Random Evolving Lotteries and Intrinsic Preference for Information[†]

Faruk Gul

Paulo Natenzon

and

Wolfgang Pesendorfer[‡]

March 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. We characterize information seeking behavior 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 distaste for waiting for news.

[†] This research was supported by grants 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 in time, 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 for terminal prizes are the decision maker's retirement assets at a certain age, her child's professional career, or her health status. At each moment in 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 t 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 and use it to study preference for (non-instrumental) information and history-dependent attitudes to risk consumption. We provide two representation theorems, one a special case of the other and characterizations of *information seeking* (or its opposite, *information aversion*). Moreover, we show that our rich set of choice objects also allows nuanced attitudes to information, including a preference for savoring the prospect of positive surprises, or a distaste for waiting for news.

To understand our notion of *information seeking*, or equivalently, preference for early resolution of uncertainty, consider the following examples: there are only two prizes, the better prize is 2 and the inferior prize is 1 and current consumption is constant throughout. In this context, we can identify each risky prospect α with an element of the unit interval; that is, the probability of winning prize 2. A decision maker has a choice of when to receive a particular piece information; for example, suppose the initial probability of winning the better prize is .7; moreover, she can either learn whether or not her chance of winning the prize will go up to .8 or fall to .6 at time t or she can receive this same piece of information at time $t - \varepsilon$. If she prefers to receive the information earlier in all such problems, we say that the decision maker *information seeking.* If she has the opposite preference in all such choice problems, we say that she is *information averse.*

Our first theorem yields a simple theory of risk consumption with four parameters; a utility index u° that determines the decision makers attitude to current consumption risk, a utility index $u^\bullet$ that determines the decision makers instantaneous risk attitude towards the terminal prize, a real valued function v that transforms instantaneous utilities, and hence, determines the decision makers attitude toward timing of resolution of uncertainty, and finally, a cumulative distribution function λ which aggregates the trajectory of transformed instantaneous utilities and hence determines the relative importance of receiving any given piece of information at time t or $t - \varepsilon$ versus s or $s - \varepsilon$.

We call this benchmark representation *separable risk-consumption utility* (SRU). The class of preferences represented by SRUs cannot accommodate some empirically relevant and more nuanced patterns of behavior. For example, there is evidence that shows investors check their stock portfolios (without transacting) more frequently after rises in the stock index than after drops (Karlsson, Loewenstein and Seppi (2009)). Such behavior suggests a greater desire for information after recently receiving some good news.[1] More generally, decision makers may care not only about when they receive a particular piece of information but also about what they have experienced before receiving that information.

A second class of phenomena outside the scope of the separable model relate to the implications of *savoring* or *dampening* behavior on the timing of information. For example, motivated by a tendency to savor the prospect of a good draw, the decision maker may prefer to receive information with delay; alternatively, the agent may be fearful of a bad draw and, as a result, prefer information immediately or not at all.

Our main theorem characterizes a class of such utility functions, which we call *non-separable risk-consumption utilities* (NRUs), that can accommodate such behavior. The NRU replaces the cumulative distribution function λ with a capacity η and aggregates trajectories of transformed instantaneous utilities by identifying each such trajectory with its Choquet integral. The greater flexibility that a capacity affords compared to a measure enables the NRU model to accommodate behaviors such as savoring or dampening and associates these behaviors with properties of the capacity.

[1] 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.

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 $\mathbb{R}^k$. 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.[2]

Our model and the KP model are not nested. Random evolving lotteries rule out the possibility that a decision maker may have an intrinsic value for information about what information she will have in the future when 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.

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 patients health status without the results of the genetic screening; it only provides information about whether or not information

[2] Alternatively, they require fewer assumptions when relating to data.

will be available on the afternoon of October 15. Therefore, the decision to have 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 to 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 where that differ only in the date of the enzyme test.

Caplin and Leahy (2001) offer a 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). The idea to use 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. Finally, our proofs use a characterization of total monotonicity that was similarly used in Gilboa and Schmeidler (1994).

2. Model

A *probability* (on some Ω), is a function $\theta : \Omega \to [0,1]$ such that $\{\omega \in \Omega \,|\, p(\omega) > 0\}$ is finite and $\sum \theta(\omega) = 1$. Given any such 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 Ω and probability θ on Ω, we let $E_\theta[f]$ denote the expectation of f; that is, $E_\theta[f] = \sum f(\omega)\theta(\omega)$. If f is takes values in $I\!R^k$, 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 lottery over terminal prizes. Let Δ_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 α° for the consumption lottery and $\alpha^\bullet$ for the prize lottery of α. When convenient, we identify Δ with the $k_1 + k_2 - 2$ dimensional simplex.

A *second-order lottery*, is a probability on Δ. We let M denote the set of all second-order lotteries and write $p, q \in M$ for its generic elements. A function on the unit interval to $I\!R^n$ is a *step-function* if it is right-continuous, continuous at 1 and takes on finitely many distinct values. An *evolving lottery* is a step-function $x : [0,1] \to \Delta$ such that, for each terminal prize i, if $x(t)(i) = 0$ then $x(s)(i) = 0$ 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 $\bar{\Pi}$ denote the set of all probabilities on D. For $P \in \bar{\Pi}$, 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 Π 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_α 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^α denote the degenerate REL such that $R^\alpha(x_\alpha) = 1$ for some α; thus, the REL R^α reveals no information along the way and the decision-maker consumes α 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. The benchmark model is characterized by three parameters; a linear preference over gambles, a function describing the decision maker's preference for information and a distribution function that describes the relative importance of each time interval.

We consider a binary relation $\succeq$ on Π; 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 $Q \not\succeq P]$. For any $P, Q \in \Pi$ and $a \in [0,1]$, let $aP + (1-a)Q$ denote the standard mixture operation. Clearly, with this operation Π is a mixture space. The next axiom is the independence axiom on this mixture space:

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

We endow Π with the Prohorov metric.[3] Our next axiom is continuity:

Axiom 3: *The sets $\{P \in \Pi \,|\, P \succeq Q\}$ and $\{P \in \Pi \,|\, Q \succeq P\}$ are closed for every $Q \in \Pi$.*

The restriction of $\succeq$ to $\{R^\alpha \in \Pi \,|\, \alpha \in \Delta\}$ induces a preference on Δ. The next Axiom guarantees that this induced preference satisfies independence.

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

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 $u^\circ : \Delta_1 \to [0,1]$ and $u^\bullet : \Delta_2 \to [0,1]$ such that $u(\alpha) = u^\circ(\alpha^\circ) + u^\bullet(\alpha^\bullet)$ for all $\alpha \in \Delta$. A utility is linear if $u(a \cdot \alpha + (1-a) \cdot \beta) = au(\alpha) + (1-a)u(\beta)$. Let Φ 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 Λ be the set of all continuous, strictly increasing functions from $[0,1]$ onto itself. A path utility w is stationary-separable if there exists $v, \lambda \in \Lambda$ such that $w(\phi) = \int v(\phi_t)\, d\lambda$. We identify the corresponding (v, λ) with the path utility.

[3] More precisely, for $A \subset D$ and $\epsilon > 0$, let $A^\epsilon = \{x \in D \,|\, \inf_{y \in A} d(x,y) < \epsilon\}$. Then, let

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

The function $d_p : \mathcal{D} \times \mathcal{D} \to I\!R_+$, where $\mathcal{D}$ is the set of all finite nonempty subsets of D, is the *Prohorov metric.*

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 stationary-separable path utility (v, λ) such that

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

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, λ) 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 Π, the instantaneous utility is a linear function on the mixture space Δ and, finally, the utility of each path is a linear function of the flow of (instantaneous) utilities.

Recall that we have normalized u such 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 $v_t(0) = 0, v_t(1) = 1$ for all t. Given these normalizations, Theorem 1 shows that the parameters (u, v, λ) 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 λ measures how much weight is given to a time interval.

2.1 Preference for Information

In this section, we consider three notions of preference for information. We show that in the case of an SRU representation all three concepts are captured by the curvature of the function v.

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

$$\varepsilon(x)(t) = \begin{cases} (x^\circ(t), x^\bullet(t+\varepsilon)) & \text{for } t \in [0, 1-\varepsilon] \\ (x^\circ(t), x^\bullet(1)) & \text{otherwise} \end{cases}$$

Then, define $\varepsilon[P] \in \Pi$ as follows:

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

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

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 *full disclosure*: $\succeq$ prefers full 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, $\bar{x}(P) = \sum_{x \in D} xP(x)$. 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 full concealment* 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) v is convex (concave); (ii) $\succeq$ is information-seeking (information averse); (iii) $\succeq$ prefers full disclosure (is averse to full disclosure); (iv) $\succeq$ is averse to full concealment (prefers full concealment).*

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 captured by the curvature of v; an agent who prefers full disclosure is always information seeking. Thus, there are no SRU preferences that wish to receive the information either very quickly or not at all; similarly, there can be no preferences that prefer a gradual resolution of uncertainty over full disclosure. Thus, 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 to information.

