{j \leq i} (S_i)$$

as desired.

For any $P \in \Pi_i$, $z = f(P)$, $\iota = (S_1, \ldots, S_n)$ and choose $s_i \in S_i$ for all $i$. Then, set $\eta(S_0) = 0$ and $\eta(S_{n+1}) = 1$. Then, equations (A2) and (A3) imply

$$V(P) = \sum_{i=1}^{n} z_i \left[ \eta(\bigcup_{j=1}^{i} S_j) - \eta(\bigcup_{j=1}^{i-1} S_j) \right] = \sum_{i=1}^{n} E_{P_i}[v(u(x_{s_i}))] \cdot \left( \eta(\bigcup_{j=1}^{i} S_j) - \eta(\bigcup_{j=1}^{i-1} S_j) \right)$$

$$= E_P \left[ \sum_{i=1}^{n} v(u(x_{s_i})) \left( \eta(\bigcup_{j=1}^{i} S_j) - \eta(\bigcup_{j=1}^{i-1} S_j) \right) \right] = \sum_{x} \int v(u(x))d\eta_P(x)$$

as desired. \qed

The proof of the if statement is again straightforward. Let $(u, v, \eta)$ and $(\hat{u}, \hat{v}, \hat{\eta})$ be two representations of $\succeq$. Then, arguments for showing $u = \hat{u}$, $v = \hat{v}$ and $V = \hat{V}$ are the same as the corresponding arguments in Theorem 1. To prove $\eta = \hat{\eta}$, replace $\lambda(t), \hat{\lambda}(t)$ with $\eta, \hat{\eta}$ and for all $S \in \mathcal{S}$, replace $x^t$ with $y$ such that $y_s = \beta$ for all $s \in S$ and $y_s = \alpha$ for $s \notin S$. Then, following the corresponding part of the proof of Theorem 1 yields $\eta(S) = \hat{\eta}(S)$ as desired. \qed

10. Appendix D: Proofs of Theorems 4 and 5

10.1 Matching Lemma

Let $X, Y$ be nonempty, finite and disjoint sets. A function $\rho : X \times Y \rightarrow \{0, 1\}$ is a bipartite graph and $b : X \cup Y \rightarrow [0, 1]$ is a resource constraint. We call the bipartite graph, resources constraint $(\rho, b)$, a matching problem.

For $Z \subset X$, let $Y_Z(\rho) = \{ j \in Y \mid \rho(i, j) = 1 \text{ for some } i \in Z \}$. If,

$$\bigcup_{i \in Z} b(i) \leq \sum_{j \in Y_Z(\rho)} b(j)$$
for all $Z \neq \emptyset$, we say that $(\rho, b)$ is feasible. We say that the resource constraint is tight if $\sum_{i \in X} b(i) = \sum_{j \in Y} b(i)$. A function $\chi : X \times Y \to [0, 1]$ is a solution to $(\rho, b)$ if $\sum_{j \in Y} \chi(i, j) = b(i)$ for all $j$, $\sum_{i \in X} \chi(i, j) \leq b(j)$ for all $i$ and $\chi(i, j) \leq \rho(i, j)$ for all $(i, j) \in X \times Y$. The solution $\chi$ is tight if $\sum_{j \in Y} \chi(i, j) = b(i)$ for all $i$. Note that if $\chi$ is a solution to a matching problem with a tight resource constraint, then it too must be tight.

We say that the matching problem $(\rho, b)$ is $n$-integer if $n > 0$ is an integer and $nb(i), nb(j)$ are integers for all $i \in X, j \in Y$. We say that the solution $\chi$ is $n$-integer if $n\chi(i, j)$ is an integer for all $(i, j) \in X \times Y$. The following is a restatement of Hall’s well-known solution to the marriage problem.

