Random Evolving Lotteries and Intrinsic Preference for Information[†]

Faruk Gul

Paulo Natenzon

and

Wolfgang Pesendorfer[‡]

April 2017

Abstract

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

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

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

1. Introduction

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

In this paper, we formulate such a model of risk consumption. We use it to study preference for (non-instrumental) information and the trade off between smooth consumption paths and path risk. Our model has four parameters; a utility index u_1 that determines the decision makers attitude to current consumption risk, a utility index u_2 that determines the decision makers instantaneous risk attitude towards the terminal prize, a real valued function v that transforms instantaneous utilities, and finally, a capacity η and aggregates trajectories of transformed instantaneous utilities by identifying each such trajectory with its Choquet integral. Our main result provides a representation theorem for this model.

We provide three applications of our model. The first analyzes agents who *savor* or *dread* news that will arrive in the near future.[1] To illustrate savoring, Lovallo and Kahnemann (2000) give the example of buying lottery tickets. They point out that sellers advertise the purchase of a lottery tickets as *buying a dream.* This suggests that owners of a lottery ticket benefit if some time passes between purchase of the ticket and the lottery drawing to give them time to dream. However, dreaming about lottery winnings is likely to yield a greater benefit if the lottery drawing is near than if it is in the far distant future

[1] Lowenstein (1987) introduces the concepts of savoring and dread in the context of consumption. A person may delay consuming ice cream because she wants to savor it. Kahnemann and Lovallo (2000) extend this idea to lotteries and identify savoring with a desire to delay information about the outcome of a lottery.

so that a very long delay between ticket purchase and lottery drawing is suboptimal as well. Our model is consistent with an optimal intermediate time at which the information is revealed. More specifically, we provide conditions on the the model parameters that lead to savoring or to its counterpart, dread.

To illustrate the second application, consider the following scenario. An individual invests his retirement savings in a target date fund and never changes his allocation. Nonetheless, the individual regularly checks his account balances. On a typical day, he first checks the stock market index on his phone and then decides whether or not to log into his individual account. On days when the market is up, he is eager to check how the market increase has affected his individual balances, whereas on days when the market is down, he is less inclined to check his account. Thus, the agent is more eager to obtain additional information after good news than after bad news. Karlsson, Loewenstein and Seppi (2009) provide evidence that this type of behavior is fairly common and use the term "Ostrich effect" to describe it.[2] In our second application, we give conditions under which our model yields the Ostrich effect.

The third application is about consumption rather than information. In our model, the capacity measures time non-separability. Thus, our model makes prediction about the agent's preference over consumption streams and, in particular, about the trade-off between path smoothness and path risk. Consider two distinct stochastic consumption plans, each leading to an equal chance of high or low consumption in every period. The first has two paths, one with high consumption the other with low consumption; the second has two paths that both alternate between high and low consumption. The first random evolving lottery offer smooth paths but exposes the agent to path risk; the second hedges path risk at the expense of path smoothness. In our final application, we generalize this example and develop a notion of *preference for hedging* or *smoothing* and relate this definition to features of the capacity. Specifically, we show that agents with a totally monotone capacity choose smooth paths at the expense of hedging.

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

1.1 Related Literature

Our approach is related to Gilboa (1989) who was the first to use capacities to model time non-separability. Gilboa's variation averse preferences satisfy our definition of a preference for smoothing adapted to his setting. Our definition represents a generalization of his that can be applied to a broader class of preferences. Models of habit formation (Pollak (1970)) are designed to capture related phenomena; to relate to this literature, we give conditions under which RCU utility can be interpreted as a model with a history dependent flow utility. The advantage of the RCU representation is that its parameters can be related to the agent's preference for smoothing (or hedging) in a straightforward way.

Kreps and Porteus (1978) (henceforth KP) formulate the first model of preference for temporal of resolution of uncertainty. The 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.[3]

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

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

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

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 will be available on the afternoon of October 15. Therefore, the decision to have the enzyme test on October 8 versus October 15 has no effect on the decision maker's beliefs about her health status on October 8 or October 15.

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

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

Caplin and Leahy (2001) offer a theoretical model *anticipatory feelings.* They develop a two-period KP-style model which they call *psychological expected utility theory* (PEU). In PEU, a pair consisting of the decision maker's consumption in period 1 and uncertain consumption in period 2 is mapped into a mental state. Caplin and Leahy relate properties of this mapping to various psychological phenomena, including dynamic uncertainty. The two-period version of our model is equivalent to the corresponding two-period KP model. Moreover, our model is stated entirely in terms of uncertain distributions over consequences without any reference to mental states. Nevertheless, our model is similar to Caplin and Leahy's since we follow their lead in postulating that only the decision maker's sequence of beliefs (in each period) over physical consequences is relevant for her payoffs and not the entire path describing the resolution of uncertainty. Grant, Kaji and Polak (2000) consider preference for information in the Kreps-Porteus framework. They show that an unambiguous preference for early (or late) resolution of uncertainty is inconsistent with a number of non-expected utility theories. Similarly, we show that agents with non-separable preferences typically do not exhibit an unambiguous preference for earlier (or later) information. Dillenberger (2010) analyzes preferences over two stage lotteries that exhibit a preference for one-shot resolution of uncertainty. His main result relates violations of the independence axiom to an aversion for a gradual resolution of uncertainty. Although our model maintains independence, the relaxation of time separability allows us to capture similar behavior.[4] Dillenberger and Rozen (2015) consider a multi-period KP-style model to analyze history dependent risk aversion. While the models and objectives are different, we share the feature that past realizations affect current attitudes; in their case, attitudes to risk, in our case attitudes to information.

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

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

[4] Specifically, if v is linear and the capacity is supermodular, then our agents exhibit a preference for one-shot resolution of uncertainty.

2. Random Evolving Lotteries

Let Ω be a non-empty set. A *probability* (on Ω) is a function $\theta : \Omega \to [0,1]$ with finite support $\{\omega \in \Omega \,|\, p(\omega) > 0\}$ and such that $\sum \theta(\omega) = 1$. For $A \subset \Omega$ we let $\theta A = \sum_{\omega \in A} \theta(\omega)$ and define a sum over the null set as 0. A probability is *degenerate* if it has a single element in its support. For any real-valued function $f : \Omega \to I\!R$, we let $E_\theta[f]$ denote the expectation of f; that is, $E_\theta[f] = \sum f(\omega)\theta(\omega)$. If f 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 be a non-empty finite set of (flow) consumption levels, and let K_2 be a non-empty finite set of terminal prizes. A *lottery* is a pair $\alpha = (\alpha_1, \alpha_2)$ where α_1 is a probability on K_1 and α_2 is a probability on K_2. 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 refer to a elements of Δ_1 as *consumption lotteries* and to elements of Δ_2 as *prize lottery.* When convenient, we identify each Δ_i with the corresponding $|K_i| - 1$ dimensional simplex.

A function on the unit interval taking values in $I\!R^n$ 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 $x : [0,1] \to \Delta$ which maps each *time* t into a lottery $x(t) = (x_1(t), x_2(t)) \in \Delta$. 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 $\overline{\Pi}$ be the set of probabilities on D. For any probability $P \in \overline{\Pi}$ and subset $A \subset D$ such that $PA > 0$, let P_A be the conditional probability of P given A, that is:

$$P_A(y) = \begin{cases} \frac{P(y)}{PA} & \text{if } y \in A \\ 0 & \text{otherwise} \end{cases}$$

A probability P on D is a *random evolving lottery (REL)* if it satisfies the following *martingale property*: for any given finite sequence of lotteries $\alpha^1, \ldots, \alpha^n \in \Delta$ and times $s_1 < \ldots < s_n < t \in [0,1]$, let $A = \{x \in D \,|\, x(s_i) = \alpha^i\}$. Then, $PA > 0$ implies

$$E_{P_A}[x_2(t)] = \alpha_2^n$$

Let Π be the set of RELs. It follows from the martingale property (and the law of iterated expectations) that $E_P[x_2(t)] = E_P[x_2(0)]$.

For each lottery $\alpha \in \Delta$, let x^α denote the constant evolving lottery such that $x^\alpha(t) = \alpha$ for all $t \in [0,1]$. By the martingale property, if $P(x) = 1$ for some x, then $x_2 = x_2^\alpha$ 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.

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. For each REL P and each $t \in [0,1]$, define $P_t \in M$ as follows:

$$P_t(\alpha) = P\{x \in D \,|\, x(t) = \alpha\}$$

Hence, P_t is the t-th coordinate distribution of P. For any second-order lottery $p \in M$, 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.

Let $\succeq$ be a binary relation on Π; that is, a subset of $\Pi \times \Pi$. We say that $\succeq$ is *degenerate* if $R^\alpha \sim R^\beta$ whenever $\alpha_1 = \beta_1$ or if $R^\alpha \sim R^\beta$ whenever $\alpha_2 = \beta_2$. 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 usual mixture of probabilities. Clearly, with this operation Π is a mixture space. We impose the independence axiom on this mixture space:

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

We endow Π with the Prohorov metric.[5] 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 $Q \in \Pi$.*

[5] 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.*

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 $R^{a\alpha+(1-a)\gamma} \succ R^{a\beta+(1-a)\gamma}$.*

We obtain an induced preference on M by restricting $\succeq$ to $\{R^p \in \Pi \,|\, p \in M\}$. This induced preference over second-order lotteries satisfies independence by Axiom 2.

For $P, Q \in \overline{\Pi}$, we say that P *dominates* Q if $R^{P_t} \succeq R^{Q_t}$ for all t. In other words, P dominates Q whenever the t-th coordinate distribution of P is preferred to the t-th coordinate distribution of Q for every t. P *strictly dominates* Q if P dominates Q and Q does not dominate P. The following axiom implies separability across time intervals:

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

The goal of our paper is to capture phenomena, such as the Ostrich effect, that are inconsistent with Axiom 5*. Evidence documented by Karlsson, Loewenstein and Seppi (2009) suggests that the investor is more eager to learn the current portfolio value if the stock market has been increasing in value than if it has been decreasing. This suggests that even if two RELs P and Q have identical coordinate distributions at each date t, one may be more attractive than the other if the former reveals more information following good news than the latter.[6] Our weakening of Axiom 5* allows for this non-indifference but maintains dominance under a more stringent condition.