3. History-Dependent Attitudes to Risk Consumption

In this section, we weaken the separability requirement of the SRU model (Axiom 5) to facilitate the analysis of history dependent attitudes in risk consumption. To motivate the analysis below, consider RELs that have three stages: let $s = 1/3$ and

$$D_3 = \{x \in D \,|\, x_t = x_0, x_{t+s} = x_s, x_{t+2s} = x_{2s} = x_1 \text{ for all } t \in [0, s)\}$$

Consider RELs with paths in D_3 that have a constant terminal prize lottery but varying intermediate consumption lottery. If $K_1 = \{1, 2\}$, then, ignoring the (constant) prize lottery, we can identify each path with a vector $(\alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ)$, where α_j° is the probability of "high" consumption 2. Let

$$P(x) = \begin{cases} 1/2 & \text{if } x^\circ = (1, 1, 3/4) \\ 1/2 & \text{if } x^\circ = (0, 0, 0) \end{cases} \quad Q(x) = \begin{cases} 1/2 & \text{if } x^\circ = (1, 0, 1) \\ 1/2 & \text{if } x^\circ = (0, 1, 0) \end{cases} \tag{$*$}$$

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, P differs from Q in that P exposes the agent to greater across-path volatility while Q exposes the agent to greater on-path volatility. If a subject is averse to variation of consumption along a path relative to the risk inherent in selecting a particular path then P is preferred to Q. A subject who is less concerned with on-path variation than with across-path variation may prefer Q over P. By contrast, when comparing R and R' below, there is no clear trade-off between across-path and on-path variation:

$$R(x) = \begin{cases} 1/2 & \text{if } x^\circ = (1/4, 1/3, 3/4) \\ 1/2 & \text{if } x^\circ = (1/5, 1/4, 1/2) \end{cases} \quad R'(x) = \begin{cases} 1/2 & \text{if } x^\circ = (1/4, 1/3, 1/2) \\ 1/2 & \text{if } x^\circ = (1/5, 1/4, 1) \end{cases} \tag{$**$}$$

Below, we generalize Axiom 5 to allow Q to be preferred to P but we continue to assume that the agent prefers R' over R. Note that all paths of R and R' are co-monotone. Whenever one path of R yields a better consumption lottery at t than at t', so does every path of both R and R'. When all paths of two RELs are co-monotone, episodes of improvements and deteriorations over past consumption coincide along every path of the two RELs. In that case, we assume that agent satisfies monotonicity, that is, R' is preferred 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

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

Definition: *P rank-dominates Q if $PA_\iota = QA_\iota = 1$ for some ι 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 $\mathcal{S}$ be the set of subset sets of the unit interval that can be expressed as the finite union intervals and let l be the Lebesgue measure on $\mathcal{S}$. A function $\eta : \mathcal{S} \to [0, 1]$ 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 $l(T) > 0]$.

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

$$\int f d\eta := \int \eta\{t \,|\, f(t) \geq \zeta\}\, d\zeta$$

Recall that Φ 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 η, define the path utility as follows:

$$w(\phi) = \int v(\phi_t)\, d\eta$$

We identify the function w with the corresponding (v, η) and call it a Choquet path utility. The function $U : \Pi \to I\!R$ is an *non-separable risk-consumption utility* (NRU) if there exists a utility u and a Choquet path-utility (v, η) 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, η). Like an SRU, a NRU is a linear function on Π and the instantaneous utility is a linear function on Δ 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 above described phenomena.

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.*

In the following, we relate the parameters of NRU preferences to the Ostrich effect, to the trade-off between path risk and path volatility and, more broadly, to the agent's preference for information.

4. On-Path versus Across-Path Volatility

RELs may differ in across-path and on-path variability of consumption while having identical marginal distributions at each time. Next, we propose two criteria that capture agents who prefer one type of volatility over the other. To motivate our definitions, we introduce several examples of RELs. Specifically, we illustrate our definitions with RELs that have a constant terminal prize lottery and differ only in their immediate consumption lottery. Moreover, in each example the REL has equiprobable paths in D_3. A $n \times 3$ matrix is a convenient way to describe these RELs: each row $(\alpha_1, \alpha_2, \alpha_3)$ of the matrix describes a path x where the entries of the row correspond to the three distinct consumption lotteries of path x. To further simplify the exposition, we consider only lotteries that yield one of two prizes and describe the lottery by the probability of the better prize. Thus,

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

describes a REL with three equally likely paths; each path has a constant terminal-prize lottery (that is suppressed). The first row describes a path that yields "low" immediate consumption in the time interval $[0, 1/3)$, an equal chance of high or low consumption in the time interval $[1/3, 2/3)$, and high immediate consumption in the time interval $[2/3, 1]$.

Consider a comparison between P, above, and Q:

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

Both RELs imply the same coordinate distribution at each time. Clearly, Q has less on-path variability and greater across-path variability than P. Thus, an agent who is more averse to on-path variability than across path variability will prefer Q to P and, conversely, an agent who is more averse to across-path variability than on-path variability will exhibit the reverse preference.

Next, consider the following two RELs with four equally likely paths:

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

Again, the RELs P' and Q' imply the same marginal distribution over consumption lotteries at each time t. Note that Q' can be obtained from P' by making two changes: (i) the consumption lotteries for times $[2/3, 1]$ in the first and second rows switch places, and (ii) the consumption lotteries for times $[2/3, 1]$ in the third and fourth row switch places. Consider an agent who is more averse to on-path variation than across-path variation. For this agent switch (i) is an unambiguous improvement; it creates a path with consumption 1 at all times at the expense of reducing the probability of high consumption along a path where consumption varies. However, switch (ii) reduces her utility because the high consumption in the final period is more valuable in the third than in the fourth path. *Upper domination*, defined below, says that the benefit of switch (i) outweighs the cost of switch (ii). The reason is that switch (i) increases the probability of high consumption following a long history of high consumption while switch (ii) reduces the probability of high consumption (by the same amount) following a short history of high consumption. Thus, agents who prefer upper dominating RELs exhibit a greater desire to maintain a high consumption after a long history of high consumption than after a short history of high consumption.

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 path that are no worse than α at each time $t \in S$. In the example above, setting $S = [0, 1], 0 < \alpha \leq 1$, we have $Q'A^{S\alpha} = 1/4$ while $P'A^{S\alpha} = 0$.

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

It is easy to verify that in the examples above, REL Q upper dominates REL P and Q' upper dominates P'. Moreover, the REL P as defined in (*) upper dominates the REL Q defined in the same section.

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

$$A_{S\alpha} = \{x \,|\, R^\alpha \succeq R^{x_s} \text{ for all } s \in S\}$$

be the paths that are no better than α at each time $t \in S$.

Definition: *The REL P lower-dominates Q if and only if $PA_{S\alpha} \leq QA_{S\alpha}$ for all $S \in \mathcal{S}$ and all $\alpha \in \Delta$.*

It is straightforward to verify that P lower dominates Q and that Q' lower dominates P. For example, if $S = [0,1]$ and $\alpha = 1/2$, then $PA_{S\alpha} = 0$ and $QA_{S\alpha} = 2/3$, $P'A_{S\alpha} = 1/4$ and $Q'A_{S\alpha} = 0$. Clearly, any agent who is more averse to across path variation than on-path variation prefers P to Q in the example above. When this agent compares P' to Q' there are again two effects. Switch (ii) is an improvement since it reduces across-path volatility at the expense of on-path volatility. By contrast, switch (i) reduces the agent's utility since it increases across-path volatility. Agents who prefer lower-dominating RELs put a greater weight on the gain implied by switch (ii) than on the loss implied by switch (i). To motivate this trade-off, note that switch (ii) increases consumption along a path with a long history of low consumption and increasing consumption after a long history of low consumption is more valuable than the corresponding loss implied by switch (i).

Theorem 3, below, shows that a preference for upper-domination corresponds to *total monotonicity* of the capacity η while a preference for lower-domination corresponds to *dual total monotonicity* of the capacity. The capacity η is totally monotone if for all $k \geq 1$ and all families of sets $\{S_1, \ldots, S_k\}$ such that $S_i \in \mathcal{S}$,

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

The capacity $\eta^{\#}$ is the dual of η if $\eta^{\#}(S) = 1 - \eta([0,1]\backslash S)$ and η is dual-totally monotone if $\eta^{\#}$ is totally monotone. If η were a probability measure then the inequalities above would be equalities. In that case, the right hand side of the above inequality is simply the probability of $S = \cup_{i=1}^{k} S_i$ implied by the inclusion-exclusion principle applied to the family $(S_1, \dots, S_k)$. Total monotonicity requires that the capacity of any set $S = \cup_{i=1}^{k} S_i$ is at least what the capacity of S would be were the exclusion-inclusion 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 two statements are equivalent:*

(i) η is totally monotone (dual-totally monotone)

(ii) $P \succeq Q$ if P upper-dominates (lower-dominates) Q.

Theorem 4 shows that if an NRU preference has a totally monotone capacity then the agent seeks to avoid on-path variation in favor of across-path variation. This NRU agent shares some similarities with agents whose utility index directly depends on the consumption history, as in models of habit formation (Pollak (1970)). Unlike models of habit formation, NRU preferences retain a certain amount of separability as expressed in Axiom 5*. Furthermore, NRU preferences satisfy the following deterministic dominance condition: if P and Q are deterministic RELs with paths x and y respectively, and, if x_t is preferred to y_t at each t, that is, $R^{x_t} \succ R^{y_t}$ for all t, then $P \succ Q$.