**Hall’s Theorem:** An $n$-integer matching problem has an $n$-integer solution if and only if it is feasible.

**Matching Lemma:** Every tight feasible matching problem has a solution.

**Proof:** Let $(\rho, b)$ be a feasible matching problem. Choose $x_*, y_* \not\in X \cup Y$ and let $X_* = X \cup \{x_*\}$, $Y_* = Y \cup \{y_*\}$, $\rho_*(i, j) = \rho(i, j)$ if $i \in X$ and $j \in Y$ and $\rho_*(i, j) = 1$ otherwise. Let

$$b_n(i) = \max\{k \mid k \text{ is an integer and } k \leq nb(i)\}/n$$

for $i \in X \cup Y$ and $b_n(x_*) = \sum_{j \in X} b(j) - \sum_{i \in X} b_n(i)$, $b_n(y_*) = \sum_{j \in Y} b(j) - \sum_{j \in Y} b_n(j)$. Clearly, $(\rho_*, b_n)$ is a feasible, $n$-integer matching problem and hence, by Hall’s Theorem, has a solution $\chi_n$. Since the sequence $\chi_n$ lies in a compact set, it must have a limit point $\hat{\chi}$. Without loss of generality, assume $\chi_n$ converges to $\hat{\chi}$. Let $\chi$ be the restriction of $\hat{\chi}$ to $X \times Y$. Since $\lim b_n(i) = b(i)$ for all $i \in X \cup Y$, we must have $\lim b_n(x_*) = \lim b(y_*) = 0$ and hence $\chi$ must be a solution to $(\rho, b)$.

**10.2 Proof of Theorem 4**

**Lemma 7:** Let $Z$ be a finite set and let $\triangleleft$ be a partial order on $Z$. Then, for any real-valued function $H$ on $Z$, there exists a unique function $h$ such that

$$H(x) = \sum_{y \triangleleft x} h(y)$$

33
Proof: Define \( h \) inductively as follows: set \( h(x) = H(x) \) for the first element \( x \) and \( h(x) = H(x) - \sum_{x \neq y \in x} h(y) \). The uniqueness of \( h \) is obvious.

When the display equation in Lemma 7 holds for all \( x \), we call \( h \) the \( \triangleleft \)-derivative of \( H \).

Suppose \( Z \) is the set of all subsets of some set and \( H \) is a capacity. Then, Dempster (1967) shows that \( H \) is totally monotone if and only if the \( \subset \)-derivative of \( H \) is nonnegative. This result extends immediately to any capacity on any finite lattice. The following extension to any \( \eta \) defined on the infinite lattice \( S \) is also immediate: let \( \theta = \{S_1, \ldots, S_n\} \) be any finite partition of \([0, 1]\) into sets in \( S \) and let \( S_\theta \) be the smallest subalgebra of \( S \) that contains \( \theta \). The capacity \( \eta \) is totally monotone if for any such \( S_\theta \), there exists a nonnegative function \( h : S_\theta \to [0, 1] \) such that

\[
\eta_S = \sum_{T \in S_\theta} h(T)
\]

for all \( S \in S_\theta \).

The characterization of total monotonicity above is related to the following characterization of the Choquet integral: let \( \eta \) be any continuous capacity and let \( H \) denote its restriction to \( S_\theta \). Let \( \eta^+ \) be the dual of \( \eta \) and let \( H^+ \) denote its restriction to \( S_\theta \). Finally, let \( h, h^+ \) be the \( \subset \)-derivatives of \( H \) and \( H^+ \) and let \( f \) be any real-valued, \( S_\theta \) measurable function on the unit interval. Then,

\[
\int fdH = \sum_{T \in S_\theta} h(T) \min_{s \in T} f(s) = \sum_{T \in S_\theta} h^+(T) \max_{s \in T} f(s)
\]

(A4)

To see why equation (A4) holds, let \( \{r_1, r_2, \ldots, r_k\} \) be the values that \( f \) takes listed in decreasing order, let \( T_i = \{t \mid f(t) = r_i\} \) and set \( T_0 = T_{n+1} = \emptyset \). Then, it is easy to verify that

\[
\int fdH = \sum_{i=1}^{n} r_i[H(\bigcup_{j=0}^{i} T_j) - H(\bigcup_{j=0}^{i-1} T_j)]
= \sum_{i=1}^{n} r_{n+1-i}[H^+(\bigcup_{j=0}^{i} T_{n+1-j}) - H^+(\bigcup_{j=0}^{i-1} T_{n+1-j})]
\]
Let \( S^i = \{ T \in S_\theta \mid T \subset \bigcup_{j=0}^i T_j \text{ and } T \not\subset \bigcup_{j=0}^{i-1} T_j \} \) and let \( S^{+i} = \{ T \in S_\theta \mid T \subset \bigcup_{j=0}^i T_{n+1-j} \text{ and } T \not\subset \bigcup_{j=0}^{i-1} T_{n+1-j} \} \)