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

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

be the $\iota-$paths. We say that P *rank-dominates* Q if $QA_\iota = PA_\iota$ for all ι and $PA_\iota > 0$ implies and P_{A_ι} dominates Q_{A_ι}. Thus, we require that P and Q assign the same probability to $\iota-$paths and that P's marginal dominates Q's marginal for such paths. 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 $P \succ Q$.*

[6] We construct RELs with this feature in section 3.3 below.

Our utility representation has three parameters; a von Neumann-Morgenstern utility $u : \Delta \to [0,1]$; a second-stage index $v : [0,1] \to [0,1]$ that measures the agent's attitude to two stage lotteries; and a capacity η to aggregate the instantaneous utility flows along a path.

The continuous function $u : \Delta \to [0,1]$ is a *utility* if it is onto and separable; that is, if there exist $u_1 : \Delta_1 \to [0,1]$ and $u_2 : \Delta_2 \to [0,1]$ such that $u(\alpha) = u_1(\alpha_1) + u_2(\alpha_2)$ for all $\alpha \in \Delta$. A utility is linear if $u(a\alpha + (1-a)\beta) = au(\alpha) + (1-a)u(\beta)$. Let Λ be the set of all continuous, strictly increasing functions from $[0,1]$ onto itself. A second stage index is a function $v \in \Lambda$.

To define the capacity η, we first describe the appropriate $\sigma-$algebra. We call $A \subset [0,1]$ an interval if $A = [s,t)$ for $1 > t > s \geq 0$ or if $A = [s,1]$ for $1 > s \geq 0$. Let $\mathcal{S}$ be the set of subsets of the unit interval that can be expressed as the finite union intervals. Let l denote Lebesgue measure. Then, a function $\eta : \mathcal{S} \to [0,1]$ is a continuous capacity if

(1) $\eta\emptyset = 0$, $\eta[0,1] = 1$; $\eta S \leq \eta T$ if $S \subset T$;

(2) $[S \subset S_k$ for all k and $l\,[\bigcap_{k=1}^{\infty} S_k] = lS]$ implies $\lim_n \eta \bigcap_{k=1}^{n} S_k = \eta S$.

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

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

A function V represents $\succeq$ if $P \succeq Q$ if and only if $V(P) \succeq V(Q)$. Such a function is a *risk consumption utility* (RCU) if there is a linear utility u, a second-stage index v, and a continuous capacity η such that

$$V(P) = E_P\left[\int v(u(x(t)))\, d\eta\right]$$

for all P. If $\succeq$ can be represented by an RCU, we call it a *risk consumption preference* (RCP). If the V above represents $\succeq$, we identify it with both (u, v, η) and $\succeq$.

Theorem 1: *A non-degenerate $\succeq$ satisfies Axioms 1–5 if and only if it is a risk consumption preference. Moreover, its RCU representation is unique.*

An RCU is a linear function on Π and the instantaneous utility u is a linear function on Δ. But the utility of a path need not be separable across time periods. If we replace Axiom 5 with Axiom 5* then η is an additive measure:

Corollary 1: *The RCP (u, v, η) satisfies Axiom 5* if and only if η is an additive measure.*

We say that a risk consumption preference is *separable* if it satisfies Axiom 5*. In that case, the utility of each path is a linear function of the flow of (instantaneous) utilities. We use the acronym SRCP (SRCU) for the separable risk consumption preferences (utilities) of Corollary 1. As we show in the next two sections, the non-separability of RCPs enables us to model more nuanced attitudes to information such as savoring, dread, and the ostrich effect that cannot be captured by SRCPs.

3. Preference for Information

Next, we consider three notions of preference for information. To focus on information, we assume in this section that the consumption lottery is fixed and constant over time. Thus, we assume that every REL is an element of the set

$$\Pi_c := \{P \in \Pi \mid P(x) \cdot P(y) > 0 \text{ implies } x_1(t) = y_1(s) \text{ for all } s, t\}$$

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

$$\varepsilon(x)(t) = \begin{cases} (x_1(t), x_2(t+\varepsilon)) & \text{for } t \in [0, 1-\varepsilon] \\ (x_1(t), x_2(1)) & \text{otherwise} \end{cases}$$

Then, define $\varepsilon[P] \in \Pi$ such that $\varepsilon[P]A = P\varepsilon(A)$ for all $A \subset D$, where $\varepsilon(A) = \{\varepsilon(x) \mid x \in A\}$. We say that $\succeq$ is *information seeking (information averse)* if $\varepsilon[P] \succeq P$ ($P \succeq \varepsilon[P]$) for all P. If $\varepsilon = 1$, then $\varepsilon[P]$ reveals all the information of P at date 0. Thus, a weaker notion of preference for information is a preference for *immediate disclosure*: $\succeq$ prefers immediate disclosure if $1[P] \succeq P$ for all P. Finally, define $\bar{x}(P)$ to be the expected path of the REL

P. That is, $\bar{x}(t)(P) = \sum_{x \in D} x_t P(x)$ for all t. The REL $R^{\bar{x}(P)}$ reveals no information about the prize lottery. We say that the preference is *averse to (prefers) no disclosure* if $P \succeq R^{\bar{x}(P)}$ ($R^{\bar{x}(P)} \succeq P$) for all P.

For a separable risk consumption utility, that is, if Axiom 5* holds, the three notions are equivalent:

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

A symmetric counterpart of Theorem 2 also holds: the counterpart can be derived from Theorem 2 by replacing prefers with averse to, averse to with prefers, information seeking with information averse and convex with concave in the above statement. Thus, there are no SRCUs that prefer information either very quickly or not at all; similarly, there is no SRCU that prefers a gradual resolution of uncertainty over both no disclosure and immediate resolution. In other words, SRCU preferences are not rich enough to analyze applications that go beyond a categorical preference for information.

3.1 Preference for Information, Savoring and Dread

As the following example illustrates, non-seperable risk consumption utilities allow for more nuanced attitudes to information. For two paths $x, y \in D$ we write xty for the path z such that $z(s) = x(s)$ for $s < t$ and $z(s) = y(s)$ for $s \geq t$. Let α, β be two lotteries that yield the same immediate consumption but differ in the prize lotteries, let $\gamma = \alpha/2 + \beta/2$ and let $Q^t \in \Pi_c$ be the REL that has two equiprobable paths and reveals the uncertainty about the prize lottery (α or β) at time t:

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

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

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

Theorem 3, below, gives necessary and sufficient conditions for an RCU agent to prefer (or be averse to) no disclosure and sufficient conditions for a preference for immediate disclosure. The example above shows that, unlike in the SCRU case, this characterization does not extend to information seeking.

For the capacity μ we define its *dual*, $\mu^\sharp$, such that $\mu^\sharp(S) = 1 - \mu([0,1] \backslash S)$.

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

Theorem 3: *Let $\succeq = (u, v, \eta)$ be an RCU. Then,*

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

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

Theorem 3 is analogous to Theorem 2, above and, like Theorem 2, has a symmetric counterpart that can be obtained from Theorem 3 by switching the places of averse to with prefers, convex with concave and submodular with supermodular throughout the statement above.

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

Below, we take advantage of this conflict between preference for immediate disclosure and aversion to no disclosure to develop a model of *savoring* and *dread*. Loewenstein

(1987) introduced these notions to facilitate his analysis of anticipatory utility in dynamic consumption. Lovallo and Kahnemann (2000) extend these notions to the analysis of resolution of uncertainty. The latter authors examine subjects' willingness to live with uncertainty as a function of the attractiveness of the gamble confronting them. Lovallo and Kahnemann find that subjects are willing to delay the resolution of uncertainty for sufficiently attractive gambles and interpret this behavior as a savoring positive future outcomes. Below, we take advantage of RCUs more nuanced attitude towards information to formulate a definition of savoring distinct from aversion to information.

For $t \in (0,1)$, let Π_c^t be the set of all RELs with constant immediate consumption that provide significant information at time t and only at time t. Thus, $P \in \Pi_c^t$ has paths of the form $x^\alpha t x^\beta$ and the information must be significant; that is, there is β, β' such that $R^\beta \succ R^{\beta'}$ and $P(x^\alpha t x^\beta) \cdot P(x^\alpha t x^{\beta'}) > 0$.

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

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

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

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

Theorem 4, below, shows that a convex v together with a sufficiently submodular η can make the decision maker savor any $P \in \Pi_c^t$. The requirement that η is sufficiently submodular is needed to overcome the effect of the convexity of v which tends to make immediate disclosure more attractive. Similarly, a concave v and sufficiently supermodular η yield dread.

Theorem 4: *Assume $P \in \Pi_c^t$ and v is convex. Then, there is η^* such that (u, v, η) savors P whenever η is more submodular than η^*. Conversely, if v is concave, then there is η^* such that (u, v, η) dreads P whenever η is more supermodular than η^*.*

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

3.2 The Ostrich Effect

In this section, we define what it means for one REL to have more "news after good news" than another REL and use this definition to relate the ostrich effect to the RCU parameters. To illustrate the effect, consider constant consumption RELs with two terminal prizes, a good prize and a bad prize. For each path x, $x(t) \in [0, 1]$ is the probability of receiving the good prize. The REL Q, defined below, has four equiprobable paths such that each path is constant on $[0, t_1), [t_1, t_2)$ and on $[t_2, 1]$. To simplify notation, we describe each path as a vector of lotteries (α, β, γ) with the interpretation that α is the lottery on $[0, t_1)$, β is the lottery on $[t_1, t_2)$ and γ is the lottery on $[t_2, 1]$. The following matrix describes the REL Q:

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

The matrix does not specify the times t_1, t_2 when information is revealed because these times play no role in the discussion below. The REL Q captures the following situation: at time 0 the agent learns whether she has a .6 or a .2 chance of winning the good prize. At time $t > 0$ she receives additional information: she learns whether her probability is .8 or .4 (if she started at .6) or whether her probability is .4 or 0 (if her starting point was .2).