Next, we provide some intuition about the relationship between total monotonicity and upper dominance (or dual total monotonicity and lower dominance). A totally monotone capacity η can be interpreted as a probability measure defined on the subsets of $[0,1]$ (more precisely, on $\mathcal{S}$). For concreteness, assume that the REL is in D_3 as described above. In that case, there are three relevant time intervals: $S_1 = [0, 1/3), S_2 = [1/3, 2/3)$ and $S_3 = [2/3, 1]$. A totally monotone capacity on the three element set $\mathcal{T} = \{S_1, S_2, S_3\}$ can be described by a probability h on the power set of T, so that for all $S \subset T$

$$\eta(S) = \sum_{S' \subset T \cap S} h(S')$$

Each path $(\alpha_1, \alpha_2, \alpha_3) \in D_3$ yields a utility path $\phi = (u(\alpha_1), u(\alpha_2), u(\alpha_3))$ and the NRU agent's path utility (with the totally monotone capacity η) is

$$w(\phi) = \sum_{S \subset T} h(S) \min_{\{i: S_i \in S\}} \phi_i \tag{\dagger}$$

Thus, with each S the agent associates the worst utility realization on S. In this formulation, it is easy to see that if P upper-dominates Q then P will yield a greater utility than Q. If η is dual-totally monotone then a similar characterization obtains; again there is a probability h on the power set of T but the minimization in (†) is replaced by a maximization. Thus, with each $S \subset T$ the agent associates the best utility realization on S. Again, it follows quite easily that if P lower-dominates Q then P will yield a greater utility than Q.

5. Information Seeking and Savoring

In Theorem 2, above, we show that for SRU preference, three definitions of "information seeking" coincide: if the agent has a convex v, then she prefers early resolution of uncertainty, she prefers full disclosure of information and she is avers to full concealment of information. As the following example illustrates, NRU agents allow for more nuanced attitudes to information.

Let α, β be two lotteries that yield the same immediate consumption but differ in the prize lotteries and let Q^t be the REL that reveals the uncertainty about the prize lottery (α or β) at time t; that is, Q^t is defined as

$$Q^t(z) = \begin{cases} 1/2 & \text{if } z = x^\gamma t x^\alpha \text{ or } z = x^\gamma t x^\beta \\ 0 & \text{otherwise} \end{cases}$$

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

Let μ be the capacity such that $\mu(S) = 2l(S) - l(S)^2$ (where l is Lebesgue measure) and let $\mu^\sharp$ be its dual. A straightforward calculation shows that if $\eta = \mu$ then $U(Q^r) > U(Q^t)$ for $t \neq r$. Thus, if $\eta = \mu$ 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, $r \leq 1/2$, then the agent is *averse to full concealment*, that is, disclosure at any time t is preferred to no disclosure at all. Conversely, if v is concave and, therefore, $r \geq 1/2$ then the agent is *averse to full disclosure*, that is, disclosure at any time t is preferred to disclosure at time 0.

If $\eta = \mu^{\sharp}$, then $U(Q^0) > U(Q^t)$ for all $t \in (0, 1)$ if $r \leq 1/2$ and $U(Q^1) > U(Q^t)$ for all $t \in (0, 1)$ if $r \geq 1/2$. Thus, the agent either wants the information immediately or not at all. Therefore, if v is concave ($r \geq 1/2$) then the agent prefers full concealment and if v is convex ($r \leq 1/2$) then the agent prefers full disclosure. However, in neither case is the agent is information seeking since full concealment and full disclosure are both preferred over any intermediate disclosure date for a linear v, that is, when $r = 1/2$.

Theorem 5, below, gives necessary and sufficient conditions for an NRU agent to prefer (or be averse to) full concealment and sufficient conditions for a preference for full disclosure. The examples above show that, unlike in the SRU case, this characterization does not extend to information seeking.

Definition: *The capacity η is supermodular (*submodular*) if $\eta(S \cup T) + \eta(S \cap T) \geq \eta(S) + \eta(T)$ ($\eta(S \cup T) + \eta(S \cap T) \leq \eta(S) + \eta(T)$) for all $S, T \in \mathcal{S}$; η is strictly supermodular (strictly submodular) if this inequality is strict for all disjoint S, T such that $l(S)l(T) > 0$.*

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

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

(i) $\succeq$ is averse to (prefers) full concealment if and only if v is convex (concave) and η is submodular (supermodular);

(ii) If v is convex (concave) and η is supermodular (submodular) then $\succeq$ prefers (is averse to) full disclosure in every REL with constant immediate consumption.

Theorem 5 provides a limited analogue to Theorem 2, above. SRU agents with a convex v are information seeking, prefer full disclosure, and are averse to full concealment.

By contrast, NRU agents with a convex v and a submodular capacity are averse to full concealment and prefer full disclosure if immediate consumption is constant. Information seeking cannot be guaranteed in either case. Part (ii) of Theorem 6 is weaker than the corresponding characterization in Theorem 2 because it requires a constant intermediate consumption and because it provides only sufficient conditions for a preference for full disclosure. The reason we need to restrict to a constant consumption is that, otherwise, immediate consumption may counteract any effect of information disclosure on the utility paths.

Next, we analyze more closely conditions under which NRU preferences with a concave v and a submodular capacity seek to avoid information disclosure that is too rapid. Lovallo and Kahnemann (2000) examine subjects willingness to live with uncertainty as a function of the attractiveness of the underlying gamble. They 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. The next result captures a related phenomenon in our setting; under certain conditions on the parameters, agents prefer to delay the resolution of uncertainty of attractive lotteries. To analyze this question we consider a subclass of RELs, we call *binary RELs*, defined below.

There is a single intermediate consumption good ($K_1 = \{1\}$) and two terminal prizes (that is, $K_2 = \{1, 2\}$). We ignore immediate consumption and let $\alpha \in \Delta = [0, 1]$ denote the probability of the more desirable prize. Thus, each path is a function $x : [0, 1] \to [0, 1]$. Let B be the set of paths $(x, y) \in D \times D$ such that $x_0 = y_0$, x is non-decreasing and y is non-increasing in t. Any pair $(x, y) \in B$ generates a unique REL P with the property that P_t has the binary support x_t, y_t at each time t:

Lemma B: *For $(x, y) \in B$ there is a unique $P \in \Pi$ such that $P_t\{x_t, y_t\} = 1$ for all t.*

For any $(x, y) \in B$, let P^{xy} denote the corresponding binary REL. For example, if $x, y \in D_3$ such that $x = (1/2, 5/8, 3/4)$ and $y = (1/2, 3/8, 1/4)$ then:

$$P^{xy}(z) = \begin{cases} 3/8 & \text{if } z = (1/2, 5/8, 3/4) \text{ or } z = (1/2, 3/8, 1/4) \\ 1/8 & \text{if } z = (1/2, 3/8, 3/4) \text{ or } z = (1/2, 5/8, 1/4) \\ 0 & \text{otherwise.} \end{cases}$$

Let $(x,y),(w,z) \in B$ with the same initial lottery $w_0 = x_0$. Then, (w,z) is *more informative* than (x,y) if and only if $w \geq x$ and $y \geq z$. Notice that, by the martingale property, we have:

$$P_t^{xy}(x_t)x_t + P^{xy}(y_t)y_t = P_t^{wz}(w_t)w_t + P_t^{wz}(z_t)z_t$$

and, therefore, at each time t, the distribution of the REL P_t^{wz} is a mean preserving spread of P_t^{xy} if and only if $w \geq x$ and $y \geq z$.

The binary REL $(x,y) \in B$ *resolves favorable information too quickly* if there exists $(w,y) \in B$ such that $w \leq x$ and $U(P^{wy}) > U(P^{xy})$. Note that $w \leq x$ implies that $P_t^{wy}(w_t) \geq P^{xy}(x_t)$ and, therefore, the REL (w,y) has a greater probability that at time t the attractive lottery is realized. Hence the REL (w,y) allows a greater chance to "savor" the attractive lottery than the REL (x,y). For any two paths $x,y \in D$ we use the notation xsy for the path that is equal to x prior to time t and equal to y after time t. Thus, $xsy \in D$ is defined as:

$$[xsy]_t = \begin{cases} x_t & \text{if } t < s \\ y_t & \text{if } t \geq s \end{cases}$$

Theorem 6: *Let $\succeq = (u,v,\eta)$ be an NRU such that v is concave, η is dual totally monotone and strictly submodular. For every decreasing $z \in D, \alpha \in \Delta$ such that $1 > \alpha > z_0$ there exists $\epsilon > 0$ such that $(x^\alpha t x^1, x^\alpha t z) \in B$ resolves favorable uncertainty too quickly if $t \leq \epsilon$.*

Theorem 6 considers $(x,y) \in B$ with the property that after time ϵ all favorable information resolves immediately; that is, if the agent receives favorable news after time ϵ then she is sure to get the good prize. Theorem 6 gives conditions under which the agent would prefer an alternative REL that resolves favorable news less quickly, that is, an REL for which some uncertainty remains even after favorable news at some time $t > \epsilon$.

6. News after Good News

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 motivate the definition below, consider the following example. The RELs below all have identical and constant immediate consumption along all paths. As an example, consider RELs with paths in D_3 and $K_2 = \{1, 2\}$. Ignoring the (constant) intermediate consumption, will write $(\alpha_1^\bullet, \alpha_2^\bullet, \alpha_3^\bullet)$ to denote a generic element of D_3 where $\alpha_i^\bullet \in [0, 1]$ is the probability of the more desirable terminal prize 2. Consider the REL R, below, with four equally likely paths

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

In the REL R, the agent receives information about the terminal prize at time $t = 1/3$. The first and second paths start with a probability of .6. At time $t = 1/3$, the agent learns whether the probability 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 $t = 1/3$ while along the third path, the agent arrives at .4 following good news at time $t = 1/3$.

