WORKING PAPERS

N° TSE-585

June, 18, 2015

“Anticipating Preference Reversal”

Yves Le Yaouanq

Toulouse School of Economics

Anticipating preference reversal

Yves Le Yaouanq *

June 18, 2015.

Abstract

This paper studies the consistency between a decision-maker's choices over menus in a first period and inside menus at a later date. The main result shows that the comparison of commitment decisions and actual subsequent choices reveals whether future taste contingencies are correctly anticipated: a sophisticated individual chooses exactly the right commitment options, whereas a naive decision-maker overlooks some profitable opportunities. The paper provides absolute and comparative measures of naivete and shows under which conditions pessimistic behavior can be attributed to the presence of self-control costs. Finally, I implement an experimental protocol based on the theoretical analysis and find substantial evidence of naivete at the individual level.

Keywords: self-control, naivete, temptation, stochastic choice, random Strotz.
JEL codes: C91, D81, D90.

*Le Yaouanq: Toulouse School of Economics. TSE, Université de Toulouse, Manufacture des Tabacs, 21 allée de Brienne, Fr - 31000 Toulouse. E-mail: yves.le-yaouanq@m4x.org. An earlier version of this article was circulated under the title "Anticipating Temptation". I am deeply indebted to Roland Bénabou, Christian Gollier and Jean Tirole for their guidance, advice and encouragement. I am also grateful to Faruk Gul, Thomas Mariotti, Stephen Morris, Wolfgang Pesendorfer, John Quah, Takuro Yamashita, and seminar participants at Princeton University and TSE. I thank the Department of Economics of Princeton University for its hospitality and the Corps des Mines for financial support. This research received approval from the TSE Research Ethics Committee.

1 Introduction

The literature on time-inconsistent preferences distinguishes two types of individuals according to their beliefs regarding their own future behavior. Sophisticated agents correctly anticipate their future choices, while naive individuals underestimate their propensity to deviate from their long-term goals. A usual observation is that present bias is not an issue *per se* as soon as it is correctly forecast: a sophisticated agent has the opportunity to compensate future deviations by engaging in appropriate actions (signing optimal contracts, buying commitment devices, etc.). In contrast, the combination of self-control issues and unrealistic expectations harms decision-makers. For instance, in applications of the quasi-hyperbolic discounting model (Laibson, 1997), fixing β the present-bias parameter, the assumptions made regarding the agent's expectation of the future value of β play an important role in welfare and policy analysis (Heidhues and Köszegi, 2010; Köszegi, 2014; Rabin and O'Donoghue, 1999; Eliaz and Spiegler, 2006).

This paper proposes an axiomatic framework in which the sophistication hypothesis can be tested by observing choice data only. The results are used to design an experimental method that offers several advantages: it allows to measure naivete at the individual level; it allows to detect partial naivete; it is nonparametric, robust to risk aversion and does not rely on functional forms or specific hypotheses about the intertemporal preferences.

The axiomatic framework takes as a primitive the choices made by a decision-maker at two successive periods. At a first stage, the decision-maker chooses the set of options that will be available in the future, as in Kreps (1979) and Dekel et al. (2001), according to a binary preference $\succeq$ over menus. At a subsequent period, the agent makes a stochastic choice inside the available menu according to a random choice rule λ. The *ex ante* preference $\succeq$ is represented by a Random Strotz model (Dekel and Lipman, 2012): in this interpretation, the agent has a certain normative preference over the prizes but expects possible deviations in the future, which creates a preference for smaller menus. The relation $\succeq$ identifies the agent's beliefs about the realization of future taste contingencies. The *ex post* choice rule λ is represented by a Random Expected Utility model (Gul and Pesendorfer, 2006): in this interpretation, the choice in the menu is driven by the realization of an uncertain taste contingency, whose distribution is also uniquely identified from the data.