Note that there are two distinct paths that lead to a probability of .4: along the second path, the agent arrives at .4 following bad news at time t while along the third path, the agent arrives at .4 following good news at time t.

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

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

Experiments by Karlsson, Loewenstein and Seppi (2009) suggest that some decision makers prefer Q^g to Q^b. Our objective is to show that this ranking is compatible with RCU utility and relate it to the parameters of RCU. Let $Q \in \Pi(x, y)$ if there exist $\alpha \succ_0 \beta$ and $0 < t, a < 1$ such that:

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

(ii) $x(s) = y(s) = a\alpha + (1-a)\beta$ if $s \in [t, 1]$.

(iii) $x(s) \succeq_0 \alpha \succ_0 \beta \succ_0 y(s)$ for $s < t$

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

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

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

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

[9] We assume that x is uniformly above α and y is uniformly below β. It would be sufficient to require, for all $s < t$, that $x(s)$ is uniformly above $y(s)$ and that neither $x(s)$ nor $y(s)$ are "between" α and β.

y. That is,

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

The REL Q^b reveals the same information along the path x. That is,

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

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

The capacity η is totally monotone if, for all $k \geq 1$ and all families of sets $\{S_1, \dots, S_k\}$ such that $S_i \in \mathcal{S}$,

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

The capacity $\eta^\#$ is the dual of η if $\eta^\# S = 1 - \eta([0,1]\backslash S)$ for all S and η is dual-totally monotone if $\eta^\#$ is totally monotone.[10]

Theorem 5: *An RCU (u, v, η) displays the ostrich effect if v is convex and η is dual totally monotone or if v is concave and η is totally monotone.*

Theorem 5 provides conditions under which RCU agents will exhibit the ostrich effect noted in Karlsson, Loewenstein and Seppi (2009). It provides a tighter control of the circumstances under which the effect is observed than Karlsson et al. In particular, Q^g and Q^b provide exactly the same information; they only differ in the history preceding the information. Also, the information is small relative to the good news or bad news that

[10] If η were a probability measure, the inequalities above would be equalities. To see this, note that the inclusion-exclusion principle applied to the family $(S_1, \dots, S_k)$ implies that the right-hand side of the above inequality is simply the probability of $S = \cup_{i=1}^k S_i$. 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 inclusion-exclusion principle to hold. Dempster (1967) and Nguyen (1978) analyze the properties of totally monotone capacities.

precedes it. This last constraint follows from the requirement that x is above α and y is below β prior to t.

If v is concave and η is totally monotone, the agent prefers no disclosure (Theorem 5) and, therefore, is averse to information. In that case, additional information at time τ 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, 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 5 also implies that the agent displays the ostrich effect when v is linear and $\eta = b\eta^1 + (1-b)\eta^2$ for some totally monotone η^1 and dual totally monotone η^2 and $b \in [0,1]$.

4. Hedging versus Smoothing

In this section, we analyze the agent's preference over consumption paths. Specifically, we introduce two criteria; the first specifies what it means for an agent to prefer smooth consumption paths and the second what it means for an agent to be averse to path risk. In Theorem 6, below, we relate those criteria to properties of the capacity.

To simplify the exposition of our motivating examples and some of the definitions below, we will adopt the following conventions and notation. The examples below will have equiprobable paths with a constant and identical prize lottery. Thus, paths differ only with respect to the consumption lottery. Furthermore, we will assume the consumption lottery yields one of two prizes; prize 2, the better prize, and prize 1, the worse prize. Let Π^* be the set of all RELs with these properties for some fixed prizes terminal prize lottery. We can write every REL in Π^* as a matrix:

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

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

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

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

Here X is a $m \times k$ matrix, Y is $m \times k'$, X^* is $m' \times k$ and Z is $m' \times k'$. The REL R has $m + m'$ paths. Each column represents the consumption lotteries of the $m + m'-$paths for a set of states in which all paths are constant. We refer to the matrices X, X^*, Y, Z as *fragments*; let $0_m, 1_m$ denote $m-$dimensional vector of zeros and ones respectively and let O be the zero fragment.

Our definition of a preference for hedging (and smoothing) is in terms of RELs in Π^* with the additional property that each path yields zero or one at every $t \in [0, 1]$. Thus, each path yields one of two possible deterministic consumptions at each time. Recall that 1 denotes the more desirable consumption. Let

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

Clearly, X *is smoother than* Y while Y *has less path risk (provides a better hedge) than* X. Next, let $S_1^* = S_1 \cup S_2$ and consider the following two fragments:

$$Y^* = \begin{pmatrix} S_1^* & S_3 \\ Y & 1_2 \\ X & 0_2 \end{pmatrix} \quad \text{and} \quad X^* = \begin{pmatrix} S_1^* & S_3 \\ Y & 0_2 \\ X & 1_2 \end{pmatrix}$$

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

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

Below, we generalize this idea to provide an inductive definition of "less risky" and "smoother." We say that Y (as defined above) is 1-less risky than X and write $Y \succeq_h^1 X$. Similarly, we say that X is 1-smoother than Y and write $X \succeq_s^1 Y$. For $k > 1$, we say W is k-less risky than Z, ($W \succeq_h^k Z$), if

$$W = \begin{pmatrix} W' & 1_m \\ Z' & 0_m \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} W' & 0_m \\ Z' & 1_m \end{pmatrix}$$

where and W', Z' are $n \times m$ matrices such that $Z' \succeq_h^{k-1} W'$. Replacing $Z' \succeq_h^{k-1} W'$ with $W' \succeq_s^{k-1} Z'$, we obtain an analogous definition of "k-smoother." These two constructions facilitate the definition below. We write $P = (Z, O)$ for an REL characterized by fragment Z on $S_1 \cup \ldots \cup S_m$ and and the $0-$matrix on $S_{m+1} \cup \ldots \cup S_n$.

Definition: *An RCU decision maker has a preference for hedging (smoothing) if she prefers $P = (X', O)$ to $Q = (Y', O)$ whenever $X' \succeq_h^k Y'$ ($X' \succeq_s^k Y'$) for some $k \geq 1$.*

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

The definitions of preference for hedging and preference for smoothing above consider only a small subset of RELs and hence are weak. Below, we provide alternative, stronger conditions. Theorem 5 establishes that for RCUs, the two characterizations are equivalent. The alternative conditions are easier to state and to apply to all RELs.

For any set $S \in \mathcal{S}$, let

$$A^{S\alpha} = \{x \,|\, R^{x(s)} \succeq R^\alpha \text{ for all } s \in S\}$$

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

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

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

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

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 $\alpha \in \Delta$.*

To illustrate these definitions, consider the following examples:

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

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

Note that Q is smoother than P while P has less path risk. Theorem 6, below, establishes that this relationship between preference for smoothing and upper domination is general and shows that, in an RCU, a preference for upper domination corresponds to the *total monotonicity* of the capacity while a preference for lower-domination corresponds to its *dual total monotonicity.*

Theorem 6: *Let $\succeq = (u, v, \eta)$ be a RCU preference. Then, the following three statements are equivalent:*

(i) $\succeq$ has a preference for smoothing (hedging)

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

(iii) η is totally monotone (dual-totally monotone).

Next, we provide some intuition regarding the equivalence of (ii) and (iii) in Theorem 6. Let $\mathcal{T}$ denote a subalgebra of $\mathcal{S}$. A totally monotone capacity η can be identified with a collection of probabilities $h_{\mathcal{T}} : \mathcal{T} \to [0,1]$, one for each finite subalgebra $\mathcal{T}$ of $\mathcal{S}$, such that for all $S \in \mathcal{T}$,

$$\eta S = \sum_{\substack{T \subset S \\ T \in \mathcal{T}}} h_{\mathcal{T}}(T) \tag{1}$$

The probabilities $h_\mathcal{T}$ yield the following characterization of an RCU path utility: let $\mathcal{T}$ be the (set inclusion) algebra generated by the partition $\{S_1, \ldots, S_n\}$ and let $x = (\alpha_1, \ldots, \alpha_n)$ be a path such that $x(t) = \alpha_i$ whenever $t \in S_i$. Hence, the corresponding utility path is $\phi = (u(\alpha_1), \ldots, u(\alpha_3))$. Then, the utility of this path for an agent with the totally monotone capacity η is

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

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

Theorem 5 shows that an RCU agent with a totally monotone capacity prefers smooth utility over time. This RCU agent shares some similarities with agents whose utility index depends on the consumption history, as in models of habit formation (Pollak (1970)). To illustrate the connection between the models, we show that Choquet path utilities with a totally monotone (or a dual totally monotone) capacity can be represented as a history dependent path utility.

5. Relation to habit models

In this paper, we have introduced a class of preferences that are flexible enough to capture nuanced attitudes to (non-instrumental) information and to the trade-off between smooth consumption paths and path risk. The novel feature of our preferences is a time-nonseparability measured by a capacity. Models of habit formation (Pollak (1970)) also feature time non-separability but this non-separability is modeled through a history dependent utility index. In this section, we give conditions under which our model can be restated as an example of a habit model.