Then, Lemma 7 and the display equation above imply

\[
\int f dH = \sum_{i=1}^n r_i \sum_{T \in S^i} h(T) = \sum_{i=1}^n r_{n+1-i} \sum_{T \in S^{+i}} h^+(T)
\]

But, since \( r_i > r_{i+1} \) for all \( i \), we conclude that \( T \in S^i \) if and only if \( \min_{s \in T} f(s) = r_i \) and \( T \in S^{+i} \) if and only if \( \max_{s \in T} f(s) = r_i \) for all \( T \in S_\theta \). Thus, we have shown that equation (A4) holds.

Recall that \( \Phi \) is the set of all step functions from \([0, 1]\) to \([0, 1]\). Let \( \Pi' \) be the set of all probabilities on \( \Phi \). For any \( x \in D \), define \( \rho^x \in \Phi \) as follows: \( [\rho^x]_t = v(u(x_t)) \). Then, each REL \( P \in \Pi \) can be mapped to a unique REL \( P' \in \Pi' \) such that \( P(x) = P'(\rho^x) \) for all \( x \in D \). Furthermore, since \( u \) and \( v \) are onto, for every \( P' \in \Pi' \), there exists a corresponding \( P \in \Pi \). Next, we restate upper and lower domination in terms “utility” RELs: for any \( c \in [0, 1] \) and \( S \in \mathcal{S} \), let \( A^{Sc} = \{ u \in \Phi, |u_t | \geq c \text{ for all } t \in S \} \). Similarly, let \( A_{S^c} = \{ u \in \Phi, |u_t | \leq c \text{ for all } t \in S \} \). For \( P, Q \in \Pi' \), we say that \( P \) u-dominates (l-dominates) \( Q \) if \( PA^S_c \geq QA^S_c \) (\( PB^S_c \leq QB^S_c \)) for all \( S \in \mathcal{S} \) and \( c \in [0, 1] \). Then, we can restate (ii) above as follows:

(iv) \( P \) u-dominates (l-dominates) \( Q \) implies \( E_P \int ud\eta \geq E_Q \int ud\eta \) for all \( P, Q \in \Pi' \).

Hence, the following lemma establishes the equivalence of (ii) and (iii) in Theorem 4:

**Lemma 8:** \( \eta \) is totally monotone (dual totally monotone) if and only if (iv) holds.

**Proof:** Suppose \( P \) u-dominates \( Q \) implies \( E_P \int ud\eta \geq E_Q \int ud\eta \) for all \( P, Q \in \Pi' \). Let \( \theta = \{ S_1, \ldots, S_n \} \) be any partition of \([0, 1]\) and let \( T \neq \emptyset \) be any element of \( S_\theta \), the smallest subalgebra of \( \mathcal{S} \) that contains \( \theta \). Hence, without loss of generality, assume \( T = \bigcup_{i=1}^{m+1} S_i \) for some \( m \) such that \( 1 \leq m < n - 1 \). Let \( N = \{1, \ldots, n\} \), \( M = \{1, \ldots, m\} \), \( \mathcal{M} = \{ L \subset M \} \), \( \mathcal{N} = \{ L \subset N \} \) and let \( S_L = \bigcup_{i \in L} S_i \). For \( L \subset M \) let \( v^L_t = 1 \) if \( t \in S_L \) or \( (t \in S_{m+1} \text{ and } m - |L| + 2 \text{ is even}); \) otherwise, \( v^L_t = 0 \). Similarly, let \( w^L_t = 1 \) if \( t \in S_L \) or \( (t \in S_{m+1} \text{ and } m - |L| + 2 \text{ is even}); \) otherwise, \( w^L_t = 0 \).
and $m - |L|$ is odd); otherwise, let $w_t^L = 0$. Let $P(u) = 2^{-m}$ if $u = v^L$ for some $L$ and $P(u) = 0$ otherwise. Similarly, let $Q(u) = 2^{-m}$ if $u = w^L$ for some $L$ and $Q(u) = 0$ otherwise.