The aim of the paper is to compare the *ex ante* anticipation of taste contingencies suggested by $\succeq$ with the actual realization of *ex post* preferences identified from λ. Section 2 provides precise definitions of sophistication and naivete in this framework. A sophisticated agent has the right model in mind about her future choices, meaning that her beliefs over future taste contingencies are correct. In contrast, a naive decision-maker underestimates her future deviations. To give content to this notion, I build on Dekel and Lipman (2012) and parameterize each possible deviation by two parameters, its *direction* and its *intensity*, the latter being a measure of the frequency of preference reversals with respect to long-term goals. Naivete is defined as the situation in which the agent underestimates in a first-order stochastic sense the intensity of future deviations in each direction.

The main result characterizes the behavioral content of these definitions under the form of conditions on the pair of preferences $\{\succeq, \lambda\}$. The axioms rely on simple tests of the following form: observing the decision-maker's preference between a choice set $\{p, q\}$ and a commitment device of the form $\{\kappa p + (1-\kappa)q\}$, where κ is an exogenous probability chosen by the experimenter. $\{\kappa p + (1-\kappa)q\}$ can be interpreted as delivering p with probability κ and q with probability $1-\kappa$. If p is *ex ante* preferred to q and if κ is lower than $\lambda^{\{p,q\}}(p)$, the actual probability with which the decision-maker chooses p *ex post*, the commitment device is rejected by a sophisticated agent who understands that she would choose p with a better probability from the whole choice set. This condition, called *No Commitment to Inferior Lotteries*, rules out pessimistic anticipations. The second axiom, *Commitment to Superior Lotteries*, rules out optimistic expectations: if $\kappa > \lambda^{\{p,q\}}(p)$, a sophisticated decision-maker accepts the commitment device. The main result of section 3 is that extensions of these axioms to larger menus characterize sophistication. In contrast, naive agents never commit to inferior lotteries but fail to seize commitment opportunities that appear profitable in light of their *ex post* behavior.

Another contribution of the paper is to introduce a subclass of representations, the Unidimensional Random Strotz models, in which attention can be restricted to menus of two elements. Section 2 defines and provides foundations for this class of representations whose relevant deviations share a same direction but might vary in intensity. This refers to situations where deviations occur for a unique reason, but where the intensity of preference reversals is unknown *ex ante*. Since most experimental settings are likely to belong to this class, assessing naivete and sophistication is easily done by varying κ and observing the choices between menus of the form $\{p, q\}$ and $\{\kappa p + (1-\kappa)q\}$.

Comparing commitment choices with *ex post* behavior identifies the degree to which the decision-maker's *ex ante* anticipations are naive. The subjective expectation regarding $\lambda^{\{p,q\}}(p)$ can be identified as the threshold κ that makes the decision-maker indifferent between $\{p, q\}$ and $\{\kappa p + (1 - \kappa)q\}$. Comparing this value with $\lambda^{\{p,q\}}(p)$ reveals whether the agent is sophisticated, naive or pessimistic, and in these latter cases to which degree. Section 3 defines a local measure of naivete that represents, for each menu, the first-order stochastic distance between the *ex ante* and *ex post* beliefs regarding the choices made inside the menu. This index is intuitively related to properties of the joint representations: fixing the *ex post* behavior, a uniform increase in the index of naivete is equivalent to a first-order stochastic downward shift in *ex ante* beliefs about the intensity of deviation, and, equivalently, to a lower demand for commitment. Fixing the preferences over menus, a (uniform) increase in the index of naivete is equivalent to a first-order stochastic upward shift in *ex post* realized deviations and, equivalently, to a higher propensity to make *ex post* choices that are suboptimal from the *ex ante* perspective.