The RCU agent (u, v, η) maps a path, $x \in D$, to a utility path, $\phi : [0,1] \to [0,1]$, such that $\phi(t) = v(u(x(t))$ and aggregates the utility path ϕ using the capacity η. Let Φ

represents the set of utility paths, that is, right continuous functions from the unit interval to the unit interval. A *path utility* $w : \Phi \to [0,1]$ assigns a value to each utility path. The RCU path utility is represented by the Choquet integral, that is, $w(\phi) = \int \phi d\eta$

In a habit model, the flow utility at time t depends on consumption at time t and on the *consumption history*. The function $V : [0,1] \times \Phi \to [0,1]$ is a *history dependent utility* if it is right continuous in t for all ϕ and if, for all $t \in [0,1]$,

$$\begin{aligned} V_t(\phi) &\geq V_t(\phi') \text{ if } \phi \geq \phi'; \\ V_t(\phi) &= V_t(\phi') \text{ if } \phi(s) = \phi'(s) \text{ for } s \leq t. \end{aligned} \tag{3}$$

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

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

(where the integral above is the standard Riemann integral). In Appendix E, we show that all Choquet path utilities with a totally monotone or a dual totally monotone capacity have a habit representation.[11] Here, we illustrate this fact with a simple example.

Consider the Choquet path utility with parameters (v, η) such that $\eta(S) = [l(S)]^2$ and let w be the corresponding path utility, that is, $w(\phi) = \int v(\phi(t)) d\eta$. Note that η is totally monotone. Define the path ϕ^t such that $\phi^t(s) = \min\{\phi(t), \phi(s)\}$ and define the history dependent utility V^η such that

$$V_t^\eta(\phi) = 2 \int_0^t v(\phi^t(s)) ds$$

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

$$w(\phi) = \int_0^1 V_t^\eta(\phi) dt$$

[11] More generally, any RCU utility with a capacity $\eta = a\mu^1 + (1-a)\mu^2$ where $a \in [0,1]$, μ^1 totally monotone and μ^2 dual totally monotone has a habit representation.

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

Next, consider the RCU path utility with parameters (v, μ) such that $\mu(S) = 2l(S) - [l(S)]^2$ and let w' be the corresponding path utility, that is, $w'(\phi) = \int v(\phi(t))d\mu$. Note that μ is the dual of η and is, therefore, dual totally monotone. Define the path $\bar{\phi}^t$ such that $\bar{\phi}^t(s) = \max\{\phi(t), \phi(s)\}$ and define the history dependent utility function V^μ such that

$$V_t^\mu(\phi) = 2\int_0^t v(\bar{\phi}^t(s))ds$$

Again, we obtain that

$$w'(\phi) = \int_0^1 V_t^\mu(\phi)dt$$

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

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

6. 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 $u : \Delta \to [0,1]$, $u_1 : \Delta_1 \to [0,1]$ and $u_2 : \Delta_2 \to [0,1]$ such that (i) $R^\alpha \succeq R^\beta$ if and only if $u(\alpha) \geq u(\beta)$, (ii) $u(\alpha) = u_1(\alpha_1) + u_2(\alpha_2)$, 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 $d_p(R^\alpha, R^\beta) = ||\alpha - \beta|| = d(x_\alpha, x_\beta)$, Axiom 3 implies that $\succeq_*$ is continuous. 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 $\overline{\alpha} \in \arg\max_\Delta u(\cdot)$ and $\underline{\alpha} \in \arg\min_\Delta u(\cdot)$ such that $u(\overline{\alpha}) = 1$ and $u(\underline{\alpha}) = 0$. For any $\alpha, \beta \in \Delta$, the linearity of u implies

$$\frac{1}{2}u(\alpha) + \frac{1}{2}u(\beta) = u\left(\frac{1}{2}\alpha_1 + \frac{1}{2}\beta_1, \frac{1}{2}\alpha_2 + \frac{1}{2}\beta_2\right) = \frac{1}{2}u(\alpha_1, \beta_2) + \frac{1}{2}u(\beta_1, \alpha_2).$$

Hence,

$$u(\alpha) + u(\beta) = u(\alpha_1, \beta_2) + u(\beta_1, \alpha_2). \qquad (A1)$$

Then, let $u_1(\alpha_1) = u(\alpha_1, \underline{\alpha}_2)$ and let $u_2(\alpha_2) = u(\underline{\alpha}_1, \alpha_2)$. Equation (A1) implies $u(\alpha) = u_1(\alpha_1) + u_2(\alpha_2)$ as desired. $\square$

Since $P \in \Pi$ has finite support, there exists a finite set of ordered partitions I such that $PA_\iota > 0$ if and only if $\iota \in I$. Define $b_\iota = PA_\iota$ and $P_\iota = P_{A_\iota}$. Then, $P = \sum_{\iota \in I} b_\iota P_\iota$. We refer to $(I, (b_\iota, P_\iota)_{\iota \in I})$ as the decomposition of P. We will write $P = (I, (b_\iota, P_\iota)_{\iota \in I})$ if $(I, (b_\iota, P_\iota)_{\iota \in I})$ is a decomposition of P. Note that every REL has a unique decomposition. We can restate rank domination in terms of the decompositions of RELs: $P = (I, (b_\iota, P_\iota)_{\iota \in I})$ rank dominates $Q = (I', (b'_\iota, Q_\iota)_{\iota \in I'})$ if $I = I'$ and, for all $\iota \in I$, $b_\iota = b'_\iota$ and P_ι dominates Q_ι.

Lemma 2: *Let $P = (I, (b_\iota, P_\iota)_{\iota \in I})$, $Q = (I', (b'_\iota, Q_\iota)_{\iota \in I})$, $a, b \in [0,1]$, and let $\alpha, \beta \in \Delta$. If $I = I'$ and, for all $\iota \in I$, $b_\iota = b'_\iota$ and $aP_\iota + (1-a)R^\alpha$ dominates $bQ_\iota + (1-b)R^\beta$, then $aP + (1-a)R^\alpha \succeq bQ + (1-b)R^\beta$.*

Proof: For each $\iota = (S_1^\iota, \ldots, S_k^\iota) \in I$, define the evolving lotteries x^ι, y^ι by

$$x^\iota(t) = (2^{-j}\overline{\alpha}_1 + (1-2^{-j})\underline{\alpha}_1, \overline{\alpha}_2) \text{ if } t \in S_j^\iota$$

$$y^\iota(t) = (2^{-j}\overline{\alpha}_1 + (1-2^{-j})\underline{\alpha}_1, \underline{\alpha}_2) \text{ if } t \in S_j^\iota$$

For $n \geq 1$, let

$$w^{\iota n} = 2^{-n}x^\iota + (1-2^{-n})x^\alpha$$

$$z^{\iota n} = 2^{-n}x^\iota + (1-2^{-n})x^\beta$$

Note that $x^\iota, y^\iota, w^{\iota n}, z^{\iota n} \in A_\iota$.

Since $aP_\iota + (1-a)R^\alpha$ dominates $bQ_\iota + (1-b)R^\beta$ for each $\iota \in I$, $aP + (1-a)R^\alpha$ dominates $bQ + (1-b)R^\beta$. Hence, for each $n \geq 1$, $n^{-1}\sum_{\iota \in I} b_\iota R^{x^\iota} + (1-n^{-1})[aP + (1-a)R^\alpha]$ strictly dominates $n^{-1}\sum_{\iota \in I} b_\iota R^{y^\iota} + (1-n^{-1})[bQ + (1-b)R^\beta]$. Thus for each fixed $n \geq 1$ there exists an integer $M(n)$ such that $n^{-1}\sum_{\iota \in I} b_\iota R^{x^\iota} + (1-n^{-1})[aP + (1-a)\sum_{\iota \in I} b_\iota R^{w^{\iota k}}]$ strictly rank-dominates $n^{-1}\sum_{\iota \in I} b_\iota R^{y^\iota} + (1-n^{-1})[bQ + (1-b)b_\iota R^{z^{\iota k}}]$, for every $k \geq M(n)$. Let $P^n = n^{-1}\sum_{\iota \in I} b_\iota R^{x^\iota} + (1-n^{-1})[aP + (1-a)\sum_{\iota \in I} b_\iota R^{w^{\iota M(n)}}]$ and $Q^n = n^{-1}\sum_{\iota \in I} b_\iota R^{y^\iota} + (1-n^{-1})[bQ + (1-b)b_\iota R^{z^{\iota M(n)}}]$ for all $n \geq 1$. Then P^n strictly rank-dominates Q^n for all n, and $P^n \succ Q^n$ by Axiom 5. Moreover, $d_p(P^n, aP + (1-a)R^\alpha) \to 0$ and $d_p(Q^n, bP + (1-b)R^\beta) \to 0$, and Axiom 3 implies $aP + (1-a)R^\alpha \succeq bQ + (1-b)R^\beta$. □

Lemma 3: *There is a continuous, linear and onto function* $V : \Pi \to [0,1]$ *that represents* $\succeq$ *such that* $V(R^{\overline{\alpha}}) \geq V(P) \geq V(R^{\underline{\alpha}})$.

Proof: The set Π is a mixture space under the usual mixture operation. 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. Lemma 2 implies that $R^{\overline{\alpha}} \succeq P$ and $P \succeq R^{\underline{\alpha}}$. It follows that the range of $\hat{V}$ is a compact interval. Axiom 5 implies that $R^{\overline{\alpha}} \succ R^{\underline{\alpha}}$. 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 3 ensure that v is a well-defined element of Λ. We call P an $\iota-$REL if $PA_\iota = 1$. Let Π_ι be the set of all $\iota-$RELs. Let Φ denote the set of all step-functions $\phi : [0,1] \to [0,1]$. Define $f : \bar{\Pi} \to \Phi$ as follows:

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

For any ordered partition ι we write $|\iota|$ for the cardinality of the partition and $\iota = (S_1^\iota, \ldots, S_{|\iota|}^\iota)$ for the sets. Let

$$\Phi_\iota = \{\phi \in \Phi : \phi(t) > \phi(s) \text{ if and only if } t \in S_i^\iota, s \in S_j^\iota \text{ such that } i < j\}$$

Let $\hat{\Phi}_\iota = \{f(P) \,|\, P \in \Pi_\iota\}$. Note that Lemma 2 ensures that $f(P) = f(Q)$ for $P, Q \in \Pi_\iota$ implies $V(P) = V(Q)$. The linearity of V ensures that $az + (1-a)z' \in \Phi_\iota$ whenever $z, z' \in \Phi_\iota$.

