Random Evolving Lotteries and Intrinsic Preference for Information[†]

Faruk Gul

Paulo Natenzon

and

Wolfgang Pesendorfer[‡]

October 2020

Abstract

We introduce random evolving lotteries to study preference for non-instrumental information. Each period, the agent enjoys a flow payoff from holding a lottery that will resolve at the terminal date. We provide a representation theorem for non-separable risk-consumption preferences and use it to characterize agents' attitude to non-instrumental information. To address applications, we characterize *peak-trough* utilities that aggregate trajectories of flow utilities linearly but, in addition, put weight on the best (peak) and worst (trough) lotteries along each path. We show that the model is consistent with recent experimental evidence on attitudes to information, including a preference for gradual arrival of good news and the ostrich effect, i.e., decision makers' tendency to prefer information after good news to information after bad news.

[†] This research was supported by grant SES-1426252 from the National Science Foundation. We thank Yusufcan Masatlioglu and Erkut Ozbay for insightful comments and suggestions.

[‡] Faruk Gul, Department of Economics, Princeton University, fgul@princeton.edu; Paulo Natenzon, Olin Business School, 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 assets at retirement, 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 outcome. 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 while the relevant choice objects are *random evolving lotteries*; that is, lotteries defined on such functions.

In this paper, we formulate a model of risk consumption and use it to study dynamic preference for non-instrumental information. Our model has three parameters; a utility index u that determines the decision maker's preferences over static consumption and terminal prize lotteries, a real valued function v that specifies the flow utility associated with any current consumption and beliefs over terminal prizes, and finally, a capacity η that aggregates trajectories of flow utilities by identifying each such trajectory with its Choquet integral. Our first result provides a characterization of *risk-consumption preferences* (RCPs) and shows their *risk-consumption utility* (RCU) representation is unique.

We apply our model to study an agent's (non-instrumental) attitude to information. To do so, we provide two notions of preference for more information. The first, *information seeking*, requires that an agent prefer one random evolving lottery over another if the former yields a mean-preserving spread of the latter in every period. Thus, information seeking agents are better off any time they receive more information. Our second, less demanding notion, *curiosity*, requires the agent to be better off if, after any given history, she receives some information as opposed to no information. Thus, a curious agent's utility cannot increase if she "turns off the news." In Theorem 2, we show that an agent is curious if and only if her v is convex and her η satisfies a weaker version of submodularity, which we call tail-submodularity. In Theorem 3, we show that an agent is information seeking if and only if v is convex and η is additive. Thus, an agent can only be information seeking if her evaluations of trajectories is separable.

To address applications, we characterize a special class of risk-consumption utilities which we call *peak-trough utility.* A peak-trough utility aggregates trajectories of flow utilities linearly but, in addition, puts special weight on the best (peak) and worst (trough) lotteries along each path. This formulation is analytically tractable and draws its inspiration from Fredrickson and Kahnemann (1993) who argue that, in retrospective evaluations, subjects neglect the duration of experiences and emphasize extremes.

Peak-trough utility adds two parameters to the additively separable model, a weight δ_h for the best experience along a path and another weight δ_ℓ for the worst experience along the path. To interpret these weights, consider random evolving lotteries that provide information about a binary terminal prize. The last period reveals whether the agent "won" or "lost" the prize and the initial period reveals no information other than the prior probability of winning. In between the first and last period, the random evolving lottery reveals information about the agent's chance of winning. In this context, the parameter δ_h measures the agent's attitude to "good news," that is, intermediate realizations above the initial prior. If $\delta_h < 0$, the agent dislikes partial good news and would rather resolve the uncertainty fully. Thus, $\delta_h < 0$ describes an agent who "does not want to get her hopes up." In contrast, if $\delta_h > 0$, the agent savors the gradual arrival of good news and prefers partial to full revelation at the intermediate stage. An analogous interpretation applies to the parameter δ_ℓ, albeit with signs reversed. If $\delta_\ell > 0$, the agent dreads the gradual arrival of bad news and would rather "get it over with quickly" while $\delta_\ell < 0$ describes an agent who prefers bad news to arrive gradually rather than all at once.

Gul, Natenzon, Ozbay and Pesendorfer (2020) conduct experiments to examine subjects intrinsic preference for information. The experiment asks subjects to choose between immediate and gradual (or late) resolution of uncertainty in different settings. In one treatment, partial good news arrives gradually while bad news is decisive; in the other, partial bad news arrives gradually while good news is decisive. More than 2/3 of the subjects are not indifferent to how uncertainty resolves. Of the non-indifferent subjects around 58% prefer immediate resolution of uncertainty in the partial bad news setting while approximately 60% of subjects prefer gradual resolution of uncertainty in the partial good news setting. Thus, the experiment provides evidence indicating that the distinction between partial good and partial bad news is quantitatively significant. In section 4.1, we show that peak-trough utility can explain this evidence; in particular, Proposition 2 relates the parameters of peak-trough utility to the agent's attitude towards intermediate good or intermediate bad news. In section 4.2, we characterize the optimal information structure for a peak-trough utility agent and give conditions under which it features intermediate good news.

Agents who like partial good (bad) news but dislike partial bad (good) news exhibit a preference for *skewed* information. Section 4.3 provides a definition of skewed information[1] and shows that the sign of $\delta_h + \delta_\ell$ identifies this preference. Masatlioglu, Orhun and Raymond (2017) conduct experiments to determine whether information skew is important for agent's rankings of dynamic lotteries. Translated to our setting, their experiment has 3 periods, the first period reveals the prior probability of winning while the last period reveals information fully. The experiment elicits subjects' preferences over information during the intermediate period. The authors find that there is substantial heterogeneity in subjects' attitudes regarding how uncertainty resolves and that a large majority of subjects exhibit preference for skewed information; moreover, most subjects are not close to indifferent as to how uncertainty resolves. Thus, the result in Masatlioglu et al (2017) also suggests that the effects isolated by peak-trough utility constitute a quantitatively meaningful aspect of behavior.

In our last application, we address the "ostrich effect." Consider the following scenario: individuals have invested their retirement savings in target date funds and therefore never change their asset allocations. Nonetheless, they frequently check the balances of their retirement accounts. Furthermore, individuals are more likely to check their accounts following days when the market is up than following days when the market is down. In other words, they are more inclined to obtain information after good news than after bad news. Karlsson, Loewenstein and Seppi (2009) provide evidence that this type of behavior is fairly common and call it the *ostrich effect*.[2] In section 4.4 we consider a stylized choice problem that replicates the scenario described above: the agent must choose whether to obtain information about a terminal-prize lottery. We show that peak-trough utilities with positive δ_h and δ_ℓ display the ostrich effect: that is, agents like getting additional information after good news and sometimes wish no additional information after bad news. In the online appendix, we provide a

[1] Our definition of preference for skewed information is related to the definition due to Dillenberger and Segal (2017).

[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. We follow Karlsson, Loewenstein and Seppi (2009) and use the term to describe investors who avoid information after bad news but may seek it after good news.

recursive formulation of our model and show that choices are dynamically consistent; we use the dynamic choice setup to show that the ostrich effect holds more generally.

Choices of peak-trough agents are history dependent; specifically, the decision to obtain information depends on how the current belief about the terminal lottery compares to past peaks and troughs. For agents with positive δ_h, δ_ℓ, information becomes more desirable the less likely it reaches a new trough and the more likely it reaches a new peak. Thus, good news encourages further information acquisition in subsequent periods while bad news discourages it. Agents with negative parameters δ_h, δ_ℓ exhibit the reverse behavior: they are less inclined to gather information when their current beliefs are near their historical peaks and more inclined when beliefs are near a trough.

Our analysis focuses on information that is not decision relevant. Of course, in many economic problems information has an instrumental role. The retirement savers of the above example may regularly re-balance their portfolios and individuals who learn about their health status may seek treatment or make lifestyle changes. It is straightforward to extend our model of recursive choice to capture demand for information that has instrumental as well as intrinsic value. In this context, a preference for skewed information would translate into an asymmetric response to news: following good news subjects are more inclined to gather information and therefore will be more responsive than following bad news.

1.1 Related Literature

Our approach is related to that of Gilboa (1989) who was the first to use capacities to model time non-separability.[3] In Gilboa's model, utility flows arise from flows of consumption; in ours they arise from anticipatory feelings about the ultimate realization of an uncertain outcome. This difference aside, Gilboa's variation averse preferences are a special class of our general model.

Kreps and Porteus (1978) (henceforth KP) formulate the first model of preference for the timing 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 $I\!R^k$. In KP, each path is also a sequence of probability distributions but each of these distributions is over a more complicated space of probability distributions. Since our random evolving

[3] Following Gilboa (1989), several authors have modeled complementarities between consumption flows in different time periods with capacities: Shalev (1997), De Waegenaere and Wakker (2001), Chateauneuf and Rebille (2004), Rebille (2007), Chateauneuf and Ventura (2013), and Bastianello and Chateauneuf (2016).

lotteries are defined over simpler choice objects, they are easier to relate to observables than temporal lotteries (i.e., they require fewer assumptions when relating to data.)

Our model and the KP model are not nested. The KP model enables a decision maker to value information about what information she will have in the future even if that information has no effect on her beliefs about final outcomes. Our model rules out this possibility. 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 even though 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 them susceptible to a particular type of cancer. The test, however, is only effective in patients that have a particular blood enzyme. In patients without the enzyme, the test is uninformative. The doctor assures the patient that checking for the blood enzyme is simple, painless and can be carried out either on the morning of October 8 ($t = 2$) 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 she knows regarding her 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 for the enzyme test 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 having enzyme test results on the date of the genetic screening or a week earlier.

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 papers provide experimental evidence that relates the specifics of the anticipated consumption to the decision maker's preference and identify conditions that lead the individual to savor or dread future consumption.

Caplin and Leahy (2001) offer a theoretical model of *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. 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, Kajii 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 to gradual resolution of uncertainty. Although our model maintains independence, the relaxation of time separability allows us to accommodate such 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, in

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

our model, like theirs, 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 Ely, Frankel and Kamenica (2015) study. In their model, agents derive utility from *changes* in the lottery over terminal prizes. This is motivated by agents' desire for surprise and suspense.

Formally, our analysis is related to the ambiguity literature, in particular, to Schmeidler's (1989) Choquet expected utility theory. Ignoring the temporal dimension of our model and treating the set of time periods as an abstract state space, our rank-dominance axiom can be translated to the classic Anscombe and Aumann (1963) setting as a weakening of the monotonicity axiom. Seo (2009) proposes a substantially different weakening of monotonicity in the Anscombe-Aumann setting, and obtains a foundation for the smooth ambiguity model of Klibanoff, Marinacci, and Mukerji (2005). Our temporal setting, on the other hand, has no ambiguity, and the Choquet integral facilitates preferences that are not separable across time. Kreps and Porteus (1978) and Epstein and Zin (1989) provide a different approach to modeling non-separable time preferences. Finally, our proofs use a characterization of integration with a total monotone (or dual totally monotone) capacity similar to the one provided by Gilboa and Schmeidler (1994).

2. 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 $\{\omega \in \Omega \,|\, \theta(\omega) > 0\}$ such that $\sum \theta(\omega) = 1$. For $X \subset \Omega$, we let $\theta X = \sum_{\omega \in X} \theta(\omega)$ and define a sum over the null set as 0. Let $\Delta(\Omega)$ denote the set of probabilities on Ω. 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 $E_\theta[f] = (E_\theta[f_1], \ldots, E_\theta[f_k])$.

Let A be a non-empty finite set of (flow) consumption levels, and let B be a non-empty finite set of terminal prizes. A *lottery* is a pair (α, β) where α is a consumption lottery (that is, probability on A) and β is prize lottery (a probability on B). We let $L = \Delta(A) \times \Delta(B)$ be the set of lotteries. When convenient, we identify each of $\Delta(A)$ and $\Delta(B)$ with its corresponding finite dimensional simplex.

Time is discrete with a finite horizon $t \in \{1, 2, \ldots, N\}$. With some abuse of notation, we let N denote the set $\{1, \ldots, N\}$. An *evolving lottery* (or *path*) $x = (x_1, \ldots, x_N) \in L^N$ is a sequence of lotteries $x_t = (\alpha_t, \beta_t)$. The consumption lottery α_t is the (possibly stochastic) consumption in period t; while the prize lottery β_t reflects the current information about the realization of the prize in period $N+1$. We endow the set of evolving lotteries L^N with the product topology.

For any probability P on L^N and any subset $X \subset L^N$ such that $PX > 0$, let P_X be the conditional probability of P given X; that is,

$$P_X(x) = \begin{cases} \frac{P(x)}{PX} & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}$$

A probability P on L^N is a *random evolving lottery (REL)* if it satisfies the following *martingale property*: for any time $t < N$ and any sequence of lotteries $((\alpha_1, \beta_1), \ldots, (\alpha_t, \beta_t))$, if $PX > 0$, for $X = \{x \in L^N \mid x_1 = (\alpha_1, \beta_1), \ldots, x_t = (\alpha_t, \beta_t)\}$, then

$$E_{P_X}[\beta_{t+1}] = \beta_t.$$

Let $\Pi \subset \Delta\left(L^N\right)$ be the set of RELs. It follows from the martingale property (and the law of iterated expectations) that $E_P[\beta_t] = E_P[\beta_1]$ for all $P \in \Pi$.

For a given lottery $(\alpha, \beta) \in L$ we let $x_{(\alpha,\beta)} = ((\alpha, \beta), \ldots, (\alpha, \beta)) \in L^N$ denote the constant path equal to (α, β) in every period t. By the martingale property, if $P(x) = 1$ for some path x, then there exists a fixed prize lottery β such that $x_t = (\alpha_t, \beta)$ for all t. For each lottery $(\alpha, \beta) \in L$, we let $P_{(\alpha,\beta)}$ denote the degenerate REL such that $P_{(\alpha,\beta)}\left(x_{(\alpha,\beta)}\right) = 1$; thus, the REL $P_{(\alpha,\beta)}$ reveals no information along the way and the decision-maker consumes (α, β) throughout.

A *second-order lottery* is an element of $\Delta(L)$ with generic element p. For each REL P and time t, define the second-order lottery $P_t \in \Delta(L)$ as follows:

$$P_t(\alpha, \beta) = P\left\{x \in L^N \mid x_t = (\alpha, \beta)\right\}.$$

Hence, P_t is the time-t distribution of P. For any second-order lottery $p \in \Delta(L)$, let P_p be the REL such that $P_p(x_{(\alpha,\beta)}) = p(\alpha, \beta)$. If p is non-degenerate, then the REL P_p reveals some information in the first period but reveals no information thereafter.