A possible caveat to this analysis is that the Random Strotz model is an arbitrary interpretation of commitment choices: indeed, there exist other representations of preferences over menus that are consistent with the same behavior, but that suggest a different system of beliefs. A false interpretation of the agent's cognitive process and preferences is thus a potential confound, since it could lead to reject the sophistication hypothesis by mistake. As a robustness check, section 4 analyzes sophistication and naivete under the other prominent model of menu choices, the Random Gul-Pesendorfer representation (Stovall, 2010). Under some restrictions, this model is observationally equivalent to the Random Strotz model for the preference over menus (Dekel and Lipman, 2012), but the *ex post* choices that it suggests are different. The most important difference is that the Random Gul-Pesendorfer model incorporates decision costs than cannot be observed *ex post*: a decision-maker might be willing to remove an option from a choice set even if this option is never chosen, as soon as its presence is unpleasant enough to inflict a decision cost. Unsurprisingly, choices that appear naive under the Random Strotz model are also naive under any equivalent Random Gul-Pesendorfer representations: adding *ex ante* menu-contingent decision costs cannot rationalize overconfident commitment choices. However, choices that appear pessimistic in the Random Strotz interpretation might be rationalized by a model incorporating decision costs, as is shown in section 4.

Finally, section 5 reports the results of an experimental design based on the theoretical results. While the main application of the setting relates to temptation and self-control issues, the Random Strotz model is silent about the source of preference reversals. The experimental protocol relies on this property and focuses on naivete about future memory failures. Participants have the possibility to earn a monetary prize p every day within a ten days period if they remember to log in to an experimental website during the day. They earn nothing (q) if they forget to do so. Prior to this ten days session, their indifference threshold κ^* between this procedure and a payment rule that delivers the prize with probability κ is elicited. This latter choice can be interpreted as the commitment device $\{\kappa p + (1-\kappa)q\}$. Comparing κ^* with the actual frequency of visiting the website over the ten days session provides a precise measure of the difference between *ex ante* expectations and *ex post* choices. I find that 66% of participants make naive choices, while 22% make sophisticated choices and only 12% display pessimistic expectations. The results show that considering stochastic choice substantially refines our measurements, and a lot of information would be lost by considering only perfectly sophisticated and completely naive decision-makers.

This work builds on the literature on menu choices, started by Kreps (1979) and pursued by Dekel et al. (2001). This field is interested in finding representations for preferences over menus that are interpretable in terms of anticipation of future subjective tastes. Applications usually include preference for flexibility, reflected in a taste for larger menus, or the role of temptation, which leads the agent to prefer smaller choice sets (Gul and Pesendorfer, 2001; Noor, 2007; Stovall, 2010; Dekel and Lipman, 2012; Kopylov, 2012; Kopylov and Noor, 2015; Noor and Takeoka, 2010). However, this literature is silent about the actual choice inside menus, which is usually left unmodeled, and it assumes that anticipations correctly predict future behavior. For instance, Dekel and Lipman (2012) show that, while the Random Strotz and Random Gul-Pesendorfer representations of choices over menus are equivalent, the *ex post* behaviors that they predict are different. As a result, observing *ex post* choices and assuming sophistication allows to separate them. In contrast, relaxing the sophistication hypothesis and allowing for incorrect anticipations breaks the identification: a naive Random Strotz model cannot be disentangled from a naive Random Gul-Pesendorfer model. Consequently, results that are robust to both specifications are useful since they do not require to take a stand on which model is the most accurate representation of behavior.

The first paper to explicitly model choice inside menus, and to provide tools to compare anticipations and realization of tastes, is the work by Ahn and Sarver (2013). They study the correspondence between *ex ante* and *ex post* subjective tastes in the particular case where the agent values flexibility. They use the Dekel et al. (2001) framework with a monotonicity assumption, in which the normative utility is uncertain *ex ante*, and aligned with the decision utility *ex post*. Their analysis shows that two axioms, *Consequentialism* and *Foreseen Contingencies*, are necessary and sufficient conditions for a joint sophisticated representation[1]. The contribution of this paper is to perform the same exercise under the assumption that the decision-maker values commitment instead of flexibility, which is suited to the analysis of sophistication and naivete in the context of preference reversals. Variants of *Consequentialism* and *Foreseen Contingencies* are necessary but not sufficient in the Random Strotz model, as explained in section 3. Finally, Ahn et al. (2015) study a related setting in a recent paper conceived independently from this work. They propose different absolute and comparative notions of naivete, as well as an application to intertemporal discounting models, while the theoretical results in this paper are more oriented towards testable behavioral properties.