Let $v(u(\overline{\alpha}_1, \underline{\alpha}_2)) = \bar{r}$ and note that $\bar{r} > 0$ by non-degeneracy. Fix any ordered partition $\iota = (S_1^\iota, \ldots, S_{|\iota|}^\iota)$. Define the $|\iota| \times |\iota|$ matrix A as follows: $a_{ij} = (|\iota| - j)\bar{r}/(|\iota| + 1)$ if $j > i$ and $a_{ij} = (|\iota| - j + 1)\bar{r}/(|\iota| + 1)$ if $j \leq i$. By invoking elementary properties of systems of linear equations, we can verify that A has a non-zero determinant. By Lemma 1, for each i and j there exists $\alpha^{ij} \in \Delta$ such that $v(u(\alpha^{ij})) = a_{ij}$ and $\alpha_2^{ij} = \underline{\alpha}_2$. Then, define the evolving lotteries $x^1, \ldots, x^{|\iota|}$ as follows: $x^i(t) = \alpha^{ij}$ whenever $t \in S_j$. Let $P^i(x^i) = 1$ for each i. Consider the following system of linear equations:

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

where $\mathbf{v}$ is a column vector such that $\mathbf{v}_i = V(P^i)$. Let $\eta_\iota(S_j) = \mathbf{y}_j$ where $\mathbf{y}$ is the solution to the system of equations (A1). Identify each $\phi \in \Phi_\iota$ with the appropriate $|\iota|$-vector $(\phi_1, \ldots, \phi_{|\iota|})$. We say that $Q \in \Pi_\iota$ is *normal* if $f(Q) = \phi$ and

$$\sum_{j=1}^{|\iota|} \phi_j \eta_\iota(S_j^\iota) = V(Q) \tag{A2}$$

Hence, each P^i is normal.

Lemma 4: *Every $P \in \Pi_\iota$ is normal.*

Proof: For $f(P) = \phi \in \Phi_\iota$, there exist $r_i \in I\!R$ for $i = 1, \ldots, |\iota|$ such that $\phi = \sum_{i=1}^{|\iota|} r_i a_i$. Let $r_{|\iota|+1} = 1 - \sum_{i=1}^{|\iota|} r_i$ and define $a_{|\iota|+1}$ such that $a_{|\iota|+1,j} = 0$ for all j. Let the corresponding path $x^{|\iota|+1}$ be $x^{|\iota|+1}(t) = \underline{\alpha}$ for all t. Define $r_{|\iota|+1}^+ := \max\{r_{|\iota|+1}, 0\}$, and $r_{|\iota|+1}^- := \max\{-r_{|\iota|+1}, 0\}$.

We rearrange the equation $\phi = \sum_{i=1}^{|\iota|+1} r_i a_i$, by moving all terms with $r_i < 0$ to the left-hand side and divide the resulting equation by the sum of the coefficients on the left-hand side. Specifically, let $N^- := \{i \in \{1, \ldots, |\iota|\} : r_i < 0\}$, let $N^+ := \{i \in \{1, \ldots, |\iota|\} : r_i \geq 0\}$, and let

$$b = r^+_{|\iota|+1} + \sum_{i \in N^+} r_i = 1 + r^-_{|\iota|+1} + \sum_{i \in N^-} |r_i|.$$

Then,

$$\frac{1}{b}\left(\phi + \sum_{i \in N^-} |r_i| a_i + r^-_{|\iota|+1} a_{|\iota|+1}\right) = \frac{1}{b}\left(\sum_{i \in N^+} r_i a_i + r^+_{|\iota|+1} a_{|\iota|+1}\right)$$

Define

$$\hat{Q} = \frac{1}{b - r^-_{|\iota|+1}} P + \sum_{i \in N^-} \frac{|r_i|}{b - r^-_{|\iota|+1}} P^i$$

$$\hat{R} = \sum_{i \in N^+} \frac{r_i}{b - r^+_{|\iota|+1}} P^i$$

Then, let

$$Q = \frac{b - r^-_{|\iota|+1}}{b} \hat{Q} + \frac{r^-_{|\iota|+1}}{b} R^{\underline{\alpha}}$$

$$R = \frac{b - r^+_{|\iota|+1}}{b} \hat{R} + \frac{r^+_{|\iota|+1}}{b} R^{\underline{\alpha}}$$

Note that $\hat{R}, \hat{Q} \in \Pi_\iota$, and $f(R) = f(Q)$. Lemma 2 then implies that $V(R) = V(Q)$. Using the linearity of V we can rearrange the terms again to get

$$V(P) = \sum_i r_i V(P^i) = \sum_i r_i \sum_{j=1}^{|\iota|} \alpha^{ij} \eta_\iota(S_j) = \sum_{j=1}^{|\iota|} \phi_j \eta_\iota(S_j)$$

as desired. □

Let $P = (I, (b_\iota, P_\iota)_{\iota \in I})$ and let $\phi^\iota := f(P_\iota)$. We say that P is normal if

$$V(P) = \sum_{\iota \in I} b_\iota \sum_{j=1}^{|\iota|} \phi^\iota_j \eta_\iota(S^\iota_j)$$

Lemma 5: *Every $P \in \Pi$ is normal.*

Proof: Let $P = (I, (b_\iota, P_\iota)_{\iota \in I})$ and define $Q = \bar{r}P + (1 - \bar{r})R^{\underline{\alpha}}$. Then, $V(Q) = \bar{r}V(P)$ by the linearity of V. Let $Q_\iota = \bar{r}P_\iota + (1 - \bar{r})R^{\underline{\alpha}}$ for each $\iota \in I$. Since $f(Q_\iota)(t) \in [0, \bar{r}]$

it follows that $f(Q_\iota) \in \hat{\Phi}$. Let $\hat{Q}_\iota \in \Pi_\iota$ such that $f(\hat{Q}_\iota) = f(Q_\iota)$. Then, by Lemma 2, $V(\sum_{\iota \in I} b_\iota \hat{Q}_\iota) = V(Q)$. Linearity and Lemma 4 then imply that Q is normal. Moreover,

$$\begin{aligned} V(Q) &= \sum_{\iota \in I} b_\iota \sum_{j=1}^{|\iota|} f_j(Q_\iota)\eta_\iota(S_j^\iota) \\ &= \sum_{\iota \in I} b_\iota \sum_{j=1}^{|\iota|} f_j(\bar{r}P_\iota + (1-\bar{r})R^{\underline{\alpha}})\eta_\iota(S_j^\iota) \\ &= \bar{r} \sum_{\iota \in I} b_\iota \sum_{j=1}^{|\iota|} f_j(P_\iota)\eta_\iota(S_j^\iota) \end{aligned}$$

Since $V(Q) = \bar{r}V(P)$ it follows that P is normal. □

For all $S \in \mathcal{S}$, let $\eta(S) = \eta_{\iota^*}(S)$ for $\iota^* = (S, [0,1]\backslash S)$ and set $\eta(\emptyset) = 0$.

Lemma 6: *(i) $S = \bigcup_{j \le i} S_j^\iota$ implies $\eta(S) = \sum_{j=1}^{i} \eta_\iota(S_j^\iota)$; (ii) η is a continuous capacity.*

Proof: (i) Let $\iota = (S_1^\iota, \ldots, S_n^\iota)$, let $S = \bigcup_{j \le i} S_j^\iota$, let $\iota^* = (S, [0,1]\backslash S)$, and let $P \in \Pi_\iota$ be degenerate with $P(y) = 1$ for some $y \in A_\iota$. Define $x^n \in A_\iota, x \in A_{\iota^*}$ as follows:

$$x^n(t) = \begin{cases} (1-2^{-n})(\overline{\alpha}_1, \underline{\alpha}_2) + 2^{-n}y(t) & \text{if } t \in S \\ (1-2^{-n})(\underline{\alpha}_1, \underline{\alpha}_2) + 2^{-n}y(t) & \text{if } t \in [0,1]\backslash S \end{cases}$$

$$x(t) = \begin{cases} (\overline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \in S \\ (\underline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \in [0,1]\backslash S \end{cases}$$

Let $P^n(x^n) = 1$ and $P'(x) = 1$. Then $P \in \Pi_\iota$ for all n, $P' \in P_{\iota^*}$ and $\lim d_p(P^n, P') = 0$. Therefore, Lemmas 3 and 4 imply $\eta(S) = \sum_{j=1}^{i} \eta_\iota(S_j^\iota)$.