Clearly, $PA_0^S = QA_0^S = 1$ for all $S$. For $L \in M$, $v_t^L = w_t^L$ for all $t \in S$. Also, for $S$ such that $S \not\subset S_M \cup S_{m+1}$ and $c > 0$, $PA_c^S = QA_c^S = 0$. Otherwise; that is, if $S \subset S_M \cup S_{m+1}$ and $S \cap S_{m+1} \neq \emptyset$, we have $PA_c^S = QA_c^S + 2^{-m}$ if

$$k := |\{j \mid S \cap S_j \neq \emptyset\}| = m + 1$$

and $PA_c^S = QA_c^S = 2^{-k}$ otherwise. Hence, $PA_c^S \geq QA_c^S$ for all $S$. That is, $P^i$ u-dominates $Q$. Also, by (A4),

$$E_P \left[ \int ud\eta \right] = \sum_{\emptyset \neq L \in N} h(S_L) \min\{u_t \mid t \in S_L\}$$

$$E_Q \left[ \int ud\eta \right] = \sum_{\emptyset \neq L \in N} h(S_L) \min\{u_t \mid t \in S_L\}$$

where $h$ is the derivative of $\eta$ on $S_\theta$. Straightforward but tedious calculations reveal that

$$E_P \left[ \int ud\eta \right] - E_Q \left[ \int ud\eta \right] = h(S_M \cup S_{m+1})$$

(A5)

and therefore, $E_P \int ud\eta \geq E_Q \int ud\eta$ implies $h(T) \geq 0$ for every partition $\theta$ and $T \in S_\theta$. Hence, $\eta$ is totally monotone.

For the converse, suppose $P$ u-dominates $Q$ and $\eta$ is totally monotone. For any $R \in \Pi'$, let $C(R) = \{u_t \mid t \in [0, 1] \text{ and } P(u) > 0\}$. Choose a partition $\theta = \{S_1, \ldots, S_m\}$ of $[0, 1]$ such that $P(u) + Q(u) > 0$ implies $u_t = u_s$ whenever $t, s \in S_k$ for $k = 1, \ldots, m$; that is $\theta$ renders all paths of $P, Q$ measurable. Let $h$ be the derivative of $\eta$ on $S_\theta$. Then, let $M = \{N \subset \{1, \ldots, m\} \mid N \neq \emptyset\}$ and let $S_N = \bigcup_{i \in N} S_i$. Note that for all $c$ and $S_N$ such that $h(S_N) > 0$,

$$PA_c^{S_N} = \frac{1}{h(S_N)} \sum_{u : u_t \geq c \forall t \in S_N} P(u)h(S_N)$$

Hence $P$ u-dominates $Q$ implies

$$\sum_{u : u_t \geq c \forall t \in S_N} P(u)h(S_N) \geq \sum_{u : u_t \geq c \forall t \in S_N} Q(u)h(S_N)$$

(A6)
for all \(c\).

Define the following matching problem: \(\Gamma = \{(S_N, c) \mid \emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(P)\}\), \(\Upsilon = \{(S_N, c) \mid \emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(Q)\}\), \(\rho(i, j) = 1\) if \(i = (S_N, c), j = (S_N, \hat{c})\) and \(c \geq \hat{c}\) and \(\rho(i, j) = 0\) otherwise. Finally, \(b(i) = \sum_{u:u_t=c} for all t \in S_N P(u)h(S_N)\) for all \(i = (S_N, c) \in \Gamma\) and \(b(j) = \sum_{u:u_t=c} for all t \in S_N P(u)h(S_N)\) for all \(j = (S_N, c) \in \Upsilon\). Equation (A6) ensures that this matching problem is feasible and since both \(P\) and \(Q\) are probabilities it is tight. Hence, by the matching lemma it has a solution \(\chi\).