Let $\succcurlyeq$ be a binary relation on Π; that is, a subset of $\Pi \times \Pi$. Let $\sim$ denote the symmetric part of $\succcurlyeq$, that is $P \sim P'$ whenever $P \succcurlyeq P'$ and $P' \succcurlyeq P$. We say that $\succcurlyeq$ is *degenerate* if

$P_{(\alpha,\beta)} \sim P_{(\alpha',\beta')}$ whenever $\alpha = \alpha'$ or if $P_{(\alpha,\beta)} \sim P_{(\alpha',\beta')}$ whenever $\beta = \beta'$. We require $\succcurlyeq$ to be a non-degenerate binary relation that satisfies the following axioms:

Axiom 1: *$\succcurlyeq$ is complete and transitive.*

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

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

We endow Π with the Prohorov metric.[5] Our next axiom is continuity:

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

The restriction of $\succcurlyeq$ to $\{P_{(\alpha,\beta)} \in \Pi \,|\, (\alpha,\beta) \in L\}$ induces a preference over lotteries. The next Axiom guarantees that this induced preference satisfies independence.

Axiom 4: *If $P_{(\alpha,\beta)} \succ P_{(\alpha',\beta')}$ and $a \in (0,1)$ then*

$$P_{a(\alpha,\beta)+(1-a)(\alpha'',\beta'')} \succ P_{a(\alpha',\beta')+(1-a)(\alpha'',\beta'')}.$$

Restricting $\succcurlyeq$ to $\{P_p \in \Pi \,|\, p \in \Delta(L)\}$ yields an induced preference on $\Delta(L)$. With some abuse of notation, we write $p \succcurlyeq p'$ instead of $P_p \succcurlyeq P_{p'}$. Axiom 2 ensures that this induced preference over second-order lotteries satisfies independence. For $P, P' \in \Delta(L^N)$, we say that P *dominates* P' if $P_t \succcurlyeq P'_t$ for all t. In other words, P dominates P' whenever the t-th coordinate distribution of P is preferred to the t-th coordinate distribution of P' for every t. P *strictly dominates* P' if P dominates P' and P' does not dominate P. The following axiom implies separability across time intervals:

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

The goal of our paper is to study phenomena, such as the ostrich effect, that are inconsistent with Axiom 5*. Karlsson, Loewenstein and Seppi (2009) document evidence showing

---

[5] Formally, for each $X \subset L^N$ and $\epsilon > 0$, let $X^\epsilon = \{x \in L^N \mid \inf_{y \in X} d(x,y) < \epsilon\}$. Then, the Prohorov metric, d_p, is defined as follows:

$$d_p(P,P') = \inf\{\epsilon > 0 \,|\, PX \leq P'X^\epsilon + \epsilon \text{ and } P'X \leq PX^\epsilon + \epsilon \text{ for all } X \subset L^N\}.$$

that investors are more eager to learn about their own portfolios when stock market indices have gone up than when they have gone down. This suggests that even if two RELs, P and P', have identical coordinate distributions at each date, one may be more attractive than the other if the former reveals more information following good news than the latter. Our weakening of Axiom 5^* allows for this non-indifference but maintains dominance under a more stringent condition.

A finite sequence $\iota = (S_1, \ldots, S_n)$ is an *ordered partition* of time N if the sets $S_1, \ldots, S_n$ are non-empty, pairwise disjoint and $S_1 \cup \cdots \cup S_n = N$. Given any ordered partition $\iota = (S_1, \ldots, S_n)$, let

$$X_\iota = \{x \in L^N \mid P_{x_t} \succ P_{x_{t'}} \text{ if and only if } t \in S_i, t' \in S_j \text{ for some } i < j\}$$

be the $\iota-$paths. We say that P *rank-dominates* P' if $PX_\iota = P'X_\iota$ for every ordered partition ι, and $PX_\iota > 0$ implies P_{X_ι} dominates P'_{X_ι}.

To see what rank-domination means, fix an arbitrary ordered partition ι and note the evolving lotteries in X_ι all yield better lotteries at times $t \in S_i$ than at times $t' \in S_{i+1}$. Hence, these evolving lotteries confront the decision maker with the same pattern of intertemporal variation in utility. For P to rank-dominate P', we require them to assign the same probability to each collection of such evolving lotteries, X_ι, and for P to dominate P' conditional on every X_ι. Hence, rank-domination is a version of domination that permits the decision-maker to take into consideration the pattern of intertemporal variation in utility when assessing RELs. P *strictly rank-dominates* P' if P rank-dominates P' but P' does not rank-dominate P.

Axiom 5: *P strictly rank-dominates P' implies $P \succ P'$.*

Our utility representation has three parameters; a utility over lotteries $u : L \to [0, 1]$; a second-stage index $v : [0, 1] \to [0, 1]$ that measures the agent's attitude to two-stage lotteries; and a capacity η that aggregates utility flows over time.

The continuous function $u : L \to [0, 1]$ is a *utility* if it is onto and separable; that is, if there exist $u_1 : \Delta(A) \to [0, 1]$ and $u_2 : \Delta(B) \to [0, 1]$ such that $u(\alpha, \beta) = u_1(\alpha) + u_2(\beta)$ for all $(\alpha, \beta) \in L$. A utility is linear if

$$u(a\alpha + (1-a)\alpha', a\beta + (1-a)\beta') = au(\alpha, \beta) + (1-a)u(\alpha', \beta').$$

Let Λ be the set of all continuous, strictly increasing functions from $[0,1]$ onto itself. A second-stage index is a function $v \in \Lambda$.

Let $\mathcal{S}$ be the set of all subsets of N. A *capacity* on N is a function $\eta : \mathcal{S} \to [0,1]$ such that $\eta\emptyset = 0$, $\eta N = 1$, and $\eta S \leq \eta T$ for all $S \subset T$. The capacity η is *strict* if $\eta S < \eta T$ whenever $S \subset T \neq S$. For any function $f : N \to I\!R$, the *Choquet integral* of f with respect to the capacity η is

$$\int f d\eta := \int_0^\infty \eta\{t \,|\, f(t) \geq \zeta\}\, d\zeta + \int_{-\infty}^0 (\eta\{t \,|\, f(t) \geq \zeta\} - 1)\, d\zeta$$

where each integral on the right hand side is the usual Riemann integral.

A function V represents $\succcurlyeq$ whenever $P \succcurlyeq P'$ holds if and only if $V(P) \geq V(P')$. Such a function is a *risk consumption utility* (RCU) if there is a linear utility u, a second-stage index v, and a strict capacity η such that

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

for all P. If $\succcurlyeq$ can be represented by an RCU, we call it a *risk consumption preference* (RCP).

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

In the following we will identify a risk consumption preference with the parameters (u, v, η) of its representation. Thus, when we refer to the RCP (u, v, η) we mean the preference that is represented by the RCU V with parameters (u, v, η).

A risk consumption utility V is a linear function on Π and the period utility u is a linear function on L. However, the utility of a path need not be separable across time periods. To obtain additivity, we replace Axiom 5 with the stronger Axiom 5*.

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

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 (period) utilities. We use the acronym SRCP (SRCU) for the separable risk consumption preferences (utilities) of Corollary 1. Below, we use λ to denote an additive capacity (a probability, which corresponds to standard time discounting weights) and we identify the SRCP with the parameters (u, v, λ) of its representation.

3. Preference for Information

In this section we relate the parameters of the RCU to an agent's attitude to information. Throughout this section, we assume that all RELs in every statement yield the consumption lottery α in every period. Hence, over time, the agent's physical consumption does not vary while her risk consumption varies as she learns more about her terminal prize lottery.

We present two notions of preference for more information. The first, *curiosity*, requires that the decision maker prefers the REL P to the REL P' whenever P' yields no information after some t-period history while P yields some information after that history and is otherwise the same as P'. Thus, a curious agent never increases her utility by "switching off the news." The second and broader notion is *information seeking*; it states that no matter how much information the agent has in any period, she is better off having even more information. While a curious agent prefers some information to none, an information seeking agent always wants more information. As we show below, every information seeking agent is curious but the converse is not true.

Formally, let $X^t = \{x \in L^N \mid x_s = x_t \text{ for all } s \geq t\}$; hence X^t is the set of all paths that yield no information after period t. Then, let $[x]_t = \{z \in L^N \mid z_s = x_s \text{ for all } s \leq t\}$; that is, $[x]_t$ is the set of all paths that agree with x until (and including) period t.

The REL P is *simply more informative* than P' after some t-period history if there is some $x \in X^t$ such that $P[x]_t = P'(x)$ and $P(z) = P'(z)$ for all $z \notin [x]_t$. Thus, P' is identical to P except that it stops the flow of information after history x. The REL P is *simply more informative* than P' after the null-history (or at the outset) if $P' = P_{\bar{x}(P)}$. We say P *simply more informative* than P' when the above definition holds for some history and there is no need to specify it.

Definition: *The RCP $\succcurlyeq$ is curious if P is simply more informative than P' implies $P \succcurlyeq P'$.*

Let $S_t^+ = \{t, \ldots, N\}$ and let $S_t^- = \{1, \ldots, t\}$. To relate curiosity to the parameters of RCU, we need the following weaker notion of submodularity:

Definition: *The capacity η is tail-submodular if $\eta S + \eta T \geq \eta(S \cup T) + \eta(S \cap T)$ whenever $(S \cap S_t^-) \cup \{t\} \subset T \subset S_t^-$ and $S_{t+1}^+ \subset S \subset N \backslash \{t\}$ for some t.*

A capacity is submodular if the inequality in the definition above is satisfied for all S, T. Thus, a submodular capacity is tail-submodular but the reverse is not true. As an illustration, consider the case where $N = \{1, 2, 3\}$. Then let η be the capacity such that $\eta N = \eta\{1, 3\} = \eta\{2, 3\} = \eta\{1, 2\} = 1, \eta\{2\} = \eta\{3\} = 1/2$ and $\eta\{1\} = 0$. This capacity is tail-submodular but not submodular.

Theorem 2 provides a characterization of curious RCUs:

Theorem 2: *The RCP (u, v, η) is curious if and only if v is convex and η is tail-submodular.*

If P is simply more informative than P', then P replaces one path, x, of P' that is constant from period t onwards with a number of paths that may or may not be constant on that interval and agree with x up to period t. Hence, by the martingale property, P_s is a mean-preserving spread of P'_s for all $s > t$ and P' has flatter paths; that is, paths with less intertemporal variation than P. A decision maker with a convex v prefers mean-preserving spreads, while a decision maker with a submodular capacity prefers more intertemporal variation. Hence, an RCU agent with a convex and a submodular η will prefer P to P' whenever P is simply more informative than P'. Because of the martingale property, the full force of submodularity is not needed; as Theorem 2 shows, tail-submodularity is enough.

To formulate an alternative and broader notion of preference for information, we need the following broader notion of more informativeness: P is more informative than P' if P_t is a mean-preserving spread of P'_t for all t.[6]

Definition: *The RCP $\succcurlyeq$ is information seeking if $P \succcurlyeq P'$ whenever P is more informative than P'.*

More informativeness is a weaker notion than simple more informativeness because, as we noted above, P'_t is a mean-preserving spread of P''_t for all t whenever P' is simply more informative than P''. However, the converse is not true. Consider the following three RELs: P^0, P', P''. In all three RELs, the agent is either going to get terminal prize 1 or 2 (with equal probability). The REL $P'' = P_{(\alpha,\beta)}$ has a single constant path and the agent gets no additional information in any period. The REL P' resolves all uncertainty in some period $t \neq 1, N$ while the REL P^0 resolves all uncertainty at the outset; that is, has two constant paths and assigns probability 1/2 to each.

Note that P^0 is more informative than P', which in turn, is more informative than P''. However, P^0 is not simply more informative than P'. Also note that P^0 and P'' have constant

[6] P_t is a mean preserving spread of P'_t if for every element β in the support of P'_t there is $T_\beta \in \Delta(\Delta(B))$ such that $P'_t(\alpha, \beta) = \sum P_t(\alpha, \beta')T_\beta(\beta')$ and $\beta = \sum \beta' T_\beta(\beta')$.

paths while P' has paths that display intertemporal variation. An information seeking agent must prefer P^0 to P' and P' to P''. Thus, this agent must sometimes prefer RELs with paths that have greater intertemporal variation and other times prefer paths that have less intertemporal variation. As we show in Theorem 3, below, this requires that the agent is, in fact, indifferent to interpempemporal variation; that is, the capacity η is additive. It follows that an information seeking preference must rely solely on the convexity of v:

Theorem 3: *The RCP (u, v, η) is information seeking if and only if v is convex and η is additive.*

While information seeking RCUs always favor immediate disclosure, the following example shows that curious RCUs may favor information disclosure at intermediate times. For any two evolving lotteries $x, x' \in L^N$ we write xtx' for the path x'' such that $x''_s = x_s$ for $s < t$ and $x''_s = x'_s$ for $s \geq t$. Let $(\alpha, \beta), (\alpha, \beta')$ be two lotteries that yield the same immediate consumption but differ in the prize lottery, let $\beta'' = \beta/2 + \beta'/2$ and let P^t be the REL that has two equiprobable evolving lotteries and reveals the uncertainty about the prize lottery (β or β') at time t:

$$P^t \left(x_{(\alpha,\beta'')} t x_{(\alpha,\beta)}\right) = P^t \left(x_{(\alpha,\beta'')} t x_{(\alpha,\beta')}\right) = 1/2.$$

The submodular capacity η is *quadratic* if

$$\eta T = 2|T|/N - (|T|/N)^2$$

Proposition 1: *Let (u, v, η) be an RCP. Assume η is quadratic, $u_2(\beta) > u_2(\beta')$ and define*

$$r = \frac{v(u(\alpha, \beta'')) - v(u(\alpha, \beta'))}{v(u(\alpha, \beta)) - v(u(\alpha, \beta'))}$$

Then, $P^s \succ P^t$ if $t < s \leq Nr$ or $Nr \leq s < t$.

Every path of the REL P^t has two possible lotteries, one yielding higher utility than the other. When evaluating a path, the quadratic capacity in Proposition 1 puts greater weight on a period if it yields the high-utility lottery than if it yields the low utility lottery in the path. This "savoring" of good experiences explains why the agent prefers to wait some time before information is disclosed; bad continuations will be discounted relative to the early periods while good continuations will be emphasized.