(ii) Lemmas 2 and 4 ensure that $\eta_\iota(S_j^\iota) \ge 0$ for all ι, j. Therefore, $\eta(S) \ge 0$ and, by (i) above, $\eta(S') \ge \eta(S)$ for $S \subset S'$. Since $V(R^{\overline{\alpha}}) = 1$ it follows that $\eta([0,1]) = 1$. Next, let S_k be a sequence in $\mathcal{S}$ such that $S \subset S_k$ for all k. Let $S^n = \bigcap_{k=1}^{n} S_k$ and note that $S^n \in \mathcal{S}$ for all n. Let $\iota^n = (S^n, [0,1]\backslash S^n)$. Let $x^n(t) = (\overline{\alpha}_1, \overline{\alpha}_2)$ if $t \in S^n$ and $x^n(t) = (\underline{\alpha}_1, \overline{\alpha}_2)$ if $t \notin S^n$. Let P^n be the degenerate REL with $P^n(x^n) = 1$ for each n. Similarly, let $x(t) = (\overline{\alpha}_1, \overline{\alpha}_2)$ if $t \in S$ and $x(t) = (\underline{\alpha}_1, \overline{\alpha}_2)$ if $t \notin S$. Let P' be the degenerate REL with $P'(x) = 1$. If $lS^\infty = lS$ then $\lim d_p(P^n, P') = 0$ and, therefore, $\lim V(P^n) = V(P')$. By Lemma 4 this, in turn, implies that $\lim \eta(S^n) = \eta(S)$, as required. □

Let $P \in \Pi$; for each $x \in D_P := \{x : P(x) > 0\}$, let ι_x be such that $x \in A_\iota$. Lemmas 4 and 5 and the definition of η imply that

$$\begin{aligned} V(P) &= \sum_{x \in D_P} P(x) \sum_{j=1}^{|\iota_x|} f_j(R^x) \eta_{\iota_x}(S_j) \\ &= \sum_{x \in D_P} P(x) \sum_{j=1}^{|\iota_x|} f_j(R^x) \left[\eta(\bigcup_{i=1}^{j} S_i) - \eta(\bigcup_{i=1}^{j-1} S_i) \right] \\ &= \sum_x \int v(u(x)) d\eta P(x) \end{aligned}$$

as desired. □

The proof of the if statement is straightforward. To show uniqueness, let (u, v, η) and $(\hat{u}, \hat{v}, \hat{\eta})$ be two representations of the non-degenerate $\succeq$. Pick $\overline{\alpha}, \underline{\alpha} \in \Delta$ such that $u(\overline{\alpha}) = 1$ and $u(\underline{\alpha}) = 0$. Then, we must have $\hat{u}(\overline{\alpha}) = 1$ and $\hat{u}(\underline{\alpha}) = 0$. Since $u, \hat{u}$ represent the same preference relation on Δ, agree at two distinct points $\overline{\alpha}, \underline{\alpha}$, 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 $\hat{u}$ is onto, we conclude $v = \hat{v}$. The same argument ensures that $V = \hat{V}$. Let $S \in \mathcal{S}$ and choose α, β such that $\alpha_2 = \beta_2$ and $u(\alpha) < u(\beta)$. Since $\succeq$ is non-degenerate, such α, β must exist. Let $x(t) = \beta$ if $t \in S$ and $x(t) = \alpha$ otherwise. Let P be a REL with $P(x) = 1$. Then the representation yields $(1 - \eta S)v(u(\alpha)) + \eta S\, v(u(\beta)) = V(P) = \hat{V}(P) = (1 - \hat{\eta} S)v(u(\alpha)) + \hat{\eta} S\, v(u(\beta))$ thus $\eta S = \hat{\eta} S$.

6.1 Proof of Corollary 1

Let (u, v, η) be an RCU representation and assume Axiom 5* holds. We will show that η is additive, that is, $\eta(S \cup S') = \eta(S) + \eta(S')$ for $S, S' \in \mathcal{S}$ disjoint. Let

$$x(t) = \begin{cases} (\overline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \in S \\ (\underline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \notin S \end{cases}$$

$$y(t) = \begin{cases} (\overline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \in S' \\ (\underline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \notin S' \end{cases}$$

$$z^n(t) = \begin{cases} ((2^{-n}\underline{\alpha}_1 + (1 - 2^{-n})\overline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \in S \cup S' \\ (\underline{\alpha}_1, \underline{\alpha}_2) & \text{if } t \notin S \cup S' \end{cases}$$

Then the REL Q given by $Q(x) = Q(y) = 1/2$ strictly dominates the REL P^n given by $P^n(z^n) = P^n(R^{\underline{\alpha}}) = 1/2$ for all n. Hence $Q \succ P^n$ by Axiom 5^*. Since $\lim d(f(P^n), f(Q)) = 0$ it follows that $\eta(S) + \eta(S') \geq \eta(S \cup S')$. The inequality $\eta(S) + \eta(S') \leq \eta(S \cup S')$ can be shown to hold with an analogous argument. □

7. Appendix B: Proof of Theorems 2-5

7.1 Proof of Theorem 2

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

$$
\begin{aligned}
V(\varepsilon[P]) &= \sum_x \int_0^1 v(u(x_t))d\lambda(t)\, \varepsilon[P](x) \\
&= \sum_x \sum_i v(u(x_{s_{i-1}}))\lambda_i\, \varepsilon[P](x) \\
&= \sum_i \lambda_i \sum_x v(u_1(x_1(s_{i-1})) + u_2(x_2(s_{i-1} + \varepsilon)))\, P(x) \\
&\geq \sum_i \lambda_i \sum_x v(u_1(x_1(s_{i-1})) + u_2(x_2(s_{i-1}))\, P(x) \\
&= V(P)
\end{aligned}
$$

with the convention that $x_2(t) = x_2(1)$ for every $t > 1$. The inequality above follows from the martingale property, the linearity of u_2 and the convexity of v. Hence the SRCU is information-seeking and therefore (iv) implies (iii).

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

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

for some $u_1 < u_2 \in [0, 1]$ and $a \in (0, 1)$. Without loss of generality, $u_2 - u_1 < \max\{u_2(\alpha_2) : \alpha_2 \in \Delta_2\} - \min\{u_2(\alpha_2) : \alpha_2 \in \Delta_2\}$. Then, we can take $\alpha_2, \beta_2 \in \Delta_2$ and $\gamma_1 \in \Delta_1$ such that $u_2(\alpha_2) + u_1(\gamma_1) = u_1$ and $u_2(\beta_2) + u_1(\gamma_1) = u_2$. Let x be an evolving lottery such that

$x_1(t) = \gamma_1$ for all t, $x_2(t) = a\alpha_2 + (1-a)\beta_2$ for $t < 1/2$ and $x_2(t) = \alpha_2$ for $t \geq 1/2$. Also, let y be an evolving lottery with $y_1(t) = \gamma_1$ for all t, $y_2(t) = a\alpha_2 + (1-a)\beta_2$ for $t < 1/2$ and $y_2(t) = \beta_2$ for $t \geq 1/2$. Finally, let P be the REL such that $P(x) = a = 1 - P(y)$. Hence, P offers the constant consumption lottery γ_1 throughout, and the decision maker learns if she gets the prize lottery α_2 or β_2 at time $1/2$. The display equation above yields $V(P) > V(1[P])$ and therefore (ii) implies (iv). □

7.2 Proof of Theorem 3

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

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

$$\int_{[0,1]} v(u(x))d\eta = \int_{S\in\theta} v(u(x_S))dH = \min_{\ell\in L} \sum_{S\in\theta} v(u(x_S))\ell(S) \tag{+}$$

Let ℓ_x be the probability (in L) that solves the maximization problem in (+); that is,

$$\sum_{S\in\theta} v(u(x_S))\ell_x(S) = \min_{\ell\in L} \sum_{S\in\theta} v(u(x_S))\ell(S)$$

Recall that P_1 is the marginal distribution of P at time 1. Hence, we need to show that $V(R^{P_1}) \geq V(P)$. Choose any $\ell \in L$. Since P is a martingale, P_1 is a mean preserving spread of P_t for all $t \in [0, 1]$. Therefore, the convexity of v and the linearity of u imply

$$\begin{aligned}
U(R^{P_1}) &= \sum_x v(u(x(1)))P(x) = \sum_{S\in\theta}\sum_x v(u(x(1)))P(x)\ell(S) \\
&\geq \sum_{S\in\theta}\left(\sum_x v(u(x_S))P(x)\right)\ell(S) = \sum_x \left(\sum_{S\in\theta} v(u(x_S))\ell(S)\right)P(x) \\
&\geq \sum_x \left(\sum_{S\in\theta} v(u(x_S))\ell_x(S)\right)P(x) = U(P)
\end{aligned}$$

as desired.

To prove part (ii), assume that $\succeq$ averse to no disclosure. Then, the convexity of v is an immediate consequence of the fact that the induced preference $\succeq_0$ on M must be risk loving. Finally, assume that v is convex and η is submodular. Let $P \in \Pi$ and define $t_0, \ldots, t_k$, θ, $\mathcal{S}_\theta$ and H as in the proof of part (i) above. Since current consumption is constant, the martingale property ensures that $R^{\bar{x}(P)}$ has a single path $\bar{x}(P) := x^\alpha$ for some $\alpha \in \Delta$. Since η is supermodular, there exists a compact, convex set of probabilities L on the finite set θ such that

$$\int_{[0,1]} v(u(x))d\eta = \int_{\mathcal{S}_\theta} v(u(x_S))dH = \max_{\ell \in L} \sum_{S \in \theta} v(u(x_S))\ell(S) \tag{A8}$$

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

$$\sum_{S \in \theta} v(u(x_S))\ell_x(S) = \max_{\ell \in L} \sum_{S \in \theta} v(u(x_S))\ell(S)$$

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

$$\begin{aligned} V(P) &= \sum_x \left(\sum_{S \in \theta} v(u(x_S))\ell_x(S) \right) P(x) \geq \sum_x \left(\sum_{S \in \theta} v(u(x_S))\ell(S) \right) P(x) \\ &= \sum_{S \in \theta} \left(\sum_x v(u(x_S))P(x) \right) \ell(S) \geq \sum_{S \in \theta} v(\alpha)\ell(S) = v(u(\alpha)) = V(R^{x(P)}) \end{aligned}$$

as desired. □

7.3 Proof of Theorem 4:

Assume $P \in \Pi_c^t$ and v is convex. Hence, $1[P] \succeq R^{x(P)}$ and we need only show that for some η^*, $P \succ 1[P]$ whenever η is more submodular than some η^*. By the martingale property, we can express P as a convex combination of RELs $P^1, \ldots, P^n$ such that $x(P) = x(P^i) = \alpha$ for all i and either each P^i assign's positive probability to exactly two distinct paths or $P^1 = R^\alpha$ and every P^i other than P^1 assign's positive probability to exactly two distinct paths.