Now, consider a modification of R that yields additional information at time $s = 2/3$. In modification R^{Bs}, the additional information arrives after the history $(.6, .4)$, thus after previous bad news. In modification R^{Gs}, the news arrives after the history $(.2, .4)$, thus after previous good news. Specifically, assume that in R^{Bs} the path $(.6, .4, .4)$ is replaced by the two (equally likely) paths $(.6, .4, .2)$ and $(.6, .4, .6)$ while in R^{Gs} the path $(.2, .4, .4)$ is replaced by the two (equally likely) paths $(.2, .4, .2)$ and $(.2, .4, .6)$. In that case, we say that R^{Gs} has *more news after good news* than R^{Bs}.

Experiments by Karlsson, Loewenstein and Seppi (2009) suggest that some decision makers prefer R^{Gs} over R^{Bs}. Our objective is to show that this ranking is compatible with NRU utility and relate this preference to the parameters of NRU.

Recall that for any two paths $x, y \in D$ we use the notation xsy for the path that is equal to x prior to time t and equal to y after time t. Let $C = \{(\alpha, \beta) \in \Delta^2 : \alpha \succ_0 \beta\}$. For $(\alpha, \beta) \in C$ and $0 < t, a < 1$, let $R \in \Pi(\alpha, \beta, t, a)$ if there exist paths x and y such that:

(i) $R(x) = R(y) > 0$;

(ii) $x_s = y_s = a\alpha + (1-a)\beta$ if and only if $s \in [t, 1]$ for some $t < 1$.

(iii) $x_s \succeq_0 \alpha \succ_0 \beta \succ_0 y_s$ for $s < t$

Thus, $R \in \Pi(\alpha, \beta, a, t)$ contains two equally likely paths[4] that are constant at $\gamma = a\alpha + (1-a)\beta$ on the interval $(t, 1]$. Along the path x, the lottery γ represents the worst lottery, whereas along y, γ represents the best lottery. This implies that along the path y the agent received good news at time t whereas along the path x the agent received bad news at time t.[5]

Let $R \in \Pi(\alpha, \beta, a, t)$ and suppose the agent receives additional information at time $t < s < 1$ that reveals either α or β. The REL R^{Gs} reveals this information along the path y. That is,

$$R^{Gs}(z) = \begin{cases} aR(y) & \text{if } z = ysx^{\alpha} \\ (1-a)R(y) & \text{if } z = ysx^{\beta} \\ R(z) & \text{if } z \neq y, ysx^{\alpha}, ysx^{\beta} \end{cases}$$

The REL R^{Bs} reveals the same information along the path x. That is,

$$R^{Bs}(z) = \begin{cases} aR(y) & \text{if } z = xsx^{\alpha} \\ (1-a)R(y) & \text{if } z = xsx^{\beta} \\ R(z) & \text{if } z \neq x, xsx^{\alpha}, xsx^{\beta} \end{cases}$$

Thus, R^{Bs} reveals the information after previous bad news while R^{Gs} reveals this information after previous good news. Notice that R^{Bs} and R^{Gs} reveal the same information at time s; they differ only in the history that precedes the information revelation. We say that $\succeq$ prefers news after good news if $R^{Gs} \succeq R^{Bs}$.

Theorem 7: *Let $\succeq = (u, v, \eta)$ be an NRU. Then, $\succeq$ prefers news after good news if v is concave (convex) and η is totally monotone (dual totally monotone).*

[4] The assumption that the two paths are equally likely is made for simplicity only. At the expense of a slightly more cumbersome definition, it would suffice if both paths have strictly positive probability.

[5] We assume that x is uniformly above α and y is uniformly below β. It would be sufficient to require that x_s is uniformly above y_s for all $s < t$ as long as neither x_s nor y_s are "between" α and β.

Theorem 7 provides conditions under which NRU agents will exhibit the Ostrich effect identified 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. First, R^{Gs} and R^{Bs} provide exactly the same information; they only differ in the history preceding the information. Second, the information is small relative to the good news or bad news that precedes it. We capture this constraint by the requirement that x is above α prior to t while y is below β prior to t.

If v is concave and η is totally monotone then the agent prefers full concealment (Theorem 5) and, therefore, is avers to information. In that case, additional information at time s reduces the agent's utility, but it does so less if it follows good news. If v is convex and η is dual totally monotone then 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 if v is linear and $\eta = \eta^1 + \eta^2$ where η^1 is totally monotone and η is dual totally monotone then the agent prefers news after good news.

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 $u : \Delta \to [0, 1]$, $u^\circ : \Delta_1 \to [0, 1]$ and $u^\bullet : \Delta_2 \to [0, 1]$ 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 Δ. Since $d_p(R^\alpha, R^\beta) = ||\alpha - \beta|| = d_s(x^\alpha, x^\beta)$, Axiom 3 implies that $\succeq_*$ is continuous. In addition, Axiom 4 states that $\succeq_*$ satisfies independence on the mixture space Δ. Hence, there exists a linear function u that represents $\succeq_*$.

Since Δ 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,

$$u(\alpha) + u(\beta) = u(\alpha^\circ, \beta^\bullet) + u(\beta^\circ, \alpha^\bullet). \tag{A1}$$

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 and linear function $V : \Pi \to [0, 1]$ such that V represents $\succeq$. Moreover, if $P_t \succeq_0 Q_t$ for all t, then $V(P) \geq V(Q)$.*

Proof: The set Π 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 part, 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 α, β. 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. □

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

Let the lotteries σ, ς respectively maximize and minimize u and let $\beta = (\sigma^\circ, \varsigma^\bullet)$. By non-degeneracy and Lemma 1, $u(\beta) > u(\varsigma) = 0$. For $t \in [0, 1)$, let Q^t denote the degenerate REL that assigns probability 1 to the evolving lottery x^t such that $x^t_s = \beta$ for all $s < t$ and $x^t_s = \varsigma$ otherwise. Then, let $\lambda(1) = 1$ and, for $0 \leq t < 1$, define $\lambda(t) = V(Q^t)/V(R^\beta)$. By Axiom 5, $\lambda(t)$ 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 λ 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

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

Hence, the definition of λ 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 $W(z) = \sum_i a_i \int z_i(t)\, d\lambda(t) = \int \sum_i a_i z_i(t)\, d\lambda = \int z\, d\lambda$. □

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

Proof: For any $0 = s_0 < s_1 < \cdots < s_{k-1} < s_k = 1$, let $z_i(t) = v(u(\beta))$ 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 $z_{k+1} = 1 - \sum_{i=1}^{k} a_i$. 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. □

The proof of the 'if' part of the representation theorem is straightforward. To prove the uniqueness assertion, assume that (u, v, λ) is a representation of some non-degenerate $\succeq$. Let $(\hat{u}, \hat{v}, \hat{\lambda})$ be a second representation. Pick α, β 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 Δ, agree at two distinct points α, β, 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 α, β such that $\alpha^\bullet = \beta^\bullet$ and $u(\alpha) < u(\beta)$. Since, $\succeq$ is non-degenerate, such α, β must exist. Let $x^t_s = \alpha$ for $s < t$ and $x^t_s = \beta$ for $s \geq t$. Clearly, $R^{x^t} \in \Pi$ for all $t < 1$. The representation yields

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

as desired. □

8. Appendix B: 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 ι REL if $PA_\iota = 1$. Let Π_ι be the set of all ι 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 σ such that $u(\sigma) = 1$, ς 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 $\iota = (S_1, \ldots, S_k)$ with $A_\iota \neq \emptyset$ and $r \in (0,1)$, there exists $x^n \in A_\iota$ converging to x^α such that $u(\alpha) = r$.*

Proof: First, we consider the case $r \geq u^\bullet(\sigma^\bullet)$. Choose $\alpha^\circ \in \Delta_1$ such that $u^\circ(\alpha^\circ) = r - u^\bullet(\sigma^\bullet)$. Lemma 1 ensures that such an α° exists. Let $\beta^n_i = ((1-n^{-i})\alpha^\circ + n^{-i}\varsigma^\circ, \sigma^\bullet) \in \Delta$. Then, define x^n as follows: $x^n_t = \beta^n_i$ if and only if $t \in S_i$. Clearly, $x^n \in A_\iota$ for all n and x^n converges to x^β where $\beta = (\alpha^\circ, \sigma^\bullet)$. Hence $R^{x^n} \in \Pi_\iota$ and R^{x^n} converges to R^β. Since $u(\beta) = r$, we have completed the proof for this case. The proof for the $r < u^\bullet(\sigma^\bullet)$ case is symmetric and omitted. □

Lemma 6: *There is a continuous, linear and onto $V : \Pi \to [0,1]$ such that V 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 $Q \succ P$. Lemma 5 ensures the existence of $P^n, Q^n \in \Pi_\iota$ converging respectively to x^α, x^β 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 α such that $u(\alpha) = r$. Lemmas 1 and 6 ensure that v is a well-defined element of Λ. 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:

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

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

Fix any ordered partition $\iota = (S_1, \ldots, S_k)$ and identify each $z \in D_\iota$ with $\hat{z} = (\hat{z}^1, \ldots, \hat{z}^k) \in [0,1]^k$ such that $\hat{z}^j = z_t$ for some $t \in S_j$. Proceeding as in the proof of Lemma 5 above, it is easy to find $\epsilon > 0$ and $R^{x_1}, \ldots, R^{x_{k-1}} \in \Pi_\iota$ such that $R^{x_i} \in \Pi_\iota$ implies $f(R^x) = z^j$ defined as follow: $z_n = (k+1-n)\epsilon$ for $n < j$ and $z_n = (k-n)\epsilon$ for $n \geq j$. Let $e = (1, \ldots, 1) \in [0,1]^k$. Let A be the following $k \times k$ matrix: for $< k$, the i'th row is $z_i = f(R^{x_i})$ and the k't row of A is e. Then, 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:

$$A\,\mathbf{x} = \mathbf{v} \tag{A1}$$

where $\mathbf{v}$ is a column vector such that $\mathbf{v}_i = V(R^{x_i})$. Let $\eta_\iota(S_i) = \mathbf{x}_i$ where $\mathbf{x}$ is the solution the system of equations (A1). Hence, we have $\sum_i \mathbf{x}_i = 1$ and

$$\sum_i z_i^j \eta_\iota(S_j) = V(R^{x_i})$$

for all i. Call $z \in D_\iota$ normal if

$$\sum_i z_i \eta_\iota(S_j) = V(Q) \tag{A2}$$

for $z = f(Q)$. Hence, z_i is normal for $i = 1, \ldots, k-1$. Then, since the first $k-1$ rows of A are linearly independent, the affine hull, H, of these $k-1$ vectors has dimension $k-1$. The linearity of V ensures that D_ι viewed as a subset of $I\!R^k$ has at most dimension $k-1$. Therefore, $D_\iota \subset H$. Hence, for any $z \in D_\iota$, there exists $a_i \in I\!R$ for $i = 1, \ldots, k-1$ such that $\sum_i^{k-1} a_i z_i = z$ and $\sum_i^{k-1} a_i = 1$. Then, arguing as in the proof of Lemma 3, we conclude that every $z \in D_\iota$ is normal.

For all $S \in \mathcal{S}$, let $\eta(S) = \eta_\iota$ for $\iota = (S, [0,1]\backslash S)$ and set $\eta(\emptyset) = 0$. The monotonicity of η follows from Lemma 2. Next, we show that for all $\iota = (S_1, \ldots, S_n)$, $S = S_i$ implies $\eta(S) = \sum_{j=1}^i \eta_\iota(S_j)$. To see this, for $\iota = (S_1, \ldots, S_n$, define $z(\epsilon)$ such that $z^j(\epsilon) = r - n\epsilon$ for $n \leq j$ and $z^j = r' - n\epsilon$, where $r > r'$, $r, r' \in (0,1)$ and $\epsilon < (r - r')/n$. Choose a convergent sequence $Q^m \in \Pi_\iota$ such that $f(Q^m) = z(1/m)$. Lemma 2 ensures that such sequence Q^m exists. Continuity of V ensures that $V(Q^m)$ converges to $Q \in D_{\iota^*}$ where $\iota^* = (S_i, [0,1]\backslash S_i)$. Then, equation (A2) implies

$$r \sum_{j=1}^{i} \eta_\iota(S_j) + r'(1 - \sum_{j=1}^{i} \eta_\iota(S_j) = \lim V(Q^m)$$
$$r\eta(S_i) + r'(1 - \eta(S_i) = V(Q)$$

Since V is continuous, we have $\lim V(Q^m) = V(Q)$ and hence $\eta(S_i) = \sum_\iota(S_i)$ as desired.

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

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

as desired. $\square$

The proof of the if statement is again straightforward. Let (u, v, η) 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. □

9. Appendix C: Proof of Theorem 2

The equivalence of (i) and (iv) is immediate. 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

$$\begin{aligned} V(\varepsilon[P]) &= \sum_x \int_0^1 v(u(x_t))d\lambda(t)\, \varepsilon[P](x) \\ &= \sum_x \sum_i v(u(x_{s_{i-1}}))\lambda_i\, \varepsilon[P](x) \\ &= \sum_i \lambda_i \sum_x v(u^\circ(x^\circ_{s_{i-1}}) + u^\bullet(x^\bullet_{s_{i-1}+\varepsilon}))\, P(x) \\ &\geq \sum_i \lambda_i \sum_x v(u^\circ(x^\circ_{s_{i-1}}) + u^\bullet(x^\bullet_{s_{i-1}}))\, P(x) \\ &= V(P) \end{aligned}$$

with the convention that $x^\bullet_t = x^\bullet_1$ for every $t > 1$. The inequality above follows from the martingale property, the linearity of $u^\bullet$ and the convexity of v. Hence the SRU is information-seeking and therefore (i) implies (ii) and (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, $u_2 - u_1 < \max\{u^\bullet(\alpha^\bullet) : \alpha^\bullet \in \Delta_2\} - \min\{u^\bullet(\alpha^\bullet) : \alpha^\bullet \in \Delta_2\}$. Then we can take $\alpha^\bullet, \beta^\bullet \in \Delta_2$ and $\gamma^\circ \in \Delta_1$ with $u^\bullet(\alpha^\bullet) + u^\circ(\gamma^\circ) = u_1$ and $u^\bullet(\beta^\bullet) + u^\circ(\gamma^\circ) = u_2$. Let x be an evolving lottery with $x^\circ(t) = \gamma^\circ$ for all t, $x^\bullet(t) = a\alpha^\bullet + (1-a)\beta^\bullet$ for $t < 1/2$ and $x(t) = x^{\alpha^\bullet}$ for $t \geq 1/2$. Also, let y be an evolving lottery with $y^\circ(t) = \gamma^\circ$ for all t, $y^\bullet(t) = a\alpha^\bullet + (1-a)\beta^\bullet$ for $t < 1/2$ and $y(t) = x^{\beta^\bullet}$ for $t \geq 1/2$. Finally let P be the REL with $P(\{x\}) = a =$

$1 - P(\{y\})$. Hence, P offer constant consumption lottery γ° throughout, and the decision maker learns if she gets the prize lottery $\alpha^\bullet$ or $\beta^\bullet$ at time 1/2. Now for any $0 < \varepsilon \leq 1$ we have $V(P) > V(\varepsilon[P])$. Thus (ii) (or (iii)) imply (i). □

10. Appendix D: Proofs of Theorems 4 and 5

Lemma D1: *Let X be a finite set and let $\lhd$ be a partial order on X. Then, for any real-valued function H on X, there exists a unique function h such that*

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

Proof: The h can be defined inductively as follows. Let $l(x) = 0$ if x is the first element of $\lhd$. Then, let $l(x) = n$ if and only if x has an immediate predecessor y such that $l(y) = n-1$. Then, set $h(x) = H(x)$ for the first element x and $h(x) = H(x) - \sum_{x \neq y \lhd x} h(y)$. The uniqueness of h is obvious. □

When the display equation in Lemma D1 holds for all x, we call h the $z\lhd$-derivative of H. Fix $n > 1$ and $\theta = \{t_0 < t_1, \ldots, t_{n-1} < t_n\}$ such that $0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = 1$. Let $\mathcal{S}_\theta$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as the finite union of sets of the form $[t_i, 1]$ or $[t_i, t_j)$ for some i and some $j \neq n$. It is easy to verify that $\mathcal{S}_T$ ordered by set inclusion is a lattice. Let η be any continuous capacity and let H denote its restriction to $\mathcal{S}_\theta$. Let η^+ be the dual of η and let H^+ denote its restriction to $\mathcal{S}_\theta$. Finally, let h, h^+ be the $\subset$-derivatives of H and H^+ and let f be any real-valued $\mathcal{S}_\theta$ measurable function on the unit interval. Then, we claim

$$\begin{aligned} \int f dH &= \sum_{T \in \mathcal{S}_\theta} h(T) \min_{s \in T} f(s) \\ &= \sum_{T \in \mathcal{S}_\theta} h^+(T) \max_{s \in T} f(s) \end{aligned} \qquad (A3)$$

To see why equation (A3) 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

$$\int f dH = \sum_{i=1}^{n} r_i [H(\bigcup_{j=0}^{i} T_j) - H(\bigcup_{j=0}^{i-1} T_j)]$$
$$= \sum_{i=1}^{n} r_{n+1-i} [H^+(\bigcup_{j=0}^{i} T_{n+1-j}) - H(\bigcup_{j=0}^{i-1} T_{n+1-j})]$$

Let $S^i = \{T \in S_\theta \,|\, T \subset \bigcup_{j=0}^{i} T_j$ and $T \not\subset \bigcup_{j=0}^{i-1} T_j\}$ and let $S^{+i} = \{T \in S_\theta \,|\, T \subset \bigcup_{j=0}^{i} T_{n+1-j}$ and $T \not\subset \bigcup_{j=0}^{i-1} T_{n+1-j}\}$

Then, Lemma D1 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 for all $T \in S_\theta$, $T \in S^i$ if and only if $\min_{s \in T} f(s) = r_i$ and $T \in S^{+i}$ if and only if $\max_{s \in T} f(s) = r_i$. Thus, we have shown that equation (A3) holds.