Let $u(\alpha, \beta) = g, u(\alpha, \beta') = b$. Then, $u(\alpha, \beta'') = (b + g)/2$ and

$$v((b + g)/2) = rv(g) + (1 - r)v(b)$$

Thus, $r - 1/2$ is the *probability premium* of the function v for the lottery that yields b and g with equal probabilities. If v is linear, then $r = 1/2$ and, therefore, the optimal disclosure time is $t = N/2$ (assuming N is even). For convex v the optimal disclosure time is earlier while for concave v it is later. Note that if v is convex, then the RCP is curious. Hence, Proposition 1 shows that curious agents may favor disclosure of information at intermediate dates.

Lovallo and Kahnemann (2000) observe that subjects sometimes prefer to delay the resolution of gambles to savor the possibility of winning. They study subjects' willingness to delay resolution of uncertainty of two-outcome gambles and find that subjects' inclination to delay resolution depends on the attractiveness of the gamble: more attractive lotteries lead to a greater fraction of subjects willing to delay resolution. We can replicate their comparative statics finding if we assume $-v''/v'$ is an increasing function. In the setting of Proposition 1, above, a more attractive lottery will lead to greater desired delay if the probability premium is increasing in the attractiveness of the lottery. It is well known that the probability premium increases if the function v has increasing absolute risk aversion; that is, if $-v''/v'$ is an increasing function.

4. Peak-Trough Preferences

In this section, we strengthen Axiom 5 to obtain utility functions, *Peak-Trough Utilities* (PTUs), that are tractable and enable us to analyze several applications. A PTU agent's utility from a path has two components; a standard separable component and a component that depends on the extreme experiences; that is, the best (peak) and worst (trough) points of the path. This functional form takes its inspiration from Kahnemann's emphasis on extreme experiences and duration neglect (Fredrickson and Kahnemann (1993)).

For any u, v and x, let $\overline{vu}(x) = \max_t v(u(x_t))$ and $\underline{vu}(x) = \min_t v(u(x_t))$. Recall that a separable risk consumption utility (SRCU) is an RCU for the which capacity is additive. The utility function V is a *Peak-Trough Utility* (PTU) if there is an SRCU $U := (u, v, \lambda)$, real numbers δ_h, δ_ℓ satisfying $1 - \delta_h - \delta_\ell > 0$ and, for all t,

$$(1 - \delta_h - \delta_\ell)\lambda(t) + \delta_h > 0$$

$$(1 - \delta_h - \delta_\ell)\lambda(t) + \delta_\ell > 0$$

such that, for all P,

$$V(P) = \delta_h E_P \overline{vu} + \delta_\ell E_P \underline{vu} + (1 - \delta_h - \delta_\ell)U(P) \tag{PT}$$

To verify that peak-trough utility is a special case of an RCU, define the capacity η as follows:[7] $\eta\emptyset = 0, \eta N = 1$ and, for every T such that $\emptyset \neq T \neq N$,

$$\eta T = (1 - \delta_h - \delta_\ell)\lambda T + \delta_h$$

Then, the RCU $V = (u, v, \eta)$ coincides with the peak-trough utility with parameters $u, v, \lambda, \delta_\ell$, and δ_h. Thus, the content of a PTU representation can be summarized as follows: an RCU with capacity η has a PTU representation if and only if

$$\eta(S \cup \{t\}) - \eta(S) = \eta(T \cup \{t\}) - \eta T$$

for all nonempty S, T such that $t \notin S \cup T$ and $S \cup \{t\} \neq N \neq T \cup \{t\}$. That is, an RCU has a PTU representation if and only if adding $t \notin S$ to any nonempty S such that $S \cup \{t\} \neq N$ yields the same increase in capacity.

Axiom PT, below, in conjunction with Axioms 1–4 characterizes preferences that have PTU representations. To state the axiom, we need to define the distribution of path-maximal and path-minimal outcomes. For each path x and preference $\succcurlyeq$ over RELs, let $h(x, \succcurlyeq)$ be the best lottery of x. That is, $h(x, \succcurlyeq) = x_t$ such that $P_{x_t} \succcurlyeq P_{x_s}$ for all $1 \leq s \leq N$. If there are multiple such x_t's, we choose one arbitrarily. Similarly, let $\ell(x, \succcurlyeq) = x_t$ such that $P_{x_s} \succcurlyeq P_{x_t}$ for all $1 \leq s \leq N$ be the worst lottery of x. Then, for each REL P, define the second-order lotteries P_h and P_ℓ as follows:

$$P_h(\alpha, \beta) = P\{x \,|\, h(x, \succcurlyeq) = (\alpha, \beta)\}$$
$$P_\ell(\alpha, \beta) = P\{x \,|\, \ell(x, \succcurlyeq) = (\alpha, \beta)\}$$

Thus, P_h and P_ℓ are the distributions of the path-maximal and path-minimal outcomes for the REL P.

Definition: *P and $\hat{P}$ are experience equivalent if $P_t \sim \hat{P}_t$ for all $t \in N \cup \{h, \ell\}$.*

[7] Chateauneuf, Eichberger and Grant (2007) coined the term *neo-additive* for this type of capacity.

Thus, P and $\hat{P}$ provide the same experience if the t-th coordinate distribution of P is indifferent to the t-th coordinate distribution of $\hat{P}$ for every t and if the distributions of the path-peak and path-trough of P are indifferent to the corresponding distributions of $\hat{P}$. An REL P is *degenerate* if $P(x) = 1$ for some x. The following Axiom requires that the agent's preference is monotone for degenerate RELs and indifferent between experience-equivalent RELs:

Axiom PT: *(i) If P and $\hat{P}$ are experience equivalent, then $P \sim \hat{P}$; (ii) If P and $\hat{P}$ are degenerate and P strictly dominates $\hat{P}$, then $P \succ \hat{P}$.*

Clearly, Axiom PT is weaker than Axiom 5*, discussed above. As Theorem 4, below, reveals, Axiom PT implies Axiom 5 in the presence of Axioms 1-4. We identify any PTU with the associated $(u, v, \lambda, \delta_h, \delta_\ell)$. We call an RCP that has a PTU representation a PT preference.

Theorem 4: *A non-degenerate $\succcurlyeq$ satisfies Axioms 1–4 and Axiom PT if and only if it is a PT preference.*

It is easy to verify, and we do so in the proof of Theorem 4, that the PTU representation $(u, v, \lambda, \delta_h, \delta_\ell)$ is unique whenever the number of periods is $N > 2$. We can apply Theorems 3 and 4 to PT preferences: the PTU $(u, v, \lambda, \delta_h, \delta_\ell)$ is curious if and only if v is convex, $\delta_h \geq 0$ and $\delta_\ell \leq 0$; it is information seeking if and only if v is convex and $\delta_h = \delta_\ell = 0$. Note that when $\delta_h \geq 0, \delta_\ell \leq 0$ the associated capacity, as defined above, is submodular. Thus, for PTUs tail-submodularity and submodularity coincide.

General RCU agents care about the timing of resolution of uncertainty but also about *how* uncertainty resolves. The focus of the following applications is on this latter effect; that is, the modality of information revelation rather than its timing.

4.1 Intermediate News: Savoring and Dread

Gul, et al. (2020) report the results of an experiment in which subjects choose among different modes of information disclosure. In the first treatment, a \$10 prize was randomly assigned to one of three boxes. Subjects, who did not know which box contained the prize, chose one and then decided the sequence in which the three boxes were opened. If the chosen box was opened first, the subject learned whether or not she won the prize immediately. If an unchosen box was opened first, and the contained the \$10, the subject would learn that she did not win; if the box was empty, the subject would learn that her chance of winning has gone up (from 1/3 to 1/2). After some delay, the second box would be opened and all uncertainty would be resolved. Thus, the first treatment confronts subjects with a choice between immediate resolution of uncertainty (if she opens the chosen box first) or a gradual resolution (if she opens an unchosen box first) that yields either decisive bad news or intermediate good news.

In the second treatment, a \$10 prize was randomly assigned to *two* of the three boxes. The remainder of the experiment was identical to the one described above. In this case, a subject who opens an unchosen box either learns that she won the prize (if the box is empty) or that her odds have gone down from 2/3 to 1/2 (if the box contains a prize). The second treatment confronts subjects with a choice between immediate resolution of uncertainty (if she opens the chosen box first) or a gradual resolution (if she opens an unchosen box first) that yields either decisive good news or intermediate bad news. The experiment finds that around 60% of subjects prefer opening an unchosen box in the first treatment while 60% of subjects prefer opening their own box in the second treatment.

Hence, many subjects react differently to disclosure rules that reveal intermediate good news and decisive bad news (as in the first treatment) than to disclosure rules that reveal decisive good news and intermediate bad news (as in the second treatment). In this section, we show that this differential response is consistent with PT preferences and relate it to the parameters of the PT model. Throughout the following analysis, we fix intermediate consumption and ignore it. We also assume, without loss of generality, that there are only two prizes and fix their utilities at 1 and 0 respectively. We slightly abuse notation and write x_t for the time t probability of winning the best prize along path x.

A path $x \in L^N$ is an intermediate good (bad) news path if $x_t \geq x_1$ ($x_t \leq x_1$) or $x_t = 0$ or ($x_t = 1$) for all t. Thus, along an intermediate good news path, the probability of winning is either no less than the initial probability of winning or 0. Let X^g (X^b) be the set of all intermediate good (bad) news paths. For any given probability of winning $a \in (0,1)$, let Π_a^g (Π_a^b) be the set of all RELs P such that $P(x) > 0$ implies $x \in X^g$ ($x \in X^b$), $x_1 = a$ and $x_N \in \{0,1\}$. The intersection of Π_a^g and Π_a^b consists of RELs that in each period t either resolve all uncertainty or else provide no information.

Recall the definition of $[x]_t$ from our definition of curiosity: $[x]_t := \{z \in L^N \mid z_s = x_s \text{ for all } s \leq t\}$; that is, $[x]_t$ is the set of all paths that agree with x until (and including) period t. The REL P *resolves later than* the REL $\hat{P}$ if the two assign the same probabilities to paths that do not contain a particular history and $\hat{P}$ resolves all uncertainty after that history. Formally, P resolves later than $\hat{P}$ if there is $t \geq 1$ and $x \in X^t$ such that (i) $P(y) = \hat{P}(y)$ for all $y \notin [x]_t$ and (ii) $y_{t+1} \in \{0, 1\}$ for all $y \in [x]_t$ with $\hat{P}(y) > 0$. Note that the preceding definition is *weak* in the sense that P resolves later than itself.

Consider, again, the Gul et al. experiment: in treatment 1, only one box contains the prize while in treatment 2, only one box contains no prize. Hence, treatment 1 offers a choice between an intermediate good news REL P that resolves later (in period 3 if the subject does not choose to open her own box first) or an intermediate good news REL $\hat{P}$ that resolves earlier (in period 2 if the subject chooses to open her own box first). In contrast, treatment 2 offers a choice between an intermediate bad news REL P that resolves later or an intermediate good news REL $\hat{P}$ that resolves earlier.

Part (i) of Proposition 2, below, establishes that a PTU agent with a linear v and $\delta_h, \delta_\ell \geq 0$ prefers intermediate good news to intermediate bad news. Part (ii) shows that a PTU agent with a concave v and $\delta_h \geq 0$ prefers RELs that resolve later; that is, savors intermediate good news. Finally, part (iii) shows that a PTU agent with a convex v and $\delta_\ell \geq 0$ does not like RELs that resolve later; that is, dreads intermediate bad news. A symmetric counterpart of Proposition 2 can be derived by reversing all inequalities, replacing concave with convex and convex with concave.

Proposition 2: *Let $V = (u, v, \lambda, \delta_h, \delta_\ell)$ and let $P^g, \hat{P}^g \in \Pi_a^g$ and $P^b, \hat{P}^b \in \Pi_a^b$ and assume that P^i resolves later than $\hat{P}^i$ for $i \in \{g, b\}$. Then,*

(i) $V(P^g) \geq V(P^b)$ if v is linear and $\delta_h, \delta_\ell \geq 0$;

(ii) $V(P^g) \geq V(\hat{P}^g)$ if v is concave and $\delta_h \geq 0$;

(iii) $V(\hat{P}^b) \geq V(P^b)$ if v is convex and $\delta_\ell \geq 0$.

To prove Proposition 2, we show that P^g's distribution of peaks stochastically dominates the corresponding distribution of $\hat{P}^g$ while both have the same distribution of troughs. It is easy to see that $\hat{P}_t$, (the time-t distribution of $\hat{P}$) is either a mean-preserving spread of P_t or the same as P_t for all $t \in N$. Since $\delta_h \geq 0$ and v is concave, part (ii) follows. A symmetric argument suffices to prove part (iii). Then, for part (i), take the unique $P \in \Pi_a^g \cap \Pi_a^b$ that reveals all information in period 2. Arguing as in part (ii), we get $V(P^g) \geq V(P)$. A symmetric argument with an appeal to part (iii) yields $V(P) \geq V(P^b)$ and therefore, $V(P^g) \geq V(P^b)$.

PTU agents with $\delta_h \geq 0$ savor intermediate good news. Conversely, individuals with $\delta_h \leq 0$ dislike getting their hopes up; that is, they dislike paths that exhibit a high intermediate probability of winning but terminate in a loss. A PTU with $\delta_\ell \geq 0$ describes someone who dreads intermediate bad news and hence would rather get it over with as soon as possible. By contrast, agents with $\delta_\ell \leq 0$ enjoy comebacks; that is, paths that end in 1 but have low interim odds of winning.

PT utility allows a nuanced preference for information that goes beyond a categorical preference for early or late resolution of uncertainty. The experiments in Gul, Natenzon, Ozbay and Pesendorfer (2020) reveal that a model that allows such nuance is indeed necessary to address subjects' behavior. PT utility is rich enough to address the substantial degree of heterogeneity displayed in subjects' attitude to information. The experiments reveal that only 16% of subjects are indifferent to the resolution of uncertainty across all treatments and only 30% exhibit a categorical preference for early (19%) or late (11%) resolution of uncertainty. For the majority of subjects (54%), the preference for delayed or immediate resolution depends on how information resolves. Of those, 57% savor intermediate good news; that is, they choose delayed resolution in the case where one box contains a prize.