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

$(l(S))^{\frac{1}{n}}$ for all $S \in \mathcal{S}$ where l is the Lebesgue measure. It is straightforward to verify that η_n is submodular for all n. Let $P^i(x^\alpha t x^{\beta^1}) = a, P^i(x^\alpha t x^{\beta^2}) = 1 - a$, $\beta^1 \succ_0 \beta^2$, $a \in (0,1)$, $\alpha = a\beta^1 + (1-a)\beta^2$, $v_1 = v(u(\beta^1))$, $v_2 = v(u(\beta^2))$ and finally, let $v_0 = v(u(\alpha))$. Then, for (u, v, η_n), the representation ensures that $P^i \succ 1[P^i]$ is equivalent to

$$\begin{aligned} av_1 + (1-a)v_2 < a\left[v_1\eta_n[t,1] + v_0\left(1 - \eta_n[t,1]\right)\right] \\ + (1-a)\left[v_0\eta_n[0,1-t) + v_2\left(1 - \eta_n[0,1-t)\right)\right] \end{aligned}$$

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

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

for all $S \in \mathcal{S}$. It follows that $P^i \succ 1[P^i]$ for all $P^i \neq R^\alpha$ whenever η is more submodular than η^*. Since $R^\alpha = 1[R^\alpha]$, the linearity of $\succeq$ ensures that $P \succ 1[P]$ whenever η is more submodular than η^*. The proof of the dread case is symmetric and, therefore, omitted. □

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

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

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

When the display equation in Lemma 7 holds for all x, we call h the $\triangleleft$-derivative of H. Suppose $\mathcal{Z}$ is the set of all subsets of some set and H is a capacity. Then, Dempster (1967) shows that H is totally monotone if and only if the $\subset$-derivative of H is nonnegative. This result extends immediately to any capacity on any finite lattice. The following extension to any η defined on the infinite lattice $\mathcal{S}$ is also immediate: let $\theta = \{S_1, \ldots, S_n\}$ be any finite partition of $[0,1]$ into sets in $\mathcal{S}$ and let $\mathcal{S}_\theta$ be the smallest subalgebra of $\mathcal{S}$ that contains θ. The capacity η is totally monotone if and only if for any such $\mathcal{S}_\theta$, there exists a nonnegative function $h : \mathcal{S}_\theta \to [0,1]$ such that

$$\eta S = \sum_{\substack{T \in S_\theta \\ T \subset S}} h(T)$$

for all $S \in S_\theta$. An analogous characterization for dual total monotonicity requires that $h(\bigcup_{j \le k} S_j)$ is non-negative for k odd and non-positive for k even.

The characterization of total monotonicity above is related to the following characterization of the Choquet integral: 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,

$$\int f dH = \sum_{T \in \mathcal{S}_\theta} h(T) \min_{s \in T} f(s) = \sum_{T \in \mathcal{S}_\theta} h^+(T) \max_{s \in T} f(s) \tag{A4}$$

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

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

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 7 and the display equation above imply

$$\begin{aligned} \int f dH &= \sum_{i=1}^{n} r_i \sum_{T \in S^i} h(T) \\ &= \sum_{i=1}^{n} r_{n+1-i} \sum_{T \in S^{+i}} h^+(T) \end{aligned}$$

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

7.4 Proof of Theorem 5:

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

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

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

$$x(s) \succeq_0 \alpha \succ_0 \beta \succ_0 y(s) \text{ for } s < t$$

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

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

Then,

$$\begin{aligned} \frac{1}{b}\left(U(Q^b) - U(Q)\right) &= a\int v(u(x\tau x_\alpha))d\eta + (1-a)\int v(u(x\tau x_\beta))d\eta \\ &\quad - \int v(u(x))d\eta \\ \frac{1}{b}\left(U(Q^g) - U(Q)\right) &= a\int v(u(y\tau x_\alpha))d\eta + (1-a)\int v(u(y\tau x_\beta))d\eta \\ &\quad - \int v(u(y))d\eta \end{aligned}$$

Choose $0 = t_0 < t_1, \ldots, t_k = t$ so that $Q(x) > 0$ implies $x(s') = x(s)$ whenever $s', s \in [t_i, t_{i+1})$ and $i < k$. Let $T_1 = [t, \tau)$, $T_2 = (\tau, 1]$, $\theta = \{[0, t_1), \ldots, [t_{k-1}, t), T_1, T_2\}$ and let $\mathcal{S}_\theta$ be the smallest subalgebra of $\mathcal{S}$ that contains θ. Then, let $\mathcal{S}^o \subset \mathcal{S}_\theta$ 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 $t_i < t_j$ and let $S^* = S \cup \{\emptyset\}$. Let H denote the restriction of η to $\mathcal{S}_\theta$, η^+ be the dual of η and H^+ denote its restriction to $\mathcal{S}_\theta$. Finally, let h, h^+ be the $\subset$-derivatives of H and H^+ respectively.

First, consider the case where η is totally monotone. Let $d^1 = av(\alpha)+(1-a)v(u(\beta))-v(u(\gamma))$ and let $d^2 := av(u(\gamma)) + (1-a)v(u(\beta)) - v(u(\gamma))$. By (A4),

$$\begin{aligned}\frac{1}{b}\left(U(Q^b) - U(Q)\right) &= d_2 \sum_{S\in\mathcal{S}^*} h(S\cup T_1\cup T_2)\\ &\quad + d_1 \sum_{S\in\mathcal{S}^*} h(S\cup T_2)\\ \frac{1}{b}\left(U(Q^g) - U(Q)\right) &= d_2 h(T_1\cup T_2) + d_1 h(T_2)\end{aligned}$$

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 $U(Q^g) \geq U(Q^b)$ as desired.

For the dual totally monotone case, let $d_3 := av(u(\alpha)) + (1-a)v(u(\gamma)) - v(u(\gamma))$. By (A4),

$$\begin{aligned}\frac{1}{b}\left(U(Q^b) - U(Q)\right) &= d_3 h^+(T_1\cup T_2) + d_1 h^+(T_2)\\ \frac{1}{b}\left(U(Q^g) - U(Q)\right) &= d_3 \sum_{S\in\mathcal{S}^*} h^+(S\cup T_1\cup T_2)\\ &\quad + d_1 \sum_{S\in\mathcal{S}^*} h^+(S\cup T_2)\end{aligned}$$

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, $U(Q^g) \geq U(Q^b)$ as desired. $\square$

8. Appendix D: Proof of Theorem 6

8.1 Matching Lemma

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

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

$$\bigcup_{i\in Z} b(i) \leq \sum_{j\in Y_Z(\rho)} \cdot b(j)$$

for all $Z \neq \emptyset$, we say that (ρ, b) is *feasible.* We say that the resource constraint is *tight* if $\sum_{i\in X} b(i) = \sum_{j\in Y} b(i)$. A function $\chi : X \times Y \to [0,1]$ is a *solution* to (ρ, b)

if $\sum_{j \in Y} \chi(i,j) = b(i)$ for all j, $\sum_{i \in X} \chi(i,j) \le b(j)$ for all i and $\chi(i,j) \le \rho(i,j)$ for all $(i,j) \in X \times Y$. The solution χ is *tight* if $\sum_{j \in Y} \chi(i,j) = b(i)$ for all i. Note that if χ is a solution to a matching problem with a tight resource constraint, then it too must be tight.

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

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

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

Proof: Let (ρ, b) be a feasible matching problem. Choose $x_*, y_* \notin X \cup Y$ and let $X_* = X \cup \{x_*\}$, $Y_* = Y \cup \{y_*\}$, $\rho_*(i,j) = \rho(i,j)$ if $i \in X$ and $j \in Y$ and $\rho_*(i,j) = 1$ otherwise. Let

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

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

Inductively define the $2^{k-1} \times k$ matrices X^k, Y^k such that $X^1 = 1, Y^1 = 0$ and, for $k > 1$,

$$X^k = \begin{pmatrix} X^{k-1} & 1_{2^{k-1}} \\ Y^{k-1} & 0_{2^{k-1}} \end{pmatrix}, \quad Y^k = \begin{pmatrix} X^{k-1} & 0_{2^{k-1}} \\ Y^{k-1} & 1_{2^{k-1}} \end{pmatrix}$$

We write x^k_{ij} and y^k_{ij} for the entries of the matrix X^k and Y^k. 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. Fix an ordered partition $\iota = (S_1, \ldots, S_n)$ and define $P^k, Q^k \in \Pi'$ such that

$$P^k = \begin{pmatrix} S_1 & \ldots & S_k & S_{k+1} & \ldots & S_n \\ x^k_{11} & \ldots & x^k_{1k} & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ x^k_{2^{k-1},1} & \ldots & x^k_{2^{k-1},k} & 0 & \ldots & 0 \end{pmatrix},$$

$$Q^k = \begin{pmatrix} S_1 & \ldots & S_k & S_{k+1} & \ldots & S_n \\ y^k_{11} & \ldots & y^k_{1k} & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ y^k_{2^{k-1},1} & \ldots & y^k_{2^{k-1},k} & 0 & \ldots & 0 \end{pmatrix}$$

Recall that each row represents a path and each path is equally likely. For a given capacity η, let H be the restriction of η to the (sub)sigma-algebra $\mathcal{S}_\theta$ where $\theta = \{S_1, \ldots, S_n\}$ and let h be the $\subset -$derivative of H.

Lemma 8: *(i)* $E_{P^k}\left[\int \phi d\eta\right] - E_{Q^k}\left[\int \phi d\eta\right] = h\left(\bigcup_{j=1}^k S_j\right)/2^{k-1}$*; (ii)* P^k *upper dominates* Q^k *for all* k*; (iii)* P^k *lower dominates* Q^k *for* k *odd,* Q^k *lower dominates* P^k *for* k *even.*

Proof: Let $\mathcal{S}_\theta$ be the algebra of sets generated by $\theta = \{S_1, \ldots, S_n\}$ and let $T = \bigcup_{j \le k} S_j$.