Recall that Φ is the set of all step functions from $[0,1]$ to $[0,1]$. Let Π' be the set of all probabilities on Φ. For any $x \in D$, let $\rho^x \in \Phi$ be such that $[\rho^x]_t = v(u(x_t))$. 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 $A^{Sc} = \{u \in \Phi, \,|\, u_t \geq c$ for all $t \in S\}$. Similarly, let $A_{Sc} = \{u \in \Phi, \,|\, u_t \leq c$ 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 $c \in [0,1]$. Theorem 4 is then equivalent to the following lemma:

Lemma D2: *Let $P, Q \in \Pi'$. Then, P u-dominates (l-dominates) Q implies $E_P \int u d\eta \geq E_Q \int u d\eta$ for all $P, Q \in \Pi'$ if and only if η is totally monotone (dual totally monotone).*

Proof: Suppose P dominates Q implies $E_P \int u d\eta \geq E_Q \int u d\eta$ for all $P, Q \in \Pi'$. Let $M = \{1, \ldots, m^+\}$, $M = \{1, \ldots, m\}$ for $m^+ > m \geq 1$ and $S_1, \ldots, S_{m^+} \in \mathcal{S}$ be a partition of $[0,1]$. Let $\mathcal{M} = \{N \subset M\}$, $\mathcal{M}^* = \{N \subset M^+\}$ and let $S_L = \bigcup_{i \in L} S_i$. For $L \subset M$ and

$i \in M^+ \backslash M$, let $v_t^{Li} = 1$ if $t \in S_L$ or ($t \in S_i$ and $m - |L| + 2$ is even); otherwise, $v_t^{Li} = 0$. Similarly, let $w_t^{Li} = 1$ if $t \in S_L$ or ($t \in S_i$ and $m - |L|$ is odd); otherwise, let $v_t^{Li} = 0$. Let $P^i(u) = 2^{-m}$ if $u = v^{Li}$ for some L and $P^i(u) = 0$ otherwise. Similarly, let $Q^i(u) = 2^{-m}$ if $u = w^{Li}$ for some L and $Q^i(u) = 0$ otherwise.

Clearly, $P^i A_0^S = Q_i A_0^S = 1$ for all S. For all S_L such that $L \in M$, $i \in M^+ \backslash M$, $u_t^{Li} = v_t^{Li}$ for all $t \in S$. Also, for S such that $S \not\subset S_i \cup S_M$ and $c > 0$, $P^i A_c^S = Q_i A_c^S = 0$. Otherwise; that is, if $S \subset S_i \cup S_M$ and $S \cap S_i \neq \emptyset$, we have $P^i A_c^S = Q^i A_c^S + 2^{-m}$ if

$$l := |\{j \mid S \cap S_j \neq \emptyset\}| = m$$

and $P^i A_c^S = Q^i A_c = 2^{-l}$ otherwise. Hence, $P^i A_c^S \geq Q^i A_c^S$ for all S. That is, P^i u-dominates Q^i. Also, by (A3),

$$E_{P^i}\left[\int u d\eta\right] = \sum_{\emptyset \neq L \in \mathcal{M}^+} h(S_L) \min\{u_t \mid t \in S_L\}$$
$$E_{Q^i}\left[\int u d\eta\right] = \sum_{\emptyset \neq L \in \mathcal{M}^+} h(S_L) \min\{u_t \mid t \in S_L\}$$

where h is the derivative of η on $\{S_1, \ldots, S_{m^+}\}$. Straightforward but tedious calculations reveal that

$$E_{P^i}\left[\int u d\eta\right] - E_{Q^i}\left[\int u d\eta\right] = h(S_M)$$

and therefore, $E_P \int u d\eta \geq E_Q \int u d\eta$ implies $h(T) \geq 0$ for every partition $S_1, \ldots S_{m^+}$ and $T = \bigcup_{j \in M} S_j$. Hence, η is totally monotone.

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

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

Hence P u-dominates Q implies

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

for all c.

Define the following matching problem: $X = \{(S_N, c) \,|\, \emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(P)\}$, $Y = \{(S_N, c) \,|\, \emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(Q)\}$, $\rho(i,j) = 1$ if $i = (S_N, c)$, $j = (S_N, \hat{c})$ and $c \geq \hat{c}$ and $\rho(i,j) = 0$ otherwise. Finally, $b(i) = \sum_{u: u_t = c \text{ for all } t \in S_N} P(u)h(S_N)$ for all $i = (S_N, c) \in X$ and $b(j) = \sum_{u: u_t = c \text{ for all } t \in S_N} P(u)h(S_N)$ for all $j = (S_N, c) \in Y$. Equation $(**)$ 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 χ.

By (A3),

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

The proof for the l-domination/dual totally monotone case is symmetric and omitted. To prove assertion regarding binary elements of Π', assume that P u-dominates Q implies $E_P \int u d\eta \geq E_Q \int u d\eta$ for all binary $P, Q \in \Pi'$. For any $w, w' \in X$, let $w \vee w'$ denote $\hat{w} \in X$ such that $w_t = \max\{w_t, w'_t\}$ and $w \wedge w'$ denote $\hat{w} \in X$ such that $\hat{w}_t = \min\{w_t, w'_t\}$ for all t. Also, let $w \geq w'$ mean $w_t \geq w'_t$ for all $t \in [0,1]$.

Then, let for any S, T such that $S \neq T$, consider, let $u_t = 1$ if $t \in S$ and $u_t = 0$ otherwise. Similarly, let $v_t = 1$ if $t \in T$ and $v_t = 0$ otherwise. Let $P(u \vee v) = P(u \wedge v) = Q(u) = Q(v) = 1/2$. It is easy to verify that P l-dominates Q. Hence, $.5\eta(S \cup T) + .5\eta(S \cap T) = E_P\left[\int w d\eta\right] \geq E_Q\left[\int w d\eta\right] = .5\eta(S) + .5\eta(T)$ proving that η is supermodular.

To prove the converse, suppose η is supermodular. We claim that if P u-dominates Q and both P and Q are binary, then $P(\hat{u}) = P(\hat{v}) = Q(u) = Q(v) = 1/2$ for some $\hat{u} \neq \hat{v}$ and $u \neq v$ such that $\hat{u} \geq u \vee v$ and $\hat{v} \geq u \wedge v$. To see this, take $\hat{u} \neq \hat{v}$ and $u \neq v$ such that $P(\hat{u}) = P(\hat{v}) = Q(u) = Q(v) = 1/2$. Clearly, if either $\hat{u} \not\geq u \wedge v$ or $\hat{v} \not\geq u \wedge v$, then P cannot u-dominate Q. So, we must have $\hat{u} \geq u \wedge v$ and $\hat{v} \geq u \wedge v$. But, if we have neither $\hat{u} \geq u \vee v$ nor $\hat{v} \geq u \vee v$, then again P cannot u-dominate Q. So, we must have $\hat{u} \geq u \vee v$ or $\hat{v} \geq u \vee v$. Assume, without loss of generality that $\hat{u} \geq u \vee v$ to establish the claim.

Clearly, $w \geq w'$ implies $\int w d\eta \geq \int w' d\eta$. So, to conclude the proof, it is enough to show that

$$\int u \vee v \, d\eta + \int u \wedge v \, d\eta \geq \int u d\eta + \int v d\eta$$

whenever η is supermodular. To see this, let $S^w(\zeta) = \{t \in [0,1] \,|\, w_t \geq \zeta\}$ and note that $S^{u\vee v}(\zeta) = S^u(\zeta) \cap S^v(\zeta)$ and $S^{u \wedge v}(\zeta) = S^u(\zeta) \cup S^v(\zeta)$. Hence, the super modularity of η implies $\eta(S^{u \vee w}) + \eta(S^{u \wedge v}) \geq \eta(S^u) + \eta(S^v)$. Therefore,

$$\begin{aligned}\int u \vee v \, d\eta + \int u \wedge v \, d\eta &= \int \eta(S^{u\vee v}(\zeta)) + \eta(S^{u \wedge v}(\zeta)) d\zeta \\ &\geq \int \eta(S^u(\zeta)) + \eta(S^v(\zeta)) d\zeta = \int u d\eta + \int v d\eta\end{aligned}$$

as desired.

The proof for the l-domination/submodular case is symmetric and omitted. □

Proof of Theorem 7: Let R, R^{Bs}, R^{Gs} satisfy the properties of the Theorem with $0 < b = R(x) = R(y)$. Then,

$$\begin{aligned}\frac{1}{b}\left(U(R^{Bs}) - U(R)\right) &= a \int (v(u([xsx^\alpha]_{t'}))d\eta) + (1-a)\int \left(v(u([xsx^\beta]_{t'}))d\eta\right) \\ &\quad - \int (v(u(x_{t'}))d\eta) \\ \frac{1}{b}\left(U(R^{Gs}) - U(R)\right) &= a \int (v(u([ysx^\alpha]_{t'}))d\eta) + (1-a)\int \left(v(u([ysx^\beta]_{t'}))d\eta\right) \\ &\quad - \int (v(u(y_{t'}))d\eta)\end{aligned}$$

Choose $\theta = \{0 = t_0 < t_1, \ldots, t_k = t\}$ so that $R(x) > 0$ implies $x_t = x_s$ for all $t, s \in [t_i, t_{i+1})$, for $i \leq k-1$. Let $\theta^* = \theta \cup \{s, 1\}$. Let $\mathcal{S}_\theta$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as the finite union of sets of the form $[t_i, t_j)$ for some $t_i, t_j \in \theta$ and, similarly, let $\mathcal{S}_{\theta^*}$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as the finite union of sets of the form $[t_i, t_j)$ or $[t_i, 1]$ for some $t_i < t_j \in \theta^*$. Let $T_1 = [t, s)$ and $T_2 = (s, 1]$ and note that $T_1, T_2 \in \mathcal{S}_{\theta^*} \backslash \mathcal{S}_\theta$. Let H denote the restriction of η to $\mathcal{S}_{\theta^*}$. Let η^+ be the dual of η and let H^+ denote its restriction to $\mathcal{S}_{\theta^*}$. Finally, let h, h^+ be the $\subset$-derivatives of H and H^+.