4.2 Optimal Information Disclosure: Up or Out

Next, we consider the problem of optimal information disclosure. An agent is one of two types; with prior probability a she is a productive type and with prior probability $1-a$ she is an unproductive type. The principal must design an information structure that reveals the agent's productivity. We abstract from all incentive considerations and, instead, ask what process of discovering the agent's productivity would maximize her intrinsic value of information. The principal is free to design any information structure that reveals the agent's type over the course of N periods.

As in the previous section, we ignore intermediate consumption. Hence, x_t is the period-t probability that the agent is a high type. We set $v(u(1)) = 1$ and $v(u(0)) = 0$. Let Π_a be the set of RELs such that $P(x) > 0$ implies $x_1 = a, x_N \in \{0, 1\}$; that is, every path in the support of any REL in Π_a starts at a and reveals the agent's type by period N. The principal can choose any REL in Π_a.

Let $P^* \in \Pi_a$ be the (unique) REL described below:

$$P_t^*(x_t) > 0 \text{ implies } x_t \in \left\{0, a^{\frac{N-t}{N-1}}\right\}$$

The REL P^* delivers either decisive bad news ($x_t = 0$) or intermediate good news ($x_{t-1} < x_t < 1$) in any period $t < N$. Moreover, there is a single good-news path x such that

$$x_1 = a, x_2 = a^{\frac{N-2}{N-1}}, x_3 = a^{\frac{N-3}{N-1}}, \ldots, x_{N-1} = a^{\frac{1}{N-1}}, x_N = 1$$

Note that this path spaces the good news evenly: $x_t/x_{t+1} = a^{\frac{1}{N-1}}$ for all t. The flow of information under P^* resembles an up-or-out policy in which every period the agent either gets promoted or fired. The final promotion in period N reveals that the agent is a high productivity type.

Proposition 3, below, shows that P^* is the unique optimal REL when the agent has a PT utility with a linear v and positive parameters δ_h and δ_ℓ:

Proposition 3: *Let $V = (u, v, \lambda, \delta_h, \delta_\ell)$ be a PTU with linear v, $\delta_h > 0, \delta_\ell \geq 0$. Then,*

$$V(P^*) = a\left[1 - \delta_\ell(1-a) + \delta_h(N-1)\left(1-a^{\frac{1}{N-1}}\right)\right] > V(P)$$

for all $P \in \Pi_a, P \neq P^*$.

The agent's maximal payoff increases in the number of periods since a larger N enables the principal to spread out good news. As N gets arbitrarily large, the agent's optimal payoff converges to:

$$\lim_{N\to\infty} V(P^*) = a\left[1 - \delta_\ell(1-a) - \delta_h \ln a\right]$$

In Proposition 3, we assumed positive δ_h and non-negative δ_ℓ. For $\delta_\ell < 0, \delta_h \leq 0$, the optimal disclosure policy is the mirror image of the one characterized in Proposition 3 and features decisive good news and gradual bad news. If $\delta_h < 0, \delta_\ell > 0$, it is optimal to disclose all information in a single period. In the final case, $\delta_h > 0, \delta_\ell < 0$ optimal disclosure entails both intermediate good news and intermediate bad news.

4.3 Preference for Skewed Information

In this section, we analyze agents' preference for *skewed* information. As in the previous section, we fix intermediate consumption and ignore it. We also assume, without loss of generality, that there are only two prizes and fix their utilities at 1 and 0 respectively. We consider 3-period random evolving lotteries; the third period reveals whether or not the agent wins the prize, while the first period reveals no information beyond the prior $a \in (0, 1)$. Recall that Π_a is the set of all such RELs. In the second period, the agent receives binary information and, thus, either obtains good news in which case her probability of winning the prize increases to $g \geq a$ or bad news which reduces her probability of winning to $b \leq a$.

Then, by the martingale property, either $g = a = b$ or the probability of receiving good news is the c that solves $a = cg + (1 - c)b$, that is

$$c = \frac{a - b}{g - b}.$$

Hence, we can identify each such REL in Π_a with a unique (g, b) such that $g \geq a \geq b$. We say that the REL (g, b) is *positively skewed* if $g - a > a - b$; that is, $g + b > 2a$ and it is *negatively skewed* if $g + b < 2a$. Thus, the REL (g, b) is positively (negatively) skewed if the probability of good news in period 2 is less (greater) than 1/2.

Masatlioglu, Orhun and Raymond (2017) conduct experiments to determine whether agents prefer positively or negatively skewed intrinsic (non-instrumental) information. In their experiments, a majority of subjects preferred positively skewed information while a minority of subjects preferred negatively skewed information and few subjects were indifferent to skew. Their findings suggest that there is substantial heterogeneity in subjects' attitudes to how information resolves and that these preferences are strongly held; that is, subjects are not close to indifferent.

Given any REL $P = (g, b)$, such that $g \leq 2a \leq 1 + b$, we call $P^{\#} = (2a - b, 2a - g)$ the dual of P. Note that (i) $(P^{\#})^{\#} = P$; (ii) P is positively skewed if and only if $P^{\#}$ is negatively skewed; and (iii) if the probability of getting good news in period 2 is c for P, then the probability of getting good news in period 2 is $1 - c$ for $P^{\#}$.

Focusing on the comparison between P and its dual $P^{\#}$ enables us to isolate the decision maker's preference for skewed information. Proposition 4, below, relates preference for skewness to the parameters of PTU preferences. In particular, it establishes that a PTU agent with linear v prefers a positively skewed P to its negatively skewed dual if $\delta_h + \delta_\ell < 0$ and has the opposite preference if $\delta_h + \delta_\ell > 0$. More generally, PTU agents with linear v and $\delta_h + \delta_\ell < 0$ will tend to prefer positively skewed RELs over comparable negatively skewed ones.

Proposition 4: *Let $V = (u, v, \lambda, \delta_h, \delta_\ell)$, assume that v is linear and P positively skewed. Then, $(V(P) - V(P^{\#}))(\delta_h + \delta_\ell) < 0$ whenever $\delta_h + \delta_\ell \neq 0$.*

4.4 The Ostrich Effect

Agents exhibit the ostrich effect if they seek information after good news and reject information after bad news. The simplest example of this effect requires four periods and two signals — one available in period 2, the other in period 3 — that provide information about a binary terminal outcome. The outcome yields either utility 0 or 1. The initial (period 1) probability of the good outcome is $0 < a < 1$ and in period 4 all uncertainty is resolved. The signals available in periods 2 and 3 are binary and independent. Specifically, the probability that each signal matches the outcome is $d > 1/2$.

Suppose the agent must observe a signal in period 2. At the beginning of period 3, she is offered an additional signal, which she may accept or reject. Her decision to accept or reject the additional signal can depend on the realization of period 2's signal. We say that the agent *seeks (rejects) information after good (bad) news* if she strictly prefers to observe (not observe) an additional signal in period 3 following a positive (negative) signal in period 2. The following result characterizes the signal acquiring behavior of a PTU agent with linear v:

Proposition 5: *Let $V = (u, v, \lambda, \delta_h, \delta_\ell)$ with v linear. Then, (i) the agent seeks information after good news if and only if $\delta_h > 0$; and (ii) the agent rejects information after bad news if and only if $\delta_\ell > 0$.*

Hence, a PTU agent with linear v will exhibit the ostrich effect if and only if δ_h and δ_ℓ are strictly positive. To see why the ostrich effect emerges in this case, assume for simplicity that $a = 1/2$ and consider the agent's choice in period 3 after having received bad news in period 2. A second bad signal would yield a new minimal probability of outcome 1 while good news after bad news would merely restore the original prior. Since $\delta_\ell > 0$, additional information has a potential cost but no benefit in this case. By contrast, additional good news after initial good news yields a new maximum probability of outcome 1 while bad news after good news restores the prior and does not yield a new minimum. Since $\delta_h > 0$, there is a potential benefit without cost to obtaining information. So, if the agent receives information in period 2, she will choose to get more information in period 3, if and only if the period 2 information yields good news.

In our model, the agent chooses an REL at time 0 and makes no further decisions in subsequent periods. Any dynamic choice problem, described as a decision tree, can be studied in our setting after enumerating all of the possible RELs that emerge by considering the various combinations of choices at different decision nodes. Such an approach would presume dynamic consistency; i.e., that the agent's behavior at any given node is in line with her intentions at the outset. In an online appendix, we provide a recursive formulation of our model and identify the minimal states spaces needed for dynamically consistent behavior.

In that appendix, we formulate a version of the problem above for arbitrary $N \geq 4$ and show that in this version the ostrich effect holds more generally, even if v is not linear provided δ_h and δ_ℓ are both greater than 0. For example, we show that, irrespective of the curvature of v, and agent with $\delta_h, \delta_\ell > 0$ is more inclined to obtain information if the current prize lottery is closer to the past peak and further from the past trough experiences.

5. Conclusions

In this paper, we introduce risk consumption utility to address observed regularities in the demand for non-instrumental information. We show that a simple special case, peak-trough utility, is rich enough to address a wide range of evidence. Specifically, we show how the parameters of peak-trough utility relate to savoring (and dread) of intermediate good or bad news, preference for skewed information, and the ostrich effect.

For expositional clarity, we specialize to a linear utility index v in some of the applications above. However, it is clear that a non-linear index v is needed to address some applications, such as Masatlioglu, Orhun and Raymond's (2017) finding that many subjects are not indifferent as to whether uncertainty resolves fully in the first or the second periods. For peak-trough utility, the ranking of RELs that fully resolve the same uncertainty at distinct dates is entirely determined by the utility index v. Hence, these relatively simple experiments can be used to identify or calibrate v.

In the online appendix, we provide a version of the ostrich effect that does not depend on the curvature of v: we show that if the current prize lottery is closer to the past peak and further from past trough experiences, the agent is more inclined to obtain information. The curvature of v also affects preference for skewed information with peak-trough utility. While a characterization of preference for skewed information with non-linear v is difficult, allowing for such v would enable us to model individuals who simultaneously exhibit a preference for positively skewed information and the ostrich effect.

Our model predicts particular correlation patterns across different applications. For example, individuals with a linear v who prefer intermediate good news but dislike intermediate bad news should also exhibit a preference for negative skewed information and be prone to the ostrich effect. Evidence regarding correlations of behavior across those (and other) domains can identify the parameters of peak-trough utilities and provide tests for the theory.

6. Appendix: Proofs

6.1 Proof of Theorem 1

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

Lemma 1: *There are continuous, linear functions $u : L \to [0,1]$, $u_1 : \Delta(A) \to [0,1]$, and $u_2 : \Delta(B) \to [0,1]$ such that (i) $P_{(\alpha,\beta)} \succcurlyeq P_{(\alpha',\beta')}$ if and only if $u(\alpha,\beta) \geq u(\alpha',\beta')$, (ii) $u(\alpha,\beta) = u_1(\alpha) + u_2(\beta)$, and (iii) u is onto.*

Proof: The restriction of $\succcurlyeq$ to $\{P_{(\alpha,\beta)} \in \Pi \,|\, (\alpha,\beta) \in L\}$ induces a complete and transitive preference $\succcurlyeq_o$ on L. Since $d_p(P_{(\alpha,\beta)}, P_{(\alpha',\beta')}) = d(x_{(\alpha,\beta)}, x_{(\alpha',\beta')})$, Axiom 3 implies that $\succcurlyeq_o$ is continuous. Axiom 4 states that $\succcurlyeq_o$ satisfies independence on the mixture space L. Hence, there exists a linear function u that represents $\succcurlyeq_o$.

Since L is finite dimensional and $\succcurlyeq$ is not degenerate, we can assume, without loss of generality, that there is $(\overline{\alpha},\overline{\beta}) \in \arg\max_L u(\cdot)$ and $(\underline{\alpha},\underline{\beta}) \in \arg\min_L u(\cdot)$ such that $u(\overline{\alpha},\overline{\beta}) = 1$ and $u(\underline{\alpha},\underline{\beta}) = 0$. For any (α,β) and $(\alpha',\beta') \in L$, the linearity of u implies

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

Hence,

$$u(\alpha,\beta) + u(\alpha',\beta') = u(\alpha,\beta') + u(\alpha',\beta). \tag{1}$$

Then, let $u_1(\alpha) = u(\alpha,\underline{\beta})$ and let $u_2(\beta) = u(\underline{\alpha},\beta)$. Equation (1) implies $u(\alpha,\beta) = u_1(\alpha) + u_2(\beta)$ as desired. □

Since the time horizon is finite, the set I of all ordered partitions $\iota = (S_1, \ldots, S_n)$ is also finite. Therefore every $P \in \Pi$ can be uniquely decomposed as a finite sum $P = \sum_\iota (PX_\iota) P_{X_\iota}$.

Lemma 2: *If $P, P' \in \Pi$ with $PX_\iota = P'X_\iota$ for all $\iota \in I$, $a, b \in [0,1]$, and $aP_{X_\iota} + (1-a)P_{(\alpha,\beta)}$ dominates $bP'_{X_\iota} + (1-b)P_{(\alpha',\beta')}$ whenever $PX_\iota > 0$, then $aP + (1-a)P_{(\alpha,\beta)} \succcurlyeq bP' + (1-b)P_{(\alpha',\beta')}$.*

Proof: For each ordered partition $\iota = (S_1^\iota, \ldots, S_{|\iota|}^\iota) \in I$, define the evolving lotteries $x^\iota, y^\iota \in L^N$ by

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

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

For $n \geq 1$, let

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

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

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

Since $aP_{X_\iota} + (1-a)P_{(\alpha,\beta)}$ dominates $bP'_{X_\iota} + (1-b)P_{(\alpha',\beta')}$ whenever $PX_\iota > 0$, Axiom 2 implies $aP + (1-a)P_{(\alpha,\beta)}$ dominates $bP' + (1-b)P_{(\alpha',\beta')}$.

Let P^x be a REL such that $P^x(x^\iota) = PX_\iota$ for each $\iota \in I$; let P^y be a REL such that $P^y(y^\iota) = PX_\iota$ for each $\iota \in I$. For each $n \geq 1$, $(1/n)P^x + (1-1/n)[aP + (1-a)P_{(\alpha,\beta)}]$ strictly dominates $(1/n)P^y + (1-1/n)[aP' + (1-a)P_{(\alpha',\beta')}]$.