2 Primitives

2.1 Objects of choice

Consider a finite set of prizes $\mathcal{Z}$, and $\Delta(\mathcal{Z})$ the set of all probability distributions on $\mathcal{Z}$, written $p, q, ...$ and called lotteries. $\Delta(\mathcal{Z})$ is endowed with the Euclidian topology, each element of $\Delta(\mathcal{Z})$ being identified with a vector of $\mathbb{R}^{|\mathcal{Z}|}$. $\mathcal{X}$ is the set of finite non-empty subsets of $\mathcal{Z}$, and elements of $\mathcal{X}$ are written $x, y, ...$ and called menus. $\mathcal{X}$ is endowed with the Hausdorff topology.

Let $\mathcal{U}$ be the set of all expected utilities on $\mathcal{Z}$. An element of $\mathcal{U}$ can be identified with a vector of $\mathbb{R}^{|\mathcal{Z}|}$. Consider the subset $\mathcal{W}$ containing all elements u of $\mathcal{U}$ that verify $\sum u(z) = 0$ and $\sum u(z)^2 = 1$. Each nonconstant expected utility can be identified with a unique element u of $\mathcal{W}$.

The behavior of a decision-maker is observed in two periods. At the *ex ante* stage, the agent has preferences over menus, as in Kreps (1979). A menu contains the options that will be available to the decision-maker in the future. This choice is described by a preference relation $\succeq$ defined on $\mathcal{X}$, $x \succeq y$ meaning that the agent prefers to choose inside the menu x rather than in y at the later period. As usual, $\succ$ denotes the asymmetric part of $\succeq$. At the *ex post* stage, the agent picks one element in the set. Her choice process is modeled as a random choice rule, i.e. as a function $\lambda : \mathcal{X} \to \Delta(\Delta(\mathcal{Z}))$ such that $\lambda^x(x) = 1$ for any menu $x \in \mathcal{X}$. If y is a subset of a menu x, $\lambda^x(y)$ represents the probability with which an object in y

[1] Dean and McNeill (2015) find some support for these axioms experimentally.

is picked when the agent chooses in x. To lighten the notation, $\lambda^x(\{p\})$ is simply written $\lambda^x(p)$. $(\succeq, \lambda)$ is our primitive.

To define sophistication and naivete in this setting, we need to put more structure into these objects in order to introduce anticipated and actual taste contingencies. This section describes the representations chosen to model $\succeq$ and λ.

2.2 Random Strotz

Preferences over menus are represented by a Random Strotz model (Dekel and Lipman, 2012).

Definition 2.1. The preference relation $\succeq$ admits a *Random Strotz representation* if there exists a pair (u, μ) where $u \in \mathcal{W}$ is a nontrivial expected utility and μ is a nontrivial probability distribution on $\mathcal{W}$ such that $\succeq$ is represented by

$$V(x) = \int_{\mathcal{W}} \mu(dw) \max_{p \in \mathcal{M}_w(x)} u(p) \tag{2.1}$$

where $\mathcal{M}_w(x) = \{p \in x | \forall q \in x, w(p) \geq w(q)\}$ is the set of maximizers of w in x.

Dekel and Lipman (2012) show that the pair (u, μ) that represents $\succeq$ is unique. To understand this representation, suppose first that the support of μ is a singleton w. The valuation of a menu x is equal to $u(p)$, where p is the lottery that maximizes w in x (with ties broken in favor of u). This suggests that the agent has long-term preferences given by u at the *ex ante* stage but anticipates that her choice inside menus will maximize w instead. The decision-maker displays a preference for commitment as soon as w differs from u, and does not value flexibility since the normative preference u is certain. The representation 2.1 adds some uncertainty about the future decision utility, keeping the normative preferences certain. The utility u is referred to as the commitment utility since it represents the preference over singleton menus.

2.3 Random Expected Utility

The stochastic choice made by the agent inside menus is modeled by a Random Expected Utility representation, as axiomatized by Gul and Pesendorfer (2006). This model represents λ as the result of the maximization of a stochastic utility v drawn from $\mathcal{W}$ according to a measure ν that represents the distribution of *ex post* preferences.