First, consider the case where η is totally monotone. Let $d^1 = av(\alpha) + (1-a)v(\beta) - v(\gamma))$ and let $d^2 := av(\gamma) + (1-a)v(\beta) - v(\gamma))$. By A3,

$$\frac{1}{b}\left(U(R^{Bs}) - U(R)\right) = d_2\left(\sum_{S\in\mathcal{S}_\theta} h\left(S\cup T_1\cup T_2\right) + h(T_1\cup T_2)\right) + d_1\left(\sum_{S\in\mathcal{S}_\theta} h\left(S\cup T_2\right) + h(T_2)\right)$$

$$\frac{1}{b}\left(U(R^{Gs}) - U(R)\right) = d_2 h(T_1\cup T_2) + d_1 h(T_2)$$

Since η is totally monotone, $h \geq 0$. Note further that $d_2 < 0$ and, since v is concave, $d_1 \leq 0$. It follows that $U(R^{Gs}) \geq U(R^{Bs})$ as desired.

For the case where η is dual totally monotone, let $d_3 := av(\alpha) + (1-a)v(\gamma) - v(\gamma)$. By A3,

$$\frac{1}{b}\left(U(R^{Bs}) - U(R)\right) = d_3 h^+(T_1\cup T_2) + d_1 h^+(T_2)$$

$$\frac{1}{b}\left(U(R^{Gs}) - U(R)\right) = d_3\left(\sum_{S\in S_\theta} h^+\left(S\cup T_1\cup T_2\right) + h^+(T_1\cup T_2)\right) + d_1\left(\sum_{S\in S_\theta} h^+\left(S\cup T_2\right) + h(T_2)\right)$$

Since η 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, $U(R^{Gs}) \geq U(R^{Bs})$ as desired. □

11. Appendix E: Proof of Theorems 5 and 6

Lemma E1: *Let $f : [0,1] \to [0,1]$ be strictly increasing and continuous. For every $\epsilon > 0$ there exists $\delta > 0, a, b \in [\delta, 1-\delta]$, such that*

(i) $(1+\epsilon)(f(a+\delta) - f(a)) \geq f(a) - f(a-\delta)$.

(ii) $(1+\epsilon)(f(b) - f(b-\delta)) \geq (f(b+\delta) - f(b))$.

Proof: We will prove (ii); (i) follows from a symmetric argument. If there is $a, b \in [0,1]$ such that $f((a+b)/2) \geq f(a)/2 + f(b)/2$ then the inequality is immediate. Hence, assume that $f((a+b)/2) \leq f(a)/2 + f(b)/2$ for all $a, b \in [0,1]$. Since f is continuous it follows that f is convex. Since f is increasing, it follows that f is differentiable almost everywhere.

Let $a \in (0,1)$ be such that $f'(a)$ exists. Since $0 < f(a) - f(0) \leq af'(a)$ it follows that $f'(a) > 0$. By Taylor's theorem,

$$f(a+\delta) - f(a) = f(a) - f(a-\delta) + \delta(h(\delta) - h(-\delta))$$

for some real function h such that $\lim_{\delta \to 0} h(\delta) = 0$. Since $f'(a) > 0$ and f is convex, it follows that $f(a+\delta) - f(a) \geq \delta f'(a)$. Finally, since $f'(a) > 0$, there exists $\delta > 0$ so that $h(\delta) - h(-\delta) \leq \epsilon f'(a)$. □

Proof of Theorem 5(i): First, we prove that $P \succeq R^{\bar{x}(P)}$ implies that v is convex and η is sub-modular. That v must be convex is an immediate consequence of the fact that the induced preference $\succeq_0$ on M must be risk loving. It remains to show that η is submodular. Let $S, S' \in \mathcal{S}$ such that $S \cap S' = \emptyset$ and $\eta(S) + \eta(S') < \eta(S \cup S') + \epsilon$ for some $\epsilon > 0$. Let $x, y \in D$ be the following paths:

$$x_t = \begin{cases} (\alpha^\circ, \alpha^\bullet) & \text{if } t \in S \\ (\beta^\circ, \beta^\bullet) & \text{if } t \in S^c \end{cases} \; y_t = \begin{cases} (\alpha^\circ, \alpha^\bullet) & \text{if } t \in S' \\ (\beta^\circ, \beta^\bullet) & \text{if } t \in S'^c \end{cases}$$

Let $P \in \Pi$ be such that $P(x) = P(y) = 1/2$ and let $Q \in \Pi$ be such that $Q((x+y)/2) = 1$. Choose $\delta > 0, a \in [\delta, 1-\delta]$ so that the conclusion of Lemma E1 holds for the function v and the ϵ defined above. Choose α so that $u(\alpha) = a + \delta$ and choose β so that $u(\beta) = a - \delta$. Then,

$$\begin{aligned} 2(U(P) - v(a-\delta)) =& (\eta(S) + \eta(S'))\,(v(a+\delta) - v(a-\delta)) \\ &\leq (2+\epsilon)(\eta(S) + \eta(S'))(v(a) - v(a-\delta)) \\ 2(U(Q) - v(a-\delta)) =& 2\eta(S \cup S')(v(a) - v(a-\delta)) \end{aligned}$$

and, therefore,

$$\begin{aligned} 2U(Q) - 2U(P) &\leq 2(v(a) - v(a-\delta))\left(\eta(S \cup S') - \eta(S) - \eta(S') - \frac{\epsilon}{2}(\eta(S) + \eta(S'))\right) \\ &\leq (v(a) - v(a-\delta))\epsilon \end{aligned}$$

and therefore $U(Q) > U(P)$. It follows that $\eta(S) + \eta(S') \geq \eta(S \cup S')$ for all S, S' with $S \cap S' = \emptyset$ which implies that η is submodular.

For the converse, assume that v is convex and η is submodular. Let $P \in \Pi$ and choose $\theta = \{0 = t_0 < t_1, \ldots, t_k = 1\}$ so that $P(x) > 0$ implies $x_t = x_s$ for all $t, s \in [t_i, t_{i+1})$, for $i \leq k-1$. Let $\mathcal{I}_\theta = \{[0,t_1),[t_1,t_2),\ldots,[t_{k-1},1]\}$ be the corresponding collection of intervals. Let $\mathcal{S}_\theta$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as a union of elements of $\mathcal{I}$ and let H denote the restriction of η to $\mathcal{S}_\theta$. Let $D_P := \{x : P(x) > 0\}$ be the paths in the support of P. Since each $x \in D_P$ is constant for each $I \in \mathcal{I}_\theta$, we can identify each element of D_P with a vector $(x_I)_{I\in\mathcal{I}_\theta}$ (such that $x_I := x_t$ for $t \in I$).

By a standard argument (Shapley (1971), Schmeidler (1989)), there exists a convex set of probabilities L on the finite set $\mathcal{I}_\theta$ such that

$$\int_{[0,1]} v(u(x))d\eta = \int_{\mathcal{I}_\theta} v(u(x_I))dH = \max_{\ell\in L} \sum_{I\in\mathcal{I}_\theta} v(u(x_I))\ell(I) \qquad (E1)$$

Note that for $Q = R^{\bar{x}(P)}$ the set D_Q contains a single path $\bar{x}(P)$ that is constant for all $t \in I, I \in \mathcal{I}_\theta$. Let $\hat{\ell}$ be the probability (in L) that solves

$$\max_{\ell\in L} \sum_{I\in\mathcal{I}_\theta} v(u([\bar{x}(P)]_I))\ell(I)$$

Since v is convex, u is linear and $[\bar{x}(P)]_I = \sum_{x\in D_P} x_I P(x)$ it follows that for all $I \in \mathcal{I}_\theta$

$$v(u(\hat{x}_I)) \leq \sum_{x\in D_P} v(u(x_I))P(x)$$

Therefore,

$$\begin{aligned}\sum_{x\in D_P} \max_{\ell\in L} \left(\sum_{I\in\mathcal{I}_\theta} v(u(x_I))\ell(I)\right) P(x) &\geq \sum_{I\in\mathcal{I}_\theta} \left(\sum_{x\in D_P} v(u(x_I))P(x)\right) \hat{\ell}(I) \\ &\geq \sum_{I\in\mathcal{I}_\theta} v(u(\hat{x}_I))\hat{\ell}(I)\end{aligned}$$

as desired.

The case of a concave v and a supermodular η is analogous and, therefore, omitted.□

Proof of Theorem 5(ii): Assume that v is concave and η is submodular. Let $P \in \Pi$ and choose $\theta = \{0 = t_0 < t_1,\ldots,t_k = 1\}$ so that $P(x) > 0$ implies $x_t = x_s$ for all $t,s \in [t_i,t_{i+1})$, for $i \leq k-1$. Let $\mathcal{I}_\theta = \{[0,t_1),[t_1,t_2),\ldots,[t_{k-1},1]\}$ be the corresponding collection of intervals. Let $\mathcal{S}_\theta$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as a union of elements of $\mathcal{I}$ and let H denote the restriction of η to $\mathcal{S}_\theta$. Let $D_P := \{x : P(x) > 0\}$ be the paths in the support of P. Since each $x \in D_P$ is constant for each $I \in \mathcal{I}_\theta$, we can identify each element of D_P with a vector $(x_I)_{I \in \mathcal{I}_\theta}$ (such that $x_I := x_t$ for $t \in I$). By a standard argument (Shapley (1971)) there exists a convex set of probabilities L on the finite set $\mathcal{I}_\theta$ such that

$$\int_0^1 v(u(x_t))d\eta = \int_{I_\theta} v(u(x_I))dH = \max_{\ell \in L} \sum_{I \in \mathcal{I}_P} v(u(x_I))\ell(I) \tag{E1}$$