Now for each $n \geq 1$, let P^{wn} and P^{zn} be RELs with $P^{wn}(w^{\iota n}) = P^{zn}(z^{\iota n}) = PX_\iota$ for each $\iota \in I$. Thus, for each fixed $n \geq 1$ there exists an integer $M(n)$ such that, for every $k \geq M(n)$, $(1/n)P^x + (1-1/n)[aP + (1-a)P^{wk}]$ strictly rank-dominates $(1/n)P^y + (1-1/n)[aP' + (1-a)P^{zk}]$. Hence Axiom 5 implies for each $n \geq 1$,

$$(1/n)P^x + (1-1/n)[aP + (1-a)P^{wM(n)}] \succ (1/n)P^x + (1-1/n)[bP' + (1-b)P^{zM(n)}].$$

Moreover,

$$d_p\left((1/n)P^x + (1-1/n)\left[aP + (1-a)P^{wM(n)}\right], aP + (1-a)P_{(\alpha,\beta)}\right) \to 0$$

and

$$d_p\left((1/n)P^x + (1-1/n)\left[bP' + (1-b)P^{zM(n)}\right], bP + (1-b)P_{(\alpha',\beta')}\right) \to 0$$

and Axiom 3 implies $aP + (1-a)P_{(\alpha,\beta)} \succcurlyeq bP' + (1-b)P_{(\alpha',\beta')}$ as desired. $\square$

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

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 $\succcurlyeq$. Axiom 3 also ensures that $\hat{V}$ is continuous. Lemma 2 implies that $P_{(\overline{\alpha},\overline{\beta})} \succcurlyeq P$ and $P \succcurlyeq P_{(\underline{\alpha},\underline{\beta})}$. It follows that the range of $\hat{V}$ is a compact interval. Axiom 5 implies that $P_{(\overline{\alpha},\overline{\beta})} \succ P_{(\underline{\alpha},\underline{\beta})}$. Then, a suitable affine transformation of $\hat{V}$ yields the desired V. $\square$

For $r \in [0,1]$, define $v(r) = V(P_{(\alpha,\beta)})$ for (α,β) such that $u(\alpha,\beta) = r$. Lemmas 1 and 3 ensure that v is a well-defined element of Λ. We call P an $\iota-$REL if $PX_\iota = 1$. Let Π_ι be the set of all $\iota-$RELs. Define the mapping $f : \Delta(L^N) \to [0,1]^N$ as follows:

$$f(P) = (E_P[v(u(x_1))], \ldots, E_P[v(u(x_N))]) .$$

Let $\bar{r} := v(u(\overline{\alpha}, \underline{\beta}))$ and note that $\bar{r} > 0$ by non-degeneracy. Fix any ordered partition $\iota = (S_1^\iota, \ldots, S_{|\iota|}^\iota)$. Define an $|\iota| \times |\iota|$ matrix (a_{ij}) as follows:

$$a_{ij} = \begin{cases} (|\iota| - j)\bar{r}/(|\iota| + 1) & \text{if } j > i \\ (|\iota| - j + 1)\bar{r}/(|\iota| + 1) & \text{if } j \leq i. \end{cases}$$

By invoking elementary properties of systems of linear equations, we can verify that the matrix (a_{ij}) has a non-zero determinant. By Lemma 1, for each i and j there exists a consumption lottery $\alpha^{ij} \in \Delta(A)$ such that $v(u(\alpha^{ij}, \underline{\beta})) = a_{ij}$. Then, define the evolving lotteries $x^1, \ldots, x^{|\iota|}$ as follows: $x_t^i = (\alpha^{ij}, \underline{\beta})$ whenever $t \in S_j^\iota$. Let $P^i(x^i) = 1$ for each i. Consider the following system of linear equations:

$$\sum_{j=1}^{|\iota|} a_{ij} y_j = V(P^i) \text{ for } i = 1, \ldots, |\iota| \tag{2}$$

Let $\eta_\iota S_j^\iota := y_j$ where $(y_1, \ldots, y_{|\iota|})$ is the solution to the system of equations (2).

We say that $P \in \Pi$ is *normal* if

$$V(P) = \sum_{\iota \in I} PX_\iota \sum_{j=1}^{|\iota|} E_{P_{X_\iota}}[v \circ u(x_{g(j)})]\, \eta_\iota S_j^\iota,$$

where $g(j) \in S_j^\iota$ for each j. In particular, $P \in \Pi_\iota$ is normal if and only if

$$V(P) = \sum_{j=1}^{|\iota|} E_P[v \circ u(x_{g(j)})]\, \eta_\iota S_j^\iota,$$

where again $g(j) \in S_j^\iota$ for each j. Hence, each P^i above is normal by construction.

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

Proof: For each $i = 1, \ldots, |\iota|$ let $\phi_i = f(P)_t$ for an arbitrary $t \in S_i^\iota$. Since $P \in \Pi_\iota$, each ϕ_i is well defined. There exist $r_i \in I\!R$ such that $(\phi_1, \ldots, \phi_{|\iota|}) = \sum_{i=1}^{|\iota|} r_i(a_{i1}, \ldots, a_{i|\iota|})$. Define $r_{|\iota|+1} = 1 - \sum_{i=1}^{|\iota|} r_i$; define $a_{|\iota|+1,j} = 0$ for $j = 1, \ldots, |\iota|$; let the path $x^{|\iota|+1} = x_{(\underline{\alpha},\underline{\beta})}$; let $P^{|\iota|+1} = P_{(\underline{\alpha},\underline{\beta})}$; and, finally, let $r_i^+ := \max\{r_i, 0\}$, and $r_i^- := \max\{-r_i, 0\}$ be the positive and negative parts of r_i for each $i = 1, \ldots, |\iota| + 1$.

Let the constant $c := \sum_{i=1}^{|\iota|+1} r_i^+ = 1 + \sum_{i=1}^{|\iota|+1} r_i^- \geq 1$. Then, the RELs

$$\frac{1}{c}P + \frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^- P^i \text{ and } \frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^+ P^i$$

both belong to Π_ι with

$$f\left(\frac{1}{c}P + \frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^- P^i\right) = f\left(\frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^+ P^i\right).$$

Lemma 2 then implies

$$V\left(\frac{1}{c}P + \frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^- P^i\right) = V\left(\frac{1}{c}\sum_{i=1}^{|\iota|+1} r_i^+ P^i\right).$$

Using the linearity of V, we can rearrange the terms again to obtain

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

as desired. □

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

Proof: We again let $\bar{r} = v(u(\overline{\alpha}, \underline{\beta}))$ and note that $\bar{r} > 0$ by non-degeneracy. Every $P \in \Pi$ has a unique decomposition as a finite sum $P = \sum_\iota (PX_\iota)P_{X_\iota}$. The linearity of V implies

$$V\left(\bar{r}P + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}\right) = \bar{r}V(P)$$

Since

$$f\left(\bar{r}P_{X_\iota} + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}\right) \in [0, \bar{r}]^N,$$

there exists $P^{0\iota} \in \Pi_\iota$ with $f(P^{0\iota}) = f\left(\bar{r}P_{X_\iota} + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}\right)$. Then, by Lemma 2,

$$V\left(\sum_\iota (PX_\iota)P^{0\iota}\right) = V\left(\bar{r}P + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}\right)$$

By linearity and Lemma 4,

$$\begin{aligned}
V(P) &= \bar{r}^{-1}V(\bar{r}P + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}) \\
&= \bar{r}^{-1}V\left(\textstyle\sum_\iota (PX_\iota)P^{0\iota}\right) \\
&= \bar{r}^{-1}\textstyle\sum_\iota (PX_\iota)\sum_{j=1}^{|\iota|} E_{P^{0\iota}}[v \circ u(x_{g(j)})]\eta_\iota S_j^\iota \\
&= \bar{r}^{-1}\textstyle\sum_\iota (PX_\iota)\sum_{j=1}^{|\iota|} f(P^{0\iota})_{g(j)}\eta_\iota S_j^\iota \\
&= \bar{r}^{-1}\textstyle\sum_\iota (PX_\iota)\sum_{j=1}^{|\iota|} f\left(\bar{r}P_{X_\iota} + (1-\bar{r})P_{(\underline{\alpha},\underline{\beta})}\right)_{g(j)}\eta_\iota S_j^\iota \\
&= \bar{r}^{-1}\textstyle\sum_\iota (PX_\iota)\sum_{j=1}^{|\iota|} \bar{r}E_{P_{X_\iota}}[v \circ u(x_{g(j)})]\eta_\iota S_j^\iota
\end{aligned}$$

where $g(j) \in S_j^\iota$ for each j, thus showing that P is normal. □

To construct the capacity η in the representation, we let $\eta\emptyset = 0$; for $S \in \mathcal{S}$ with $\emptyset \neq S \neq N$, we let $\eta S = \eta_{\iota^*}S$ for $\iota^* = (S, N\backslash S)$; and we let $\eta N = \eta_{\iota^*}N$ for $\iota^* = N$.

Lemma 6: *$S = \bigcup_{j \le i} S_j^\iota$ implies $\eta S = \sum_{j=1}^i \eta_\iota S_j^\iota$.*

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

$$x_t^n = \begin{cases} (1-2^{-n})(\overline{\alpha}, \underline{\beta}) + 2^{-n}x_t & \text{if } t \in S \\ (1-2^{-n})(\underline{\alpha}, \overline{\beta}) + 2^{-n}x_t & \text{if } t \in N\backslash S \end{cases}$$
$$x_t^* = \begin{cases} (\overline{\alpha}, \underline{\beta}) & \text{if } t \in S \\ (\underline{\alpha}, \overline{\beta}) & \text{if } t \in N\backslash S \end{cases}$$

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

To see that η is indeed a capacity, note that Lemmas 2 and 4 ensure that $\eta_\iota S_j^\iota \ge 0$ for all ι and all j. Therefore, $\eta S \ge 0$ and, by Lemma 6, $\eta T \ge \eta S$ for $S \subset T$. Finally since $V\left(P_{(\overline{\alpha},\overline{\beta})}\right) = 1$, it follows that $\eta N = 1$.

Let $P \in \Pi$; for each path x, let $\iota(x) \in I$ be the ordered partition such that $x \in X_{\iota(x)}$. Lemmas 5 and 6 and the definition of η imply that

$$\begin{aligned}
V(P) &= \sum_x P(x) \sum_{j=1}^{|\iota(x)|} v(u(x_{g(j)}))\ \eta_{\iota(x)} S_j^{\iota(x)} \\
&= \sum_x P(x) \sum_{j=1}^{|\iota(x)|} v(u(x_{g(j)}))\ \left[\eta \bigcup_{i=1}^{j} S_i^{\iota(x)} - \eta \bigcup_{i=1}^{j-1} S_i^{\iota(x)}\right] \\
&= E_P \left[\int v(u(x_t))\, d\eta\right]
\end{aligned}$$

where $g(j) \in S_j^{\iota(x)}$ for each j, which completes the proof of the 'only if' statement.

The proof of the 'if' statement is straightforward. To show uniqueness, let (u, v, η) and $(\hat{u}, \hat{v}, \hat{\eta})$ be two RCU representations of the non-degenerate $\succcurlyeq$. Pick lotteries $(\overline{\alpha}, \overline{\beta})$ and $(\underline{\alpha}, \underline{\beta})$ such that $u(\overline{\alpha}, \overline{\beta}) = 1$ and $u(\underline{\alpha}, \underline{\beta}) = 0$. Then, we must have $\hat{u}(\overline{\alpha}, \overline{\beta}) = 1$ and $\hat{u}(\underline{\alpha}, \underline{\beta}) = 0$. Since $u, \hat{u}$ represent the same preference relation on L, agree at two distinct points $(\overline{\alpha}, \overline{\beta})$ and $(\underline{\alpha}, \underline{\beta})$, and are both linear, we must have $u = \hat{u}$. Similarly, the utility index $v \circ u = v \circ \hat{u}$ and $\hat{v} \circ u$ represent the same linear preference over $\Delta(L)$ and agree at two distinct degenerate second order-lotteries p, p'. 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 two lotteries (α, β) and (α', β) such that $u(\alpha, \beta) < u(\alpha', \beta)$. Since $\succcurlyeq$ is non-degenerate, such lotteries must exist. Let x be an path with $x_t = (\alpha, \beta)$ if $t \in S$ and $x_t = (\alpha', \beta)$ otherwise. Let P be a REL with $P(x) = 1$. Then the representation yields

$$(1 - \eta S) v(u(\alpha, \beta)) + \eta S\, v(u(\alpha', \beta)) = V(P) = \hat{V}(P) = (1 - \hat{\eta} S) v(u(\alpha, \beta)) + \hat{\eta} S\, v(u(\alpha', \beta))$$

thus $\eta S = \hat{\eta} S$ as desired. $\square$

6.2 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, $\eta(S \cup T) = \eta S + \eta T$ for $S, T \in \mathcal{S}$ disjoint. Let

$$\begin{aligned}
x_t &= \begin{cases} (\overline{\alpha}, \underline{\beta}) & \text{if } t \in S \\ (\underline{\alpha}, \overline{\beta}) & \text{if } t \notin S \end{cases} \\
y_t &= \begin{cases} (\overline{\alpha}, \underline{\beta}) & \text{if } t \in T \\ (\underline{\alpha}, \overline{\beta}) & \text{if } t \notin T \end{cases} \\
z_t^n &= \begin{cases} ((2^{-n}\underline{\alpha} + (1 - 2^{-n})\overline{\alpha}, \underline{\beta}) & \text{if } t \in S \cup T \\ (\underline{\alpha}, \underline{\beta}) & \text{if } t \notin S \cup T \end{cases}
\end{aligned}$$

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

6.3 Proof of Theorem 2

Let η be a tail-submodular capacity, v be convex and assume P' is simply more informative than P''. Hence, either (i) $P'' = P_{\bar{x}(P')}$ or (ii) there is $t \geq 1$ and $x \in X^t$ such that $P'[x]_t = P''(x)$ and $P'(z) = P''(z)$ for all $z \notin [x]_t$. It is enough for us to prove $V(P') \geq V(P'')$ assuming either (i′) $P'' = P_{\bar{x}(P')}$ and the support of P' consists of (multiple) constant paths or (ii′) $t \geq 1$, $x \in X^t$, $P'[x]_t = P''(x)$, $P'(z) = P''(z)$ for all $z \notin [x]_t$ and ($y \in [x]_t$, $P'(y) > 0$ implies $y \in X^{t+1}$). This is sufficient because once $V(P') \geq V(P'')$ is established for the cases (i′) and (ii′), a simple inductive argument ensures that $V(P') \geq V(P'')$ holds whenever (i) or (ii) is satisfied.