A well-known issue that arises in random utility models is that a particular w selected among the possible decision utilities might admit multiple maximizers in the choice set. In that case, the choice prescribed by w is ambiguous. To overcome this issue, I follow Gul and Pesendorfer (2006) and Ahn and Sarver (2013) and assume that indifference is resolved according to a tie-breaking procedure.

Let $\mathrm{B}_{\mathcal{W}}$ be the Borel σ-algebra on $\mathcal{W}$ and $\Delta^f(\mathcal{W})$ denote the set of all finitely additive probability measures over $(\mathcal{W}, \mathrm{B}_{\mathcal{W}})$. A tie-breaking rule specifies, for each $w \in \mathcal{W}$, how the choice is made in the case where w has multiple maximizers in the choice set.

Definition 2.2. A *tie-breaking rule* is a function $\tau : \mathcal{W} \to \Delta^f(\mathcal{W})$ such that, for all $x \in \mathcal{A}$ and $p \in x$:

$$\tau_w(\{v \in \mathcal{W} | \forall q \in x \setminus \{p\}, v(p) > v(q)\}) = \tau_w(\{v \in \mathcal{W} | v(p) = \max_{q \in x} v(q)\})$$

Among the set of maximizers of w in a menu x, a lottery p is chosen if the tie-breaker τ_w chooses an expected utility $v \in \mathcal{W}$ such that p maximizes v among the maximizers of w. Hence, to be selected an element must survive a two-stage procedure: first being a maximizer of w, second being a maximizer of v among the remaining lotteries, v being chosen according to the distribution τ_w. Definition 2.2 ensures that this second comparison resolves the indifference.

Definition 2.3. λ has a *Random Expected Utility representation* if there exists a measure ν on $\mathcal{W}$ and a tie-breaking rule τ on $\mathcal{W}$ such that, for $y \subseteq x$,

$$\lambda^x(y) = \int_{w \in \mathcal{W}} \tau_w(\{v \in \mathcal{W} | \mathcal{M}_v(\mathcal{M}_w(x)) \in y\}) d\nu(w)$$

The measure ν is defined over the set of expected utilities $\mathcal{W}$, and $\lambda^x(y)$ equals the probability with which the outcome of the two-stages process described above belongs to y. This representation suggests that an expected utility w is drawn according to the measure ν, and that the agent's choice maximizes this realized preference.

Since the Random Strotz representation implicitly assumes that ties are broken in favor of u, we restrict attention to Random Expected Utility that also satisfy this property: if $p \in \mathcal{M}_w(x)$ but $p \notin \mathcal{M}_u(\mathcal{M}_w(x))$, p is chosen with probability zero by the tie-breaking rule.

Assumption 1. $\forall x \in \mathcal{A}, w \in \mathcal{W}, \tau_w(\{v \in \mathcal{W} | \mathcal{M}_v(\mathcal{M}_w(x)) \not\subset \mathcal{M}_u(\mathcal{M}_w(x))\}) = 0.$

An alternative possibility would be to include this property in the definition of sophistication. This would introduce some cumbersome notation but the axioms and the results would be unchanged. Assumption 1 is therefore maintained for the sake of simplicity, which allows to restrict attention to the comparison between μ and ν and to avoid discussing the presence of multiple maximizers.

2.4 Sophistication and naivete

2.4.1 Partial order on $\mathcal{W}$

Given those primitives, we are interested in comparing the metacognitive process of the agent, reflected in μ, with the actual realization of tastes, reflected in ν. Defining sophistication is straightforward: a sophisticated agent has exactly the right model in mind *ex ante* when she contemplates her future choices, which is equivalent to the equality $\mu = \nu$. Defining naivete, in contrast, requires to capture the fact that the decision-maker systematically underestimates the strength of her future deviations. To give content to this notion, I build on Dekel and Lipman (2012) to define a notion of *intensity* of future temptations.

Definition 2.4. Define the order $\succeq^u$ on $\mathcal{W}$ by

$