By (A4),

\[
E_P \left[ \int u d\eta \right] - E_Q \left[ \int u d\eta \right] = \sum_u \sum_{S_N} h(S_N)P(u_t) \min_{t \in S_N} u_t - \sum_u \sum_{S_N} h(S_N)Q(u_t) \min_{t \in S_N} u_t \\
= \sum_{c \in C(P)} \sum_{S_N} \sum_{u:u_t \in S_N} P(u)h(S_N)c - \sum_{c \in C(Q)} \sum_{S_N} \sum_{u:u_t \in S_N} Q(u)h(S_N)c \\
= \sum_{c \in C(P)} \sum_{c \in C(Q)} \sum_{S_N} (c - \hat{c})\chi(S_N, c, S_N, \hat{c}) \geq 0
\]

The proof for the l-domination/dual totally monotone case is symmetric and omitted.

\[\square\]

To conclude the proof of Lemma 4, we will establish the equivalence of (i) and (iii). First, we prove (iii) implies (i). First, we show that if \(P = (X, O), Q = (Y, O)\) and \(X \succeq^k Y\), then \(P, Q\) are REL described in the first paragraph of the proof of Lemma 7 for \(m = k\) and \(n = m + n'\) where \(n'\) is the number of columns in \(O\). To see why this is so, note that if \(k = 1\), then \(X = \left(\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right)\) and \(Y = \left(\begin{array}{c} 1 \\ 0 \\ 1 \end{array}\right)\). Then, the inductive step follows from the inductive definition of \(\succeq^k\).

Let \(\theta = \{S_1, \ldots, S_{m+1}, \ldots, S_n\}\) be the partition of \([0, 1]\) corresponding to the columns of \(P\) (or equivalently, \(Q\)) and let \(h\) be the derivative of \(\eta\) on \(S_\theta\). Then, we can rewrite (A5) as follows:

\[
E_P \left[ \int u d\eta \right] - E_Q \left[ \int u d\eta \right] = h \left( \bigcup_{i=1}^{m+1} S_i \right)
\]

Since \(\eta\) is totally monotone, we conclude that \(P \succeq Q\). This argument reverses to establish that (i) implies (iii). The proof for the preference for hedging/dual total monotonicity case is symmetric and omitted. \[\square\]
10.3 Proof of Theorem 5

First, we prove part (i): assume that \( \succeq \) prefers immediate disclosure. Then, the convexity of \( v \) is an immediate consequence of the fact that the induced preference \( \succeq_0 \) on \( M \) must be risk loving.

Next, assume that \( v \) is convex and \( \eta \) is supermodular. Let \( P \in \Pi \) and choose \( 0 = t_0 < t_1, \ldots, t_{k-1} < t_k = 1 \) so that \( P(x) > 0 \) implies \( x_t = x_s \) for all \( t, s \in [t_i, t_{i+1}) \) and \( i \leq k - 1 \). Let \( \mathcal{I} = \{[0, t_1), [t_1, t_2), \ldots, [t_{k-1}, 1]\} \) be the corresponding collection of intervals. Since each \( x \) in the support of \( P \) is constant on every \( I \in \mathcal{I} \), we can identify such paths with vectors \( (x_I)_{I \in \mathcal{I}_\theta} \). Let \( S_\theta \) be smallest subalgebra of \( S \) that contains \( \theta \) and let \( H \) denote the restriction of \( \eta \) to \( S_\theta \).

Since \( \eta \) is supermodular, there exists a compact, convex set of probabilities \( L \) on the finite set \( \mathcal{I}_\theta \) such that

\[
\int_{[0,1]} v(u(x))d\eta = \int_{\mathcal{I}_\theta} v(u(x_I))dH = \min_{\ell \in L} \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I) \tag{A7}
\]

Let \( \ell_x \) be the probability (in \( L \)) that solves the maximization problem in (A7); that is,

\[
\sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell_x(I) = \min_{\ell \in L} \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I)
\]

Recall that \( P_1 \) is the marginal distribution of \( P \) at time 1. Hence, we need to show that \( V(R^{P_1}) \geq V(P) \). Choose any \( \ell \in L \). Since \( P \) is a martingale, \( P_1 \) is a mean preserving spread of \( P_I \) for all \( I \in \mathcal{I}_\theta \). Therefore, the convexity of \( v \) and the linearity of \( u \) imply

\[
U(R^{P_1}) = \sum_x v(u(x_1))P(x) = \sum_{I \in \mathcal{I}_\theta} \sum_x v(u(x_1))P(x)\ell(I)
\]

\[
\geq \sum_{I \in \mathcal{I}_\theta} \left( \sum_x v(u(x_I))P(x) \right)\ell(I)
\]

\[
= \sum_{I \in \mathcal{I}_\theta} \left( \sum_x v(u(x_I))P(x) \right)\ell(I)
\]

\[
\geq \sum_{I \in \mathcal{I}_\theta} \left( \sum_x v(u(y_I)) \right)\ell_x(I)
\]

\[
= U(P)
\]
as desired.