Choose β^1, β^0 such that $u(\alpha, \beta^1) \geq u(\alpha, \beta) \geq u(\alpha, \beta^0)$ for all β. Non-degeneracy ensures that $u(\alpha, \beta^1) > u(\alpha, \beta^0)$. Let $\beta^a = a\beta^1 + (1-a)\beta^0$ for all $a \in [0,1]$. We will also assume that for all y in the support of P' or P'', $y_s = (\alpha, \beta^a)$ for all $s \in N$. Proving $V(P') \geq V(P'')$ for this case will be enough since for any β, we can choose a such that $u(\alpha, \beta^a) = u(\alpha, \beta)$.

If (i′) holds, then $V(P') \geq V(P'')$ follows from the convexity of v. Assume, (ii′) holds and $P' \neq P''$; otherwise there is nothing to prove. Then, choose $y, z \in [x_t]$, such that $P'(y) \cdot P'(z) > 0$, $y_s = (\alpha, \beta^a)$, $y_s = (\alpha, \beta^b)$ for all $s > t$, where $a > c > b$ and $x_s = (\alpha, \beta^c)$ for all $s \geq t$. The martingale property ensures that such y, z exist.

Let y^ϵ be the path obtained forum y by replacing β^a with $\beta^{a-\epsilon}$ in y_s for all $s \geq t+1$ and z^ϵ be the path obtained by replacing β^b with $\beta^{b+\lambda\epsilon}$ in y_s for all $s \geq t+1$, where $\lambda = \frac{P'(y)}{P'(z)}$. Then, let P^ϵ be the REL obtained by replacing the path y with y^ϵ and z with z^ϵ in P'. Note that for $\epsilon < \min\{a-c, c-b\}$, P^ϵ is indeed a REL. Define $\Delta(\epsilon) = \frac{V(P') - V(P^\epsilon)}{\epsilon}$. Since v is convex, its right and left derivatives exist. Hence,

$$\lim_{\epsilon \to 0} \Delta(\epsilon) = d^* \cdot [\eta(C_1 \cup S^+_{t+1}) - \eta S^+_{t+1}]P'(y) + d \cdot [\eta(C_2 \cup S^+_{t+1}) - \eta S^+_{t+1}]P'(y)$$

where d^* is the left-derivative of v at $u(\alpha, \beta^a)$ multiplied by $u(\alpha, \beta^1) - u(\alpha, \beta^0)$, d is the right-derivative of v at $u(\alpha, \beta^b)$ multiplied by $u(\alpha, \beta^1) - u(\alpha, \beta^0)$ and

$$C_1 = \{s \leq t \,|\, u(x_s) \geq u(\alpha, \beta^a)\}$$

$$C_2 = \{s \leq t \,|\, u(x_s) > u(\alpha, \beta^b)\}$$

Since, v is increasing, convex and $u(\alpha, \beta^a) > u(\alpha, \beta^b)$, we have $d^* \geq d \geq 0$. Hence, $\lim \Delta(\epsilon) \geq 0$ whenever

$$\eta(C_1 \cup S_{t+1}^+) + \eta C_2 \geq \eta C_2 \cup S_{t+1}^+ + \eta C_1$$

Let $S = C_1 \cup S_{t+1}^+$ and $T = C_2$. Since $C_1 \subset C_2$, we have $S \cup T = C_1 \cup C_2 \cup S_{t+1}^+ = C_2 \cup S_{t+1}^+$. Also, since $C_2 \cap S_{t+1}^+ = \emptyset$, $S \cap T = C_1$. Then, we can rewrite the equation above as

$$\eta S + \eta T \geq \eta(S \cup T) + \eta(S \cap T)$$

Finally, note that $(S \cap S_t^-) \cup \{t\} \subset T \subset S_t^-$ and $S_{t+1}^+ \subset S \subset N \backslash \{t\}$. Since η is tail-submodular, we conclude that $\lim \Delta(\epsilon) \geq 0$.

Hence, $V(P') \geq V(P^\epsilon)$. Then, setting $\epsilon = \min\{a - c, c - b\}$ yields a REL P^1 such that $V(P') \geq V(P^1)$, $P^1[x]_t = P'(x)$, $P'(z) = P'(z)$ for all $z \notin [x]_t$ and the cardinality of the set $\{x' \in [x]_t \backslash \{x\} \mid P^1(x') > 0\}$ is at least one smaller than the cardinality of the set $\{x' \in [x]_t \backslash \{x\} \mid P'(x') > 0\}$. Then, repeating the argument as needed yields as sequence of RELs P^j for $j = 0, 1, \ldots, m$ such that for all j, $P^0 = P'$, $P^m = P''$ and $V(P^j) \geq V(P^{j+1})$ for all j, proving sufficiency.

To prove necessity, let $P' = \lambda P_{(\alpha, \beta^a)} + (1 - \lambda) P_{(\alpha, \beta^b)}$ and $P'' = P_{(\alpha, \beta^{\lambda a + (1-\lambda) b})}$. Hence, $P'' = P_{\bar{x}(P')}$ and therefore $V(P') \geq V(P'')$ since V is curious. Then, by our normalization $u_2(\beta^0) = 0$ and hence,

$$\lambda v(u_1(\alpha) + a u_2(\beta^1)) + (1 - \lambda) v(u_1(\alpha) + b u_2(\beta^1)) \geq v(u_1(\alpha) + [\lambda a + (1 - \lambda) b] u_2(\beta^1))$$

By varying α, we can derive $\lambda v(a) + (1 - \lambda) v(b) \geq v(\lambda a + (1 - \lambda) b)$ for all $a, b, \lambda \in [0, 1]$.

To prove that η must be tail-submodular, let $S, T \subset N$ be two sets satisfying the conditions in the definition of tail-submodularity. Choose $c \in (0, 1)$ such that v is differentiable at $u(\alpha, \beta^c)$ and its derivative is strictly positive. Since v strictly increasing and convex, this can be done. Define the path x as follows:

$$x_s = \begin{cases} (\alpha, \beta^1) & \text{for } s \in S \cap T \\ (\alpha, \beta^c) & \text{for } s \in (S \backslash T) \cup (T \backslash S) \\ (\alpha, \beta^0) & \text{otherwise} \end{cases}$$

Then, define y^ϵ as follows:

$$y_s^\epsilon = \begin{cases} (\alpha, \beta^{c+\epsilon}) & \text{for } s \in S_{t+1}^+ \\ x_s & \text{otherwise} \end{cases}$$

and let $z^\epsilon = y^{-\epsilon}$.

Let P' be any REL such that $[x]_t \cap \{x' \mid P'(x') > 0\} = \{x\}$. Clearly, such a REL can be constructed. Then, let P^ϵ be the unique REL derived from P' by replacing the path x with the two paths y^ϵ, z^ϵ and assigning probability $P'(x)/2$ to each. Define $\hat{\Delta}(\epsilon) = \frac{V(P^\epsilon)-V(P')}{\epsilon}$. Then,

$$\lim_{\epsilon\to 0} \hat{\Delta} = \lambda[\eta S - \eta(S \cap T) - \eta(S \cup T) + \eta T]$$

where $\lambda = [u(\alpha, \beta^1) - u(\alpha, \beta^0)]v'(u(\alpha, \beta^c))P'(x)/2$. Since V is curious, we must have $V(P^\epsilon) \geq V(P')$ for all ϵ and hence, $\lim \hat{\Delta} \geq 0$. Therefore, $\eta S + \eta T \geq \eta(S \cup T) + \eta(S \cap T)$ as desired.$\square$

6.4 Proof of Theorem 3

Suppose $V = (u, v, \eta)$ is information seeking. It follows from Theorem 2 that v is convex. Next, we will show that η must be additive. A set $S \subset N$ is an interval if $S = \{s, s+1, \ldots, t-1, t\}$ for some s and t. A partition $\mathbf{a} = \{A_1, \ldots, A_m\} \subset 2^N \backslash \{\emptyset\}$ of N is an interval partition if each A_i is a interval. We number the intervals so that $s \in A_i, t \in A_{i+1}$ implies $s < t$. We say that $\mathbf{a}$ is an m-interval partition if it contains exactly m intervals. A set T is $\mathbf{a}$-measurable if $\mathbf{a}$ is an interval partition such that $A_i \cap T \neq \emptyset$ implies $A_i \subset T$. A set T has complexity m if it is $\mathbf{a}$-measurable for some m-interval partition $\mathbf{a}$ and is not $\mathbf{b}$-measurable for any m'-interval partition $\mathbf{b}$ such that $m' < m$. For any interval partition $\mathbf{a}$ and $\mathbf{a}$-measurable T, let $\mathbf{a}(T) = \{A_i \in \mathbf{a} \mid A_i \subset T\}$. Hence,

$$T = \bigcup_{A_i \in \mathbf{a}(T)} A_i$$

Note that for any m-interval partition $\mathbf{a} = \{A_1, \ldots, A_m\}$, there are only two $\mathbf{a}$-measurable sets that have complexity m: if m is odd, these sets are $A_1 \cup A_3 \cup \cdots \cup A_m$ and $A_2 \cup A_4 \cup \cdots \cup A_{m-1}$; if m is even, these sets are $A_1 \cup A_3 \cup \cdots \cup A_{m-1}$ and $A_2 \cup A_4 \cup \cdots \cup A_m$. All other $\mathbf{a}$-measurable sets have less complexity.

Choose β^3, β^0 such that $u_2(\beta^3) > u_2(\beta^0)$ and $\beta^2 = \frac{2}{3}\beta^1 + \frac{1}{3}\beta^0$, $\beta^1 = \frac{1}{3}\beta^1 + \frac{2}{3}\beta^0$. For any interval partition $\mathbf{a}$, with cardinality m, we define the REL $P^{\mathbf{a}}$ which corresponds to the following stochastic process: in every period in the first interval, $A_1 \in \mathbf{a}$, $P^{\mathbf{a}}$ delivers (α, β^1) for sure. After the last period in the first interval, there is either good news or bad news, each with probability 1/2. If the news is bad, the agent gets (α, β^0) in every subsequent period, if the news is good, the agent gets (α, β^2) in every period $s \in A_2$ and after the last $t \in A_2$, again there is good news with probability 1/2 and bad news with probability 1/2; if the news is good, the agent gets (α, β^3) in every subsequent period. If the news is bad, the agent gets (α, β^1) in every period $t \in A_3$ and after the last period in A_3, there is again good news or bad news and so on.

Formally, we define the paths $x^1, \ldots, x^m$ as follows: for any odd $j < m$,

$$x_t^j = \begin{cases} (\alpha, \beta^1) & \text{if } t \in A_i \text{ for some odd } i \leq j \\ (\alpha, \beta^2) & \text{if } t \in A_i \text{ for some even } i \leq j \\ (\alpha, \beta^0) & \text{otherwise} \end{cases}$$

for any even $j < m$,

$$x_t^j = \begin{cases} (\alpha, \beta^1) & \text{if } t \in A_i \text{ for some odd } i \leq j \\ (\alpha, \beta^2) & \text{if } t \in A_i \text{ for some even } i \leq j \\ (\alpha, \beta^3) & \text{otherwise} \end{cases}$$

Finally, let

$$x_t^m = \begin{cases} (\alpha, \beta^1) & \text{if } t \in A_i \text{ for } i \text{ odd} \\ (\alpha, \beta^2) & \text{if } t \in A_i \text{ for } i \text{ even} \end{cases}$$

Let $P^{\mathbf{a}}(x^j) = 2^{-j}$ for all $j < m$ and $P^{\mathbf{a}}(x^m) = 2^{1-m}$. Let $P^1 = \frac{1}{3}P_{(\alpha,\beta^3)} + \frac{2}{3}P_{(\alpha,\beta^0)}$ and note that P^1 is more informative than $P^{\mathbf{a}}$ which, in turn, is more informative than $P_{(\alpha,\beta^1)}$.

Assume for now that the restriction of v to the interval $[u(\alpha, \beta^0), u(\alpha, \beta^3)]$ is an affine function. Then, (after taking an affine transformation of v) we can assume without loss of generality that $v(u(\alpha, \beta^j)) = j$ for all j. If V is information seeking then

$$V(P^1) \geq V(P^{\mathbf{a}}) \geq V(P_{(\alpha,\beta^1)}) = V(P^1) = 1$$

Hence, $V(P^{\mathbf{a}}) = 1$ for all $\mathbf{a}$.

Next, for any interval partition $\mathbf{a}$ with cardinality m, we define the REL $Q^{\mathbf{a}}$. Let $y^1, \ldots, y^m$ be the following paths for any odd $j < m$,

$$y_t^j = \begin{cases} (\alpha, \beta^2) & \text{if } t \in A_i \text{ for some odd } i \leq j \\ (\alpha, \beta^1) & \text{if } t \in A_i \text{ for some even } i \leq j \\ (\alpha, \beta^0) & \text{otherwise} \end{cases}$$

for any even $j < m$,

$$y_t^j = \begin{cases} (\alpha, \beta^2) & \text{if } t \in A_i \text{ for some odd } i \leq j \\ (\alpha, \beta^1) & \text{if } t \in A_i \text{ for some even } i \leq j \\ (\alpha, \beta^3) & \text{otherwise} \end{cases}$$

Finally, let

$$y_t^m = \begin{cases} (\alpha, \beta^2) & \text{if } t \in A_i \text{ for } i \text{ odd} \\ (\alpha, \beta^1) & \text{if } t \in A_i \text{ for } i \text{ even} \end{cases}$$

Let $Q^{\mathbf{a}}(x^j) = 2^{-j}$ for all $j < m$ and $Q^{\mathbf{a}}(x^m) = 2^{1-m}$. Let $Q^1 = \frac{2}{3}Q_{(\alpha,\beta^3)} + \frac{1}{3}Q_{(\alpha,\beta^0)}$ and note that Q^1 is more informative than $Q^{\mathbf{a}}$ which, in turn, is more informative than $Q_{(\alpha,\beta^2)}$. Therefore, if the V has a strong preference for information, then $V(Q^1) \geq V(Q^{\mathbf{a}}) \geq V(Q_{(\alpha,\beta^2)}) = V(Q^1) = 2$. Hence, $V(Q^{\mathbf{a}}) = 2$ for all $\mathbf{a}$.