First, we observe that $P^k A^{S1} = Q^k A^{S1}$ for all $S \in \mathcal{S}_\theta$, $S \neq T$ and $P^k A^{T1} = 2^{-(k-1)}, Q^k A^{T1} = 0$. For $k = 1$ the assertion is immediate. Given it holds for $k-1$, a straightforward inductive step proves the assertion. To prove part (ii), note that $P^k A^{Sc} = P^k A^{S1}$ and $Q^k A^{Sc} = Q^k A^{S1}$ for all $c > 0$. Therefore, the first assertion above proves that P^k u-dominates Q^k for all k.

Second, we observe that $P^k A_{S0} = Q^k A_{S0}$ for all $S \in \mathcal{S}_\theta$, $S \neq T$ and $P^k A_{T0} = 2^{-(k-1)}, Q^k A_{T0} = 0$ if k is odd and $Q^k A_{T0} = 2^{-(k-1)}, P^k A_{T0} = 0$ if k is even. For $k = 1$ the assertion is immediate. Given it holds for $k-1$, again, a straightforward inductive completes the proof of the claim. To prove part (iii), note that $P^k A_{Sc} = P^k A_{S0}$ and $Q^k A_{Sc} = Q^k A_{S0}$ for all $c < 1$. Therefore, the second assertion implies that P^k lower dominates Q^k if k is odd and P^k lower dominates Q^k if k is even. To prove part (i), note that A4 and the first assertion above imply

$$\begin{aligned} E_{P^k}\left[\int \phi d\eta\right] - E_{Q^k}\left[\int \phi d\eta\right] &= \sum_{S \in \mathcal{S}_\theta} \left(P^k A^{S1} - Q^k A^{S1}\right) h(S) \\ &= 2^{-(k-1)} h\left(\bigcup_{j=1}^k S_j\right) \end{aligned}$$

which completes the proof of Lemma 8. $\square$

Proof of Theorem For any $x \in D$, define $\rho^x \in \Phi$ as follows: $[\rho^x](t) = v(u(x(t)))$. Then, each REL $P \in \Pi$ can be mapped to a unique REL $P' \in \Pi'$ such that $P(x) = P'(\rho^x)$ for all $x \in D$. Let $v(u_1(\bar{\alpha}_1)) = r > 0$ and let $\hat{\Pi}' = \{P \in \Pi' | P(\phi) > 0 \text{ implies } \phi \leq r\}$. Then, for every $P' \in \hat{\Pi}'$ there exists a corresponding $P \in \Pi$. Next, we restate upper and lower domination in terms "utility" RELs: for any $c \in [0,1]$ and $S \in \mathcal{S}$, let $A^{Sc} = \{u \in \Phi, | \, u(t) \geq c \text{ for all } t \in S\}$. Similarly, let $A_{Sc} = \{u \in \Phi, | \, u(t) \leq c \text{ for all } t \in S\}$. For $P, Q \in \Pi'$, we say that P u-dominates (l-dominates) Q if $PA^{Sc} \geq QA^{Sc}$ ($PA_{Sc} \leq QA_{Sc}$) for all $S \in \mathcal{S}$ and $c \in [0,1]$. Since Choquet integration satisfies positive homogeneity, (iv) below is equivalent to part (ii) of Theorem 6:

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

Suppose P u-dominates Q implies $E_P \int u d\eta \geq E_Q \int u d\eta$ for all $P, Q \in \Pi'$. We must show that η is totally monotone. By Lemma 8 (ii), P^k u-dominates Q^k for all k. From Lemma 8 (i) it then follows that if H is the restriction of η on $\theta = \{S_1, \ldots, S_n\}$, then for all $S \in \mathcal{S}_\theta$, $h(S) \geq 0$ where h is the $\subset -$derivative of H. It follows that η is totally monotone (from the characterization of total monotonicity discussed after the proof of Lemma 7).

Suppose P l-dominates Q implies $E_P \int u d\eta \geq E_Q \int u d\eta$ for all $P, Q \in \Pi'$. We must show that η is dual totally monotone. By Lemma 8 (ii), P^k l-dominates Q^k for k odd and P^k u-dominates Q^k for k even. From Lemma 8 (i) it then follows that if H is the restriction of η on $\theta = \{S_1, \ldots, S_n\}$, then for all $S \in \mathcal{S}_\theta$, $h(\cup S_k) \geq 0$ if and only if $(-1)^{k+1} \geq 0$ where h is the $\subset -$derivative of H. It follows that η is dual totally monotone.

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

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

Hence P u-dominates Q implies

$$\sum_{\{u:u(t)\geq c\forall t\in S_N\}} P(u)h(S_N) \geq \sum_{\{u:u(t)\geq c\forall t\in S_N\}} Q(u)h(S_N) \tag{$A6$}$$

for all c.

Define the following matching problem: $\Gamma = \{(S_N, c)\,|\,\emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(P)\}$, $\Upsilon = \{(S_N, c)\,|\,\emptyset \neq N \in \mathcal{M}, h(S_N) > 0, c \in C(Q)\}$, $\rho(i,j) = 1$ if $i = (S_N, c)$, $j = (S_N, \hat{c})$ and $c \geq \hat{c}$ and $\rho(i,j) = 0$ otherwise. Finally, $b(i) = \sum_{\{u:u(t)=c\,\forall t\in S_N\}} P(u)h(S_N)$ for all $i = (S_N, c) \in \Gamma$ and $b(j) = \sum_{\{u:u(t)=c\,\forall t\in S_N\}} P(u)h(S_N)$ for all $j = (S_N, c) \in \Upsilon$. Equation ($A6$) ensures that this matching problem is feasible and, since both P and Q are probabilities, it is tight. Hence, by the matching lemma it has a solution χ. By (A4),

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

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

To conclude the proof of Theorem 6, we will establish the equivalence of (i) and (iii).

To prove (iii) implies (i), note that $P^k \succeq_s^k Q^k$ for all k (by construction). Therefore, the same argument that establishes that (i) implies (ii) then also establishes (i) implies (iii). By Lemma 8 (i), the argument can be reversed to establish the equivalence of (i) and (iii). The proof for the preference for hedging/dual total monotonicity case is symmetric. $\square$

9. Appendix E: Relationship to Habit Models

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

Let $\mathcal{K}$ be the collection of compact subsets of the unit interval endowed with the Hausdorff metric and let $\mathcal{B}(\mathcal{K})$ be the Borel sigma algebra generated by the Hausdorff metric. Let Δ be the set of probability measures defined on $(\mathcal{K}, \mathcal{B}(\mathcal{K}))$. Let μ be a totally monotone continuous capacity. Then, there exists a unique $m \in \Delta$ such that $\mu(K) = m\{K' \subset K\}$ for all $K \in \mathcal{K}$. (Theorem 1.13, pg. 10 and Theorem 2.7 (iii), pg 27. in Molchanov (2005)). Moreover, for any $\phi \in \Phi$

$$\int_{[0,1]} (v \circ \phi) d\mu = \int_{\mathcal{K}} \min\{v(\phi(s)) : s \in K\} dm$$

(Theorem 5.1 pg. 70 in Molchanov (2005)). Note that the right hand side is the Choquet integral whereas the left hand side is a (standard) integral with respect to the unique measure $m \in \Delta$.

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

Then, define $V : [0,1] \times \Phi \to [0,1]$ such that

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

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

$$\begin{aligned}
\int_{[0,1]} V_t(\phi) d\lambda &= \int_{[0,1]} \int_{\mathcal{K}} \min\{v(\phi(s)) : s \in K\} M(dK, s) d\lambda \\
&= \int_{\mathcal{K}} \min\{v(\phi(s)) : s \in K\} dm \\
&= \int_{[0,1]} (v \circ \phi) d\mu
\end{aligned}$$

□

References

Caplin A. and J. Leahy (2001) "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics,* 116, 55–79.

Dempster, A. P. (1967) "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics,* 38, 325–339.

Dillenberger, D. (2010) "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica*, 78, 1973-2004.

Dillenberger, D. and K. Rozen (2015) "History-Dependent Risk Attitude," *Journal of Economic Theory*, 157, 445-477.

Ely, J., A. Frankel and E. Kamenica (2015) "Suspense and surprise," *Journal of Political Economy*, 123, 215–60.

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.

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, 2741–59.

Gilboa, I, (1989), Expectation and variation in multi-period decisions, *Econometrica,* 57, 1153–69.

Gilboa, I. and D. Schmeidler (1994) "Additive representations of non-additive measures and the Choquet integral, *Annals of Operations Research*, 52, 43–65.

Grant, S. , A. Kajii and B. Polak (2000) "Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models," *Econometrica*, 68, 425-434.

Hall, Philip (1935), "On Representatives of Subsets," *Journal of the London Mathematical Society,* 10, 26–30.

Karlsson, N., G. Loewenstein and D. Seppi (2009) "The ostrich effect: selective attention to information," *Journal of Risk and uncertainty,* 38, 95–115.

Loewenstein, G. (1987) "Anticipation and the valuation of delayed consumption," *The Economic Journal,* 97, 666–84.

Lovallo, D. and D. Kahnemann (2000) "Living with uncertainty: attractiveness and resolution timing, *Journal of Behavioral Decision Making*, 13, 171–90.

Molchanov, I, (2005), "Theory of Random Sets," Springer Verlag, London, UK.

Nguyen, H. T., (1978), "On random sets and belief functions," Journal of Mathematical Analysis and Applications, 65, 531–42.

Pollak, R. A., (1970), "Habit formation and dynamic demand functions." Journal of Political Economy 78, 745-763.

Schmeidler, D. (1989) "Subjective probability and expected utility without additivity," *Econometrica,* 57, 571–87.

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