Next, we prove part (ii): assume that \(\succeq\) averse to no disclosure. Then, the convexity of \(v\) is an immediate consequence of the fact that the induced preference \(\succeq_0\) on \(M\) must be risk loving.

Assume that \(v\) is convex and \(\eta\) is submodular. Let \(P \in \Pi\) and define \(t_0, \ldots, t_k, \mathcal{I}, \theta, S_\theta\) and \(H\) as in the proof of part (i) above. Since current consumption is constant, the martingale property ensures that \(R^{\bar{x}(P)}\) has a single path \(\bar{x}(P) : = x_\alpha\) for some \(\alpha \in \Delta\). Since \(\eta\) is supermodular, there exists a compact, convex set of probabilities \(L\) on the finite set \(\mathcal{I}_\theta\) such that

\[
\int_{[0,1]} v(u(x))d\eta = \int_{\mathcal{I}_\theta} v(u(x_I))dH = \max_{\ell \in L} \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I) \quad (A8)
\]

Let \(\ell_x\) be the probability (in \(L\)) that solves the maximization problem in (A8); that is,

\[
\sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell_x(I) = \max_{\ell \in L} \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I)
\]

Choose any \(\ell \in L\). Since \(v\) is convex, \(u\) is linear, the martingale property of \(P\) implies

\[
V(P) = \sum_x \left( \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell_x(I) \right) P(x) \geq \sum_x \left( \sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I) \right) P(x)
\]

\[
= \sum_{I \in \mathcal{I}_\theta} \left( \sum_x v(u(x_I))P(x) \right) \ell(I)
\]

\[
\geq \sum_{I \in \mathcal{I}_\theta} v(\alpha)\ell(I)
\]

\[
= v(u(\alpha)) = V(R^{\bar{x}(P)})
\]

as desired. \(\square\)

10.4 Proof of Theorem 6:

Assume \(P \in \Pi_t\) and \(v\) is convex. Hence, \(1[P] \succeq R^{\bar{x}(P)}\) and we need only show that for some \(\eta^*, 1[P] \succ R^{\bar{x}(P)}\) whenever \(\eta\) is more submodular than \(\eta^*\). By the martingale property, we can express \(P\) as a convex combination of RELs \(P^1, \ldots, P^n\) such that \(x(P) = x(P^i) = \alpha\) for all \(i\) and either each \(P^i\) assign’s positive probability to exactly two distinct
paths or $P^1 = R^\alpha$ and every $P^i$ other than $P^1$ assign’s positive probability to exactly two distinct paths.

We will construct $\eta^*$ such that $\eta$ more submodular than $\eta^*$ implies $P^i \succ 1[P^i]$ for all $P^i \neq R^\alpha$ and appeal to the linearity of $P$. Consider the capacity $\eta_n$ below:

$$\eta_n S = (l(S))^{\frac{1}{n}}$$

for all $S \in S$ where $l$ is the Lebesgue measure. Verifying that $\eta_n$ is submodular for all $n$ is straightforward. Let $P_i(x_\alpha t^\beta) = a, P_i(x_\alpha t^\beta) = 1 - a$, $\beta^1 > \beta^2$, $a \in (0, 1)$, $\alpha = a\beta^1 + (1 - a)\beta^2$, $v_1 = v(u(\beta^1))$, $v_2 = v(u(\beta^2))$ and finally, let $v_0 = v(u(\alpha))$. Then, for $(u, v, \eta_n)$, the representation ensures that $P^i \succ 1[P^i]$ is equivalent to

$$av_1 + (1 - a)v_2 < a[v_1\eta_n[t, 1] + v_0(1 - \eta_n[t, 1])] + (1 - a)[v_0\eta_n[0, 1 - t] + v_2(1 - \eta_n[0, 1 - t])]$$