Let $V = (\eta, u, v)$. We will prove, by induction on m, that for any m-interval-partition $\mathbf{a}$ and $\mathbf{a}$-measurable T,

$$\eta T = \sum_{A_i \in \mathbf{a}(T)} \eta A_i \tag{3}$$

This will establish that η is additive. Let $\mathbf{a} = \{A_1, A_2\}$. Then, $V(P^{\mathbf{a}}) = 1$ implies

$$\eta A_2 + \frac{1}{2}\eta A_1 + \frac{1}{2}\eta A_1 = 1$$

that is, $\eta(A_1 \cup A_2) = \eta A_1 + \eta A_2$ an thus we have the desired result for $m = 2$.

Next, assume that the result has been established for any k-interval partition such that $k < m$. Let $\mathbf{a} = \{A_1, \ldots, A_m\}$ be an m-interval partition. Note that the inductive hypothesis enables us to conclude that equation (3) holds for any $\mathbf{a}$-measurable T with complexity less than m. In particular,

$$\sum_{j=1}^{n} \eta A_j = \eta(A_1 \cup A_2) + \sum_{j=3}^{n} \eta A_j = \eta(A_1 \cup A_2 \cup A_3) + \sum_{j=4} \eta A_j = \cdots = \eta\Big(\bigcup_{j=1}^{m} A_j\Big) = 1 \tag{4}$$

Since $V(P^{\mathbf{a}}) = 1$ and $V(Q^{\mathbf{a}}) = 2$, we have

$$\sum_{j=1}^{m} \int v(u(x_t^j))d\eta = 1 \text{ and } \sum_{j=1}^{m} \int v(u(y_t^j))d\eta = 2. \tag{5}$$

First, assume m is odd and note that the only set of complexity m that comes up in computing the right-hand side of the first equation in (5) is $A_2 \cup A_4 \cup \cdots \cup A_{m-1}$ while the only equation of complexity m that comes up when computing the right-hand side of the second equation in (5) is $A_1 \cup A_3 \cup \cdots \cup A_m$. Then, invoking equation (3) for all other sets, the martingale property and some simplification yield:

$$\eta(A_2 \cup A_4 \cup \cdots \cup A_{m-1}) + \sum_{i=1}^{(m+1)/2} \eta A_{2i-1} = 1 = \eta(A_1 \cup A_3 \cup \cdots \cup A_m) + \sum_{i=1}^{(m-1)/2} \eta A_{2i} \tag{6}$$

Equations (4) and (6) yield

$$\eta(A_1 \cup A_3 \cup \cdots \cup A_m) = \eta A_1 + \eta A_3 + \cdots + \eta A_m$$
$$\eta(A_2 \cup A_4 \cup \cdots \cup A_{m-1}) = \eta A_2 + \eta A_4 + \cdots + \eta A_{m-1}$$

as desired. For m even, equation (6) becomes

$$\eta(A_2 \cup A_4 \cup \cdots \cup A_m) + \sum_{i=1}^{m/2} \eta A_{2i-1} = 1 = \eta(A_1 \cup A_3 \cup \cdots \cup A_{m-1}) + \sum_{i=1}^{m/2} \eta A_{2i}$$

and once again, the desired conclusion follows.

If v is not affine on any nondegenerate interval, we can still replicate the preceding argument after noting that v must be differentiable at some point and hence can be approximated by an affine function within an ϵ interval. Then, after renormalizing v, we can find four prize lotteries, $\hat{\beta}^j$ for $j = 0, 1, 2, 3$ such that $v(\hat{\beta}^j) = j\epsilon$ for $j = 0, 1, 2, 3$. Then, arguing as above, we have

$$\epsilon + o(\epsilon) \geq \epsilon \sum_{j=1}^{m} \int v(u(x_t^j))d\eta + o(\epsilon) \geq \epsilon + o(\epsilon)$$
$$\epsilon + o(\epsilon) \geq \epsilon \sum_{j=1}^{m} \int v(u(x_t^j))d\eta + o(\epsilon) \geq 2\epsilon + o(\epsilon)$$

where the $o(\epsilon)$'s are distinct terms of order ϵ. Dividing all terms in the above equations by ϵ, letting ϵ go to 0 again yields equation (5). Then, the rest of the argument above can be repeated to complete the proof.

For sufficiency, note that if η is additive,

$$V(P) = \sum_x \int v(u(x_t))d\eta = \sum_t \sum_x v(u(x_t))P_t(x_t)$$

for all P. If P' is more informative than P'' and v is convex, then

$$\sum_x v(u(x_t))P_t'(x_t) \geq \sum_x v(u(x_t))P_t''(x_t)$$

for all t and hence $V(P') \geq V(P'')$ as desired. □

6.5 Proof of Proposition 1

From the RCU formula, we have:

$$V(P^t) = \frac{1}{2}\left(\psi\left(\frac{N-t+1}{N}\right)v(u(\alpha,\beta)) + \left(1-\psi\left(\frac{N-t+1}{N}\right)\right)v(u(\alpha,\beta''))\right)$$
$$+\frac{1}{2}\left(\psi\left(\frac{t-1}{N}\right)v(u(\alpha,\beta'')) + \left(1-\psi\left(\frac{t-1}{N}\right)\right)v(u(\alpha,\beta'))\right)$$

where $\psi(\xi) = 2\xi + \xi^2$. Replacing ξ for $t-1/N$ and ignoring the integer constraint reveals that $V(P^t)$ is strictly concave in ξ. Straightforward calculations reveal that $V(P^t)$ is maximized at $\xi = r := \frac{v(u(\alpha,\beta''))-v(u(\alpha,\beta'))}{v(u(\alpha,\beta))-v(u(\alpha,\beta'))}$. The desired conclusion follows. $\square$

6.6 Proof of Theorem 4

The necessity part of the Theorem is immediate. We will therefore prove only the sufficiency part. In the following, assume Axioms 1-4 and PT hold.

First, note that Lemma 1 continues to hold if we replace rank dominance with dominance for degenerate paths in its proof. Thus, there exists a linear and onto function $u : L \to [0,1]$ such that $u(\alpha,\beta) = u_1(\alpha) + u_2(\beta)$ where u_1, u_2 are also linear functions and $P_{(\alpha,\beta)} \succcurlyeq P_{(\alpha',\beta')}$ if and only if $u(\alpha,\beta) \geq u(\alpha',\beta')$. Then, choose $\overline{\alpha}, \underline{\alpha}$ and $\underline{\beta}$ such that $u_1(\overline{\alpha}) \geq u_1(\alpha) \geq u_1(\underline{\alpha})$ and $u_2(\beta) \geq u_2(\underline{\beta})$ for all α, β. By nondegeneracy, $u(\overline{\alpha},\underline{\beta}) > u(\underline{\alpha},\underline{\beta})$.

Next, note that Axioms 1-3 ensure that the preference on $\Delta(L)$ that $\succcurlyeq$ induces has an expected utility representation with a continuous von Neumann-Morgenstern utility index $w : L \to I\!R$. The observation of the previous paragraph ensures that $w(\alpha,\beta) > w(\alpha',\beta')$ whenever $u_1(\alpha) + u_2(\beta) > u_1(\alpha') + u_2(\beta')$. Finally, Axioms 1-3 and the Mixture Space Theorem ensure the existence of a linear $\hat{V} : \Pi \to I\!R$ that represents $\succcurlyeq$.

We say that P path-by-path dominates $\hat{P}$ if there exist degenerate RELs $P^i, \hat{P}^i$, $a_i > 0$ such that $\sum_i a_i = 1$, $P = \sum_{i=1}^n a_i P^i$, $\hat{P} = \sum_{i=1}^n a_i \hat{P}^i$ and P^i dominates $\hat{P}^i$ for all i. Axiom PT and continuity ensure that $P \succcurlyeq \hat{P}$ ($P \succ \hat{P}$ whenever P dominates (strictly dominates) $\hat{P}$.

Step 1: The binary relation $\succcurlyeq$ satisfies Axiom 5.

First, we will show that if P strictly rank dominates $\hat{P}$, then there is $a \in (0,1]$, P^1, P^2 such that P^2 strictly path-by-path dominates P^1 and $aP + (1-a)P^1$ is experience equivalent to $a\hat{P} + (1-a)P^2$.

The observations stated immediately prior to Step 1 ensure that for every ordered partition ι such that $PX_\iota = \hat{P}X_\iota > 0$ and $t \in N$, there is $a_t \in (0,1]$ such that

$$a_t^\iota E_{p_t} w(\alpha,\beta) + (1-a_t^\iota)w(\underline{\alpha},\underline{\beta}) < a_t^\iota E_{\hat{p}_t} w(\alpha,\beta) + (1-a_t^\iota)w(\overline{\alpha},\underline{\beta})$$

where p_t ($\hat{p}_t$) is the second-order distribution that P_{X_ι} (P_{X_ι}) induces on L in period t. Choose a small enough so that the inequality above holds for all ι, t when a_t^ι is replaced by a. Then, the linearity of u_1 and u_2 and the fact that $w(\alpha, \beta) > w(\alpha', \beta')$ whenever $u_1(\alpha) + u_2(\beta) > u_1(\alpha') + u_2(\beta')$ ensure that there is some $b_t^\iota \in (0, 1]$ such that

$$aE_{p_t} w(\alpha, \beta) + (1-a)w(\underline{\alpha}, \underline{\beta}) = aE_{\hat{p}_t} w(\alpha, \beta) + (1-a)w(\alpha_t^\iota, \underline{\beta})$$

for $\beta_t^\iota = b_t^\iota \underline{\beta} + (1 - b_t^\iota)\beta_t^\iota$.

Let $P^1 = aP + (1-a)P_{(\underline{\alpha}, \underline{\beta})}$ and $P^2 = a\hat{P} + (1-a)P'$ where $P'(x) = \sum_\iota \hat{P}X_\iota \cdot P(x^\iota)$. By construction, $P_t^1 \sim P_t^2$ for all t. Also, note that for all ι there are t, s such that $\succcurlyeq_x^h = x_t$ and $\succcurlyeq_x^l = x_s$ for all x such that $x \in X_\iota$. Finally, note that for all ι, there are t, s such that $\succcurlyeq_x^h = x_t$ and $\succcurlyeq_x^l = x_s^\iota$ for $x = x_{(\alpha, \beta)}$ and all $x \in X_\iota$ such that $P'(x) > 0$. Hence, P^1 and P_2 are experience equivalent.

To conclude the proof of this step, note that by Axiom PT, $P^1 \sim P^2$; that is, $a\hat{V}(P) + (1-a)\hat{V}(P_{(\alpha,\alpha)}) = a\hat{V}(\hat{P}) + (1-a)\hat{V}(P')$. Since P strictly rank dominates $\hat{P}$, there must be at least one $b_t^\iota < 1$ and therefore P' strictly path-by-path dominates $P_{(\underline{\alpha}, \underline{\beta})}$. Hence, $\hat{V}(P_{(\alpha,\alpha)}) < \hat{V}(P')$ and therefore, $\hat{V}(P) > \hat{V}(\hat{P})$ as desired.

Since $\succcurlyeq$ satisfies Axiom 5 together with Axioms 1–4, by Theorem 1, it has a RCU representation. Let (u, v, η) be this representation.

Step 2: $\eta(T \cup \{t\}) - \eta T = \eta(S \cup \{t\}) - \eta S$ for all S, T such that $s \notin T \cup S$ and $T^c \neq \{s\} \neq S^c$.

For $N = 2$, the statement is vacuously true. So, assume that $N \geq 3$. For any $\hat{S} \subset N$, let $1^{\hat{S}}$ be the path $x^{\hat{S}}$ as follows: $x_t^{\hat{S}} = (\overline{\alpha}, \underline{\beta})$ for all $t \in \hat{S}$ and $x_t^{\hat{S}} = (\underline{\alpha}, \underline{\beta})$ for all $t \notin \hat{S}$. Recall that by our normalization, $v(u(\underline{\alpha}, \underline{\beta})) = 0$ and by nondegeneracy, $c := v(u(\overline{\alpha}, \underline{\beta})) > 0$.

For any T, S satisfying the conditions above, let $P = {}^1\!/\!_2 x^S + .{}^1\!/\!_2 x^{T \cup \{t\}}$ and $P' = {}^1\!/\!_2 x^{S \cup \{t\}} + {}^1\!/\!_2 x^{\hat{T}}$ and note that P, P' are experience equivalent. By Axiom PT, $V(P) = V(P')$; that is,

$$.5c\eta S + .5c\eta(T \cup \{t\}) = .5c\eta(S \cup \{t\}) + .5c\eta T$$

Hence, $\eta(T \cup \{t\}) - \eta T = \eta(S \cup \{t\}) - \eta S$ as desired.

Step 3: $\succcurlyeq$ has a PTU representation.

First, assume $N = 2$. Let $\lambda_t = \frac{1-\eta(3-t)}{2-\eta(1)-\eta(2)}$, $\delta_h = \eta(1) + \eta(2) - 1$ and $\delta_\ell = 0$. Let $V = (u, v, \delta_h, \delta_\ell)$. It is easy to verify that $E_P \int vud\eta = V(P)$ for all P. The restrictions on δ_h, δ_ℓ follow from the strictness of η; that is, $0 < \eta(t) < 1$ for $t = 1, 2$.

Next, assume $N > 2$. Then, we claim that there is a unique collection δ_n for $n = 1, \ldots, h, l$ such that (i) $\eta T = \delta_h + \sum_{t \in T} \delta_t$ whenever $T \neq \emptyset \neq T^c$ and (ii) $1 - \eta N \backslash \{t\} = \delta_t + \delta_\ell$ for all $t \in N$. Moreover, (iii) $\delta_t > 0$, $\delta_h + \delta_s > 0$ and $\delta_\ell + \delta_t > 0$ for all $t \in N$.

To see why this case, for all t, choose $t' \neq t$ and let $\delta_t = \eta\{t, t'\} - \eta(t)$. Then, note that for any t and s, Step 2 yields $\eta(s) + \delta_t = \eta\{s, t\} = \eta(t) + \delta_s$. Hence, there is δ_h such that $\eta(t) = \delta_h + \delta_t$ for all $t \in N$. This fact together with Step 2 yields (i).