Note that for $Q = R^{P_1}$ all paths are constant and, therefore, for all $y \in D_Q$, and all $\ell, \hat{\ell} \in L$,

$$\sum_{I \in \mathcal{I}_\theta} v(u(y_I))\ell(I) = \sum_{I \in \mathcal{I}_\theta} v(u(y_I))\hat{\ell}(I)$$

On each $I \in \mathcal{I}_\theta$, the paths $x \in D_P$ are constant. Therefore, we can define $P_I \in M$ such that $P_I := P_t$ for $t \in I$. Since P is a martingale, it follows that $P_1 = Q_I$ is a mean preserving spread of P_I for all $I \in \mathcal{I}_\theta$. Therefore, the concavity of v and the linearity of u imply that for all $I \in \mathcal{I}_\theta$,

$$\sum_{x \in D_P} v(u(x_I))P(x) \geq \sum_{y \in D_Q} v(u(x_I))Q_I(y)$$

Therefore, for an arbitrary $\hat{\ell} \in L$, we have:

$$\begin{aligned} U(P) = \sum_{x \in D_P} \max_{\ell \in L} \left(\sum_{I \in \mathcal{I}_\theta} v(u(x_I))\ell(I) \right) P(x) &\geq \sum_{I \in \mathcal{I}_\theta} \left(\sum_{x \in D_P} v(u(x_I))P(x) \right) \hat{\ell}(I) \\ &\geq \sum_{I \in \mathcal{I}_P} \left(\sum_{y \in D_Q} v(u(y_I))Q(y) \right) \hat{\ell}(I) \\ &= U(Q) \end{aligned}$$

The proof that $U(Q) \geq U(P)$ if v is convex and η is supermodular is analogous and, therefore, omitted. □

Proof of Lemma 1: Choose $\theta = \{0 = t_0 < t_1, \ldots, t_k = 1\}$ so that $x_t = x_s, y_t = y_s$ for all $t, s \in [t_i, t_{i+1})$. Let $D_{xy} := \{w : w_t \in \{x_t, y_t\}\}$ and note that each $w \in D_{xy}$ can be identified with a vector $(w_i)_{i=1}^k$ where $w_i = w_t$ for $t \in [t_{i-1}, t_i)$. For $w \in D_{xy}$,

$i \in \{2, \ldots, k\}$, define $p_i(w)$ as follows: if $x_i = y_i$ then $p_i(w) = 1$; if $x_i > y_i$ and $w_i = x_i$ then $p_i(w) = (w_{i-1} - y_i)/(x_i - y_i)$; if $x_i > y_i$ and $w_i = y_i$ then $p_i(w) = (x_i - w_{i-1})/(x_i - y_i)$. Since $w_i \in [y_{i+1}, x_{i+1}]$ it follows that $p_i(w) \in [0, 1]$. Let P be the REL such that $P(w) = \prod_{i=2}^{k} p_i(w)$ if $w \in D_{xy}$ and $P(w) = 0$ otherwise. It is easy to check that P is the unique REL that satisfies the conditions of Lemma 1.□

Proof of Theorem 6: For $k \geq 2$, choose $\theta = \{0 = t_0 < t_1 < \ldots < t_k = 1\}$ so that $z_t = z_s$ for all $t, s \in [t_i, t_{i+1})$ and. Let $\mathcal{I}_\theta = \{[0, t_1), [t_1, t_2), \ldots, [t_{k-1}, 1]\}$ be the corresponding collection of intervals. Let $\mathcal{S}_\theta$ be the subset of $\mathcal{S}$ consisting of all sets that can be written as a union of elements of $\mathcal{I}_\theta$.

Recall that $1 > \alpha > z_0$. Let $x_\epsilon = x^\alpha \epsilon x^1, y_\epsilon = x^\alpha \epsilon z$ and $w_\epsilon = \alpha\epsilon\left(x^\delta t_1 x^1\right)$ for $1 > \delta > \alpha$. Let $Q = \lim_{\epsilon \to 0} P^{x_\epsilon y_\epsilon}$ and let $Q^\delta = \lim_{\epsilon \to 0} P^{w_\epsilon y_\epsilon}$. It is straightforward to show that Q, Q^δ are RELs with $EP_t = \alpha = EQ_t^\delta$ for all t. Since U is continuous, it suffices to show that there exists $1 > \delta > \alpha$ such that $U(Q^\delta) > U(Q)$.

Let $a := Q(x^1)$ and $b := \sum_{\{y \in D: y_0 = \delta\}} Q^\delta(y)$. Then, by the martingale property of Q and Q^δ it follows that:

$$\sum_{\{y: y_{t_1} = z_{t_1}\}} Q^\delta(y) = \sum_{\{y: y_{t_1} = z_{t_1}\}} Q(y)$$

$$\sum_{\{y: y_{t_1} = 1\}} Q^\delta(y) = \sum_{\{y: y_{t_1} = 1\}} Q(y)$$

$$a + (1 - a)z_0 = b\delta + (1 - b)z_0$$

Each path y in the support of Q or Q^δ can be identified with a vector $(y_I)_{I \in \mathcal{I}_\theta}$ (such that $y_I := y_t$ for $t \in I$). Let H denote the restriction of $\eta^\sharp$ to $\mathcal{S}_\theta$ and let h be the corresponding function as defined in Lemma D1. By (A3)

$$U(Q^\delta) = \sum_{\{y \in D\}} Q^\delta(y) \sum_{S \in \mathcal{S}_\theta} h(S) \max_{I \in S} v(u(y_I))$$

$$U(Q) = \sum_{\{y \in D\}} Q(y) \sum_{S \in \mathcal{S}_\theta} h(S) \max_{I \in S} v(u(y_I))$$

and, therefore,

$$U(Q) - U(Q^\delta) = h([0, t_1)) \left[av(1) + (1 - a)v(z_0) - (bv(\delta) + (1 - b)v(z_0))\right]$$
$$- \sum_{\{y \in D: y_{t_1} = z_{t_1}\}} Q^\delta(y) \sum_{S \in \mathcal{S}_\theta, S \neq [0, t_1)} h(S) \left(\max_{I \in S} v(u(y_I)) - \max_{I \in S} v(u([z t_1 y]_I)\right)$$

Since H is totally monotone, it follows that $h \geq 0$. Furthermore, note that

$$\max_{I \in S} v(u(y_I)) \geq \max_{I \in S} v(u([zt_1 y]_I)$$

since $y_t \in \{z_0, \delta\}$ for $t \in [0, t_1)$. Let $y := x^\delta t_1 z$ be the path that is equal to δ on $[0, t_1)$ and equal to z thereafter. Since $1 - \delta > \alpha$ it follows that $Q^\delta(y) > 0$. Then,

$$\begin{aligned}
U(Q) - U(Q^\delta) &\geq h([0, t_1)) \left[av(1) + (1-a)v(z_0) - (bv(\delta) + (1-b)v(z_0))\right] \\
&- Q^\delta(y) \sum_{S \in \mathcal{S}_\theta} h(S) \left(\max_{I \in S} v(u(y_I)) - \max_{I \in S} v(u([zt_1 y]_I) \right) \\
&\geq h([0, t_1)) \left[av(1) + (1-a)v(z_0) - (bv(\delta) + (1-b)v(z_0))\right] \\
&- Q^\delta(y) h([0, t_2)(v(u(\delta) - v(u(z_0)))
\end{aligned}$$

Since $H([0, t_2)) > H([0, t_1) + H([t_1, t_2)$ by dual strict total monotonicity, it follows that $h([0, t_2)) > 0$ and, therefore,

$$-Q^\delta(y) h([0, t_2)(v(u(\delta) - v(u(z_0))) < 0$$

Since $h([0, t_1)) \geq 0$, v is concave and $a + (1-a)z_0 = b\delta + (1-b)z_0$ it follows that

$$h([0, t_1)) \left(av(1) + (1-a)v(z_0) - (bv(\delta) + (1-b)v(z_0)\right) \leq 0$$

It follows that $U(Q) < U(Q^\delta)$ as desired. $\square$

References

Ely, J; A. Frankel and E. Kamenica, (2015), Suspense and Surprise, *Journal of Political Economy*,vol. 123, no. 1, 2015, pp. 215–260.

Epstein, L.G. and Zin, S.E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* 57, 937–69.

Gilboa, I. and D. Schmeidler, (1994), Additive representations of non-additive measures and the choquet integral, *Annals of Operations Research*, March 1994, Volume 52, Issue 1, pp 43–65

Gilboa, I, (1989), Expectation and variation in multi-period decisions, *Econometrica* 57 (1989) 1153–1169.

D. Schmeidler, (1989), Subjective probability and expected utility without additivity, *Econometrica* 571–587.

Caplin A. and J Leahy, "Psychological Expected Utility Theory and Anticipatory Feelings," The Quarterly Journal of Economics, Vol. 116, No. 1 (Feb., 2001), pp. 55–79.

Lovallo, D. and D. Kahnemann, (2000), Living with Uncertainty: Attractiveness and Resolution Timing, *Journal of Behavioral Decision Making*, 13, pp. 171–190.

Nguyen, H. T., (1978), "On random sets and belief functions," Journal of Mathematical Analysis and Applications, 65, 531–542 (1978)

Shafer, G., (1976), "A Mathematical Theory of Evidence," Princeton Univ. Press, Princeton, N.J., 1976.

Galai, D. and O. Sade, (2006). The "Ostrich Effect" and the Relationship between the Liquidity and the Yields of Financial Assets. *Journal of Business* 79 (5).

Karlsson, N., G. Loewenstein, and D. Seppi, (2009) The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty* 38.2 (2009): 95–115.