As $n$ goes to infinity, both $\eta_n[t, 1]$ and $\eta_n[0, 1 - t]$ converge to 1 and hence the right-hand side of the above equation converges to $av_1 + (1 - a)v_0$ ensuring that it holds. Choose $n_i$ such that the above inequality holds whenever $n \geq n_i$. Let $n = \max n_i$ and set $\eta^* = \eta_n$. If $\eta$ is more submodular than $\eta$, then $\eta^*$, we have a concave $f$ such that

$$\eta(S) = f(\eta^*(S)) \geq \eta^*(S)$$

for all $S \in S$. It follows that $P^i \succ 1[P^i]$ for all $P^i \neq R^\alpha$ whenever $\eta$ is more submodular than $\eta^*$. Then, the linearity of $\geq$ ensures that $P \succ 1[P]$ whenever $\eta$ is more submodular than $\eta^*$.

The proof of the dread case is symmetric and, therefore, omitted. \hfill \Box

10.5 Proof of Theorem 7:

Let $Q, Q^b, Q^g$ satisfy the properties of the theorem. Hence,

$$Q(x) = Q(y) = b > 0$$

$$x_s = y_s = a\alpha + (1 - a)\beta \text{ if } s \in [t, 1]$$

$$x_s \succeq_0 \alpha \succeq_0 \beta \succeq_0 y_s \text{ for } s < t$$
\[ Q^g(z) = \begin{cases} aQ(y) & \text{if } z = y\tau x_{\alpha} \\ (1-a)Q(y) & \text{if } z = y\tau x_{\beta} \\ Q(z) & \text{if } z \neq y, y\tau x_{\alpha}, y\tau x_{\beta} \end{cases} \]

\[ Q^b(z) = \begin{cases} aQ(y) & \text{if } z = x\tau x_{\alpha} \\ (1-a)Q(y) & \text{if } z = x\tau x_{\beta} \\ Q(z) & \text{if } z \neq x, x\tau x_{\alpha}, x\tau x_{\beta} \end{cases} \]

Then,

\[
\frac{1}{b} (U(Q^b) - U(Q)) = a \int v(u(x\tau x_{\alpha}))d\eta + (1-a) \int v(u(x\tau x_{\beta}))d\eta \\
- \int v(u(x))d\eta \\
\frac{1}{b} (U(Q^g) - U(Q)) = a \int v(u(y\tau x_{\alpha}))d\eta + (1-a) \int v(u(y\tau x_{\beta}))d\eta \\
- \int v(u(y))d\eta
\]

Choose 0 = t_0 < t_1, \ldots, t_k = t so that Q(x) > 0 implies x_{s'} = x_s whenever s', s \in [t_i, t_{i+1}) and i < k. Let \( T_1 = [t, \tau], T_2 = (\tau, 1], \theta = \{[0, t_1), \ldots, [t_{k-1}, t), T_1, T_2\} \) and let \( S_{\theta} \) be the smallest subalgebra of \( S \) that contains \( \theta \). Then, let \( S^o \subset S_{\theta} \) be the subset of \( S \) consisting of all sets that can be written as the finite union of sets of the form \([t_i, t_{i+1})\) for some \( t_i < t_j \) and let \( S^* = S \cup \{\emptyset\} \). Let \( H \) denote the restriction of \( \eta \) to \( S_{\theta} \), \( \eta^+ \) be the dual of \( \eta \) and \( H^+ \) denote its restriction to \( S_{\theta} \). Finally, let \( h, h^+ \) be the \( \subset \)-derivatives of \( H \) and \( H^+ \) respectively.

First, consider the case where \( \eta \) is totally monotone. Let \( d_1 = av(\alpha) + (1-a)v(u(\beta)) - v(u(\gamma)) \) and let \( d_2 := av(u(\gamma)) + (1-a)v(u(\beta)) - v(u(\gamma)) \). By (A4),

\[
\frac{1}{b} (U(Q^b) - U(Q)) = d_2 \sum_{S \in S^*} h(S \cup T_1 \cup T_2) \\
+ d_1 \sum_{S \in S^*} h(S \cup T_2) \\
\frac{1}{b} (U(Q^g) - U(Q)) = d_2 h(T_1 \cup T_2) + d_1 h(T_2)
\]