Let $\delta_\ell = 1 - \delta_h - \sum_{s=1}^N \delta_s$. Then, by (i), $1 - \eta(N \backslash \{t\}) = 1 - \delta_h + \delta_t - \sum_{s=1}^N \delta_t = \delta_\ell + \delta_t$ proving (ii). Then, (iii) follows from the strictness of η.

Let λ be the probability on $\{1, \ldots, N\}$ defined as follows: $\lambda_t = \frac{\delta_t}{1-\delta_h-\delta_\ell}$ and let $V = (u, v, \lambda, \delta_h, \delta_\ell)$. Condition (iii) above ensures that δ_h, δ_ℓ and λ satisfy restrictions in the definition of a PTU and therefore V is a PTU. Conditions (i) and (ii) imply $V(P) = E_P vud\eta$ and hence V represents $\succcurlyeq$. The uniqueness of u, v follows from standard uniqueness arguments applied to the restriction of $\succcurlyeq$ to RELs of the form P_p for $p \in \Delta(L)$. The uniqueness of λ, δ_h and δ_ℓ follows from the uniqueness of η in a RCU representation. $\square$

6.7 Proof of Proposition 2

For part (ii), note that for any $P \in \Pi_a^g$, $P_\ell(a) = a$ and $P_\ell(0) = 1 - a$. Hence,

$$V(P) = (1 - \delta_h - \delta_\ell) \sum_t E_{P_t} v(u(\beta))\lambda(t) + \delta_h E_{P_h} v(u(\beta)) + \delta_\ell \cdot a^2$$

and therefore,

$$\begin{aligned} &V(P^g) - V(\hat{P}^g) = \\ &(1 - \delta_h - \delta_\ell) \sum_t [E_{P_t^g} v(u(\beta)) - E_{\hat{P}_t^g} v(u(\beta))]\lambda(t) + \delta_h [E_{P_h^g} v(u(\beta)) - E_{\hat{P}_h^g} v(u(\beta))] \end{aligned} \tag{7}$$

Since P^g resolves later than $\hat{P}^g$, either $P_t^g = \hat{P}_t^g$ or $\hat{P}_t^g$ is a mean preserving spread of P_t^g for all $t \in N$. Since v is concave, we conclude that the first term on the right-hand side of equation (7) is non-negative.

Next, we claim that either $P_h^g = \hat{P}_h^g$ or P_h^g stochastically dominates $\hat{P}_h^g$. To see why this is the case, let t, x be such that $P^g(y) = \hat{P}^g(y)$ for all $y \notin [x]_t$ and such that $y_{t+1} \in \{0, 1\}$ for all $y \in [x]_t$ with $\hat{P}^g(y) > 0$. Since the path-maximum of any $y \in [x]_t$ is at least $c = \max_{s \le t} x_s$, we have $\sum_{\beta \le t} P_h^g(\beta) = \sum_{\beta \le t} \hat{P}_h^g(\beta)$ for all $t < c$. Since

$$P^g\{y \in [x]_t \,|\, h(y, \succcurlyeq) = 1\} = x_t P[x]_t = \hat{P}^g\{y \in [x]_t \,|\, y_{t+1} = 1\},$$

$$P^g\{y \in [x]_t \,|\, h(y, \succcurlyeq) = c\} \le (1 - x_t)P[x]_t = \hat{P}^g\{y \in [x]_t \,|\, y_{t+1} = 0\}$$

the desired conclusion follows.

The proof of part (iii) is essentially identical and omitted. Take any $P^g \in \Pi_a^g$ and $P^b \in \Pi_a^b$ and let P be the unique REL in $\Pi_a^g \cap \Pi_a^b$ such that $P(x) > 0$ implies $x_2 \in \{0, 1\}$. Hence, P resolves all uncertainty in period 2. Therefore, P^g resolves later than P and by part (ii), $V(P^g) \ge V(P)$. By part (iii), $V(P) \ge V(P^b)$ and therefore, $V(P^g) \ge V(P^b)$. $\square$

6.8 Proof of Proposition 3

Let $X_a^* := \{x \in L^N \,|\, x_1 = a, x_{t+1} \notin (0, x_t)\}$; that is, $x \in X_a^*$ if either $x_{t+1} \ge x_t$ or $x_{t+1} = 0$. Let Π_a^* be the RELs that satisfy $P(X_a^*) = 1$. Note that Π_a is compact and therefore, there is $\hat{P} \in \Pi_a$ with $V(\hat{P}) \ge V(P)$ for all $P \in \Pi_a$; that is, V attains its maximum in Π_a. Next, we will show that for all $P \in \Pi_a \backslash \Pi_a^*$, there is $\hat{P} \in \Pi_a^*$ such that $V(\hat{P}) \ge V(P)$.

For any $P \in \Pi_a$, call (x, s) a violation if $P(x) > 0$ and $x_s \in (0, \max_{t \le s} x_t)$. Note that $P \in \Pi_a$ has at most finitely many violations and $P \in \Pi_a^*$ if and only if it has no violations. To establish the desired conclusion, we will show that for all $P \in \Pi_a \backslash \Pi_a^*$, there is $\hat{P}$ with fewer violations such that $V(\hat{P}) \ge V(P)$.

For any path x with a violation, let s be the last period such that (x, s) is a violation. Since $x_N \in \{0, 1\}$, we must have $s < N$ and $\max_{t \le s} x_t \le x_{s+1}$ or $x_{s+1} = 0$. Then, for each path $y \in [x]_s$, define the path z^y as follows:

$$z_t^y = \begin{cases} y_{t+1} & \text{if } t = s \\ y_t & \text{otherwise} \end{cases}$$

Let $\hat{P}$ be the REL such that

$$\hat{P}(z) = \begin{cases} 0 & \text{if } z \in [x]_s \\ P(z) + P(y) & \text{if } z = z^y \text{ and } y \in [x]_s \\ P(z) & \text{otherwise} \end{cases}$$

Hence $\hat{P}$ is identical to P except after the history $(x_1, \ldots, x_{s-1})$; after this history, $\hat{P}$ reveals additional information that P would have revealed in period $s+1$ after history $(x_1, \ldots, x_s)$.

Clearly, $\hat{P} \in \Pi_a$ and has fewer violations than P. We claim that $V(\hat{P}) \geq V(P)$. This follows since, for any $y \in [x]_s$, $\max_s y_s = \max_s z_s^y$ and $\min y_s \leq \min z_s^y$. We can repeat the argument above to obtain $\hat{P} \in \Pi_a^*$ such that $V(\hat{P}) \geq V(P)$.

Next, we show P^* is optimal (in Π_a^*). Note that every path $x \in X_a^*$ that terminates in 1 has a as its trough and every path that terminates in 0 has 0 as its trough. Since the probability that a path terminates in 1 or 0 is fixed for all RELs in Π_a, all $P \in \Pi_a^*$ have the same distribution of path-troughs. Thus, for $P \in \Pi_a^*$, $V(P) = (1 - \delta_h - \delta_\ell)a + \delta_\ell a^2 + \delta_h M(P)$ where $M(P)$ is the expected value of the path-peaks. Note that $M(P^*) = Na - (N-1)a^{\frac{N}{N-1}}$. To show that P^* is optimal, it suffices to show that $M(P^*) > M(P)$ for all $P \in \Pi_a^*$.

If $N = 2$, then P^* is the unique element of Π_a^* and $M(P) = 2a - a^2$ and therefore the assertion is true. Assume P^* is optimal whenever $N \leq n$ and let $N = n+1$. Let P be any optimal REL and let $p = P_2$. Hence, p is the period-2 distribution of some optimal REL P. Since $P \in \Pi_a^*$, $p(b) > 0$ implies $b = 0$ or $b \geq a$. We can associate, with each $b > 0$ such that $p(b) > 0$, an n-period REL, starting in period 2. This REL will be in the set Π_b^* for n-period RELs. Moreover, this n-period REL must have maximum expected value of path-peaks in Π_b^*, otherwise P would not be optimal. By the inductive hypothesis,

$$M(P) = \sum_{b>0} p(b)[nb - (n-1)b^{n/n-1}] + p(0)a$$

Let $f(b) = nb - (n-1)b^{n/n-1}$ and note that f is a strictly concave function. We claim that $\{b > 0 \,|\, p(b) > 0\}$ must be a singleton. To see why, suppose the set $\{b > 0 \,|\, p(b) > 0\}$ has at least two elements; that is p has at least three elements in its support. Then, since $1 - p(0) = \sum_{b>0} p(b)$, we have

$$[1-p(0)]f\left(\sum_{b>0} \frac{p(b)}{1-p(0)} b\right) > [1-p(0)] \sum_{b>0} \frac{p(b)}{1-p(0)} f(b) = \sum_{b>0} p(b) f(b)$$

Hence, p must have exactly two elements in its support and a unique element $b > 0$. This b must maximize $(1-p(0))f(b) + p(0)a$ where $p(0) = 1 - a/b$ by the martingale property. A straightforward calculation shows that $b = a^{\frac{n-1}{n}} = a^{\frac{N-2}{N-1}}$ as desired. Hence, P^* is the unique optimal REL in Π_a^*.

To conclude the proof, we will show that there is no $P \in \Pi_a \backslash \Pi_a^*$ such that $V(P) = V(P^*)$. If such a P existed, we could modify it, as described above to obtain $\hat{P} \in \Pi_a^*$ such that

$V(\hat{P}) \geq V(P)$. This $\hat{P}$ would have, in its support, a path x such that $0 < x_t = x_{t+1}$ for some $t \leq N$. Thus, $\hat{P} \neq P^*$ contradicting the fact that P^* is the unique maximizer of V in Π_a^*. $\square$

6.9 Proof of Proposition 4

For any PTU with linear v and $P = (g, b) \in \Pi_a$, we have $V(P)$ equal to

$$(1 - \delta_h - \delta_\ell)a + cg(\delta_h + \delta_\ell a) + c(1-g)\delta_h g + (1-c)b(\delta_h + b\delta_\ell) + (1-c)(1-b)a\delta_h$$

and similarly, $V(P^{\#})$ is equal to

$$(1 - \delta_h - \delta_\ell)a + (1-c)g'(\delta_h + \delta_\ell a) + (1-c)(1-g')\delta_h g' + cb'(\delta_h + b'\delta_\ell) + c(1-b')a\delta_h$$

where $cg + (1-c)b = a = (1-c')g' + c'b'$. Let $d^g = g - a, d^b = a - b$ and note that $g' = a + d^b, b' = a - d^g$. Moreover, $cd^g = (1-c)d^b$ by the martingale property. Substituting these restrictions, a straightforward calculation yields

$$V(P) - V(P^{\#}) = (\delta_h + \delta_l)\frac{d^g d^b(d^b - d^g)}{d^g + d^b}$$

Thus, $V(P) - V(P^{\#}) > 0$ if and only if $(\delta_h + \delta_l)(d^b - d^g) > 0$ proving the result. $\square$

6.10 Proof of Proposition 5

In the online appendix, we extend our framework to a recursive choice problem and we show that the agent is dynamically consistent. Hence, we may prove the result by comparing the ex-ante value of the REL generated by each behavioral strategy. Let P denote the REL generated by acquiring a signal in period 3 if and only if the period-2 signal is positive (ostrich effect behavior). Let P' denote the REL generated by never acquiring a signal in period 3; let P'' denote the REL generated by always acquiring a signal in period 3; and, finally, let P''' denote the REL generated by acquiring a signal in period 3 if and only if the period-2 signal was negative (the opposite of the ostrich effect behavior). Applying formula (PT) yields

$$V(P) - V(P') = V(P'') - V(P''') = \frac{a(1-a)^2 d(2d-1)(1-d)^3}{[a(2d-1) + 1 - d]\,[a(2d-1) + (1-d)^2]}\delta_h$$

and

$$V(P) - V(P'') = V(P') - V(P''') = \frac{(1-a)a^2(1-d)^3 d(2d-1)}{(a + d - 2ad)\,(a + d^2 - 2ad)}\delta_\ell$$

Since $0 < a < 1$ and $1/2 < d < 1$, it is readily verified that $V(P) > V(P')$ if and only if $V(P'') > V(P''')$ if and only if $\delta_h > 0$ and that $V(P) > V(P'')$ if and only if $V(P') > V(P''')$ if and only if $\delta_\ell > 0$. $\square$

References

Anscombe, F.J. and Aumann, R.J. (1963) "A definition of subjective probability," *Annals of mathematical statistics,* 34, 199–205.

Bastianello, L. and Chateauneuf, A. (2016) "About delay aversion," *Journal of Mathematical Economics,* 63, 62–77.

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

Chateauneuf, A., J. Eichberger and S. Grant (2007) "Choice under uncertainty with the best and worst in mind: Neo-additive capacities" *Journal of Economic Theory*, 137, 538–567.

Chateauneuf, A. and Rebille, Y. (2004) "Some characterizations of non-additive multi-period models," *Mathematical social sciences,* 48, 235–250.

Chateauneuf, A. and Ventura, C. (2013) "G-continuity, impatience and myopia for Choquet multi-period utilities," *Journal of Mathematical Economics,* 49, 97–105.

De Waegenaere, A. and Wakker, P. (2001) "Nonmonotonic Choquet integrals," *Journal of Mathematical Economics,* 36, 45–60.

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

Dillenberger, D. and U. Segal (2017) "Skewed Noise," *Journal of Economic Theory,* 169, 344–364.

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.

Fredrickson, Barbara L.; Kahneman, Daniel (1993) "Duration neglect in retrospective evaluations of affective episodes," *Journal of Personality and Social Psychology,* 65, 45–55.

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.

Gul, F., Natenzon, P., Ozbay, E. and Pesendorfer, W. (2020) "The Thrill of Gradual Learning," Working paper.

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

Klibanoff, P., Marinacci, M. and Mukerji, S. (2005) "A smooth model of decision making under ambiguity," *Econometrica*, 73, 1849–1892.

Kreps, D.M. and Porteus, E.L. (1978) "Temporal resolution of uncertainty and dynamic choice theory," *Econometrica*, 46, 185–200.

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.

Masatlioglu, Y., Orhun, A.Y. and Raymond, C. (2017) "Intrinsic information preferences and skewness," Ross School of Business Paper. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3232350

Rebille, Y. (2007) "Patience in some non-additive models," *Journal of Mathematical Economics,* 43, 749–763.

Seo, K. (2009). "Ambiguity and second-order belief," *Econometrica,* 77, 1575–1605.

Shalev, J. (1997) "Loss aversion in a multi-period model, *Mathematical Social Sciences,* 33, 203–226.

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