Since \( \eta \) is totally monotone, \( h \geq 0 \). Note further that \( d_2 < 0 \) and, since \( v \) is concave, \( d_1 \leq 0 \). It follows that \( U(Q^g) \geq U(Q^b) \) as desired.
For the dual totally monotone case, let \( d_3 := av(u(\alpha)) + (1 - a)v(u(\gamma)) - v(u(\gamma)) \). By (A4),
\[
\frac{1}{b} (U(Q^b) - U(Q)) = d_3 h^+(T_1 \cup T_2) + d_1 h^+(T_2)
\]
\[
\frac{1}{b} (U(Q^g) - U(Q)) = d_3 \sum_{S \in S^*} h^+(S \cup T_1 \cup T_2)
\]
\[
+ d_1 \sum_{S \in S^*} h^+(S \cup T_2)
\]
Since \( \eta \) is dual totally monotone, it follows that \( h^+ \geq 0 \). Note that \( d_3 > 0 \) and, since \( v \) is convex, \( d_1 \geq 0 \). Therefore, \( U(Q^g) \geq U(Q^b) \) as desired.

11. Appendix E: Relationship to Habit Models

In section 4 we assert that a Choquet path utility \((v, \eta)\) has a habit representation if \( \eta \) is totally monotone (or dual totally monotone). We establish this result here for \( \eta \) totally monotone.

Let \( \mathcal{K} \) be the collection of compact subsets of the unit interval endowed with the Hausdorff metric and let \( \mathcal{B}(\mathcal{K}) \) be the Borel sigma algebra generated by the Hausdorff metric. Let \( \Delta \) be the set of probability measures defined on \((\mathcal{K}, \mathcal{B}(\mathcal{K}))\). Let \( \mu \) be a totally monotone continuous capacity. Then, there exists a unique \( m \in \Delta \) such that \( \mu(K) = m\{K' \subset K\} \) for all \( K \in \mathcal{K} \). (Theorem 1.13, pg. 10 and Theorem 2.7 (iii), pg 27. in Molchanov (2005)). Moreover, for any \( \phi \in \Phi \)
\[
\int_{[0,1]} (v \circ \phi)d\mu = \int_{\mathcal{K}} \min\{v(\phi_s) : s \in K\} dm
\]
(Theorem 5.1 pg. 70 in Molchanov (2005)). Note that the right hand side is the Choquet integral whereas the left hand side is a (standard) integral with respect to the unique measure \( m \in \Delta \).

Define \( g : \mathcal{K} \rightarrow [0,1] \) such that \( g(K) = \max K \) and note that \( g \) is continuous. Let \( \lambda \) be the image of \( m \) under \( g \). That is \( \lambda(s) = m\{K : g(K) \leq s\} \). Since \( \mu \) is a continuous capacity it follows that \( \lambda \in \Lambda \). By a standard result on the existence of a conditional probability measure there exists a map \( M : \mathcal{K} \times [0,1] \rightarrow [0,1] \) such that \( M(\cdot, s) \in \Delta \) for all \( s \in [0,1] \), \( M(K, \cdot) \) is a measurable function, \( m(B \cap g^{-1}(S)) = \int_S M(B, s)d\lambda \) for all \( B \in \mathcal{B}(\mathcal{K}), S = [0, t] \) for some \( t \in [0,1] \).
Then, define $V : [0, 1] \times \Phi \to [0, 1]$ such that

$$V_t(\phi) = \int_{\mathcal{K}} \min \{ v(\phi_s) : s \in K \} M(dK, s)$$

and note that the support of $M(\cdot, s)$ is $\{ K \in \mathcal{K} : \max K \leq s \}$. Clearly, $V_t$ is non-decreasing in $\phi$ and, therefore, $V$ is a history dependent utility. It follows that

$$\int_{[0,1]} V_t(\phi) d\lambda = \int_{[0,1]} \int_{\mathcal{K}} \min \{ v(\phi_s) : s \in K \} M(dK, s) d\lambda$$

$$= \int_{\mathcal{K}} \min \{ v(\phi_s) : s \in K \} dm$$

$$= \int_{[0,1]} (v \circ \phi) d\mu$$

$\square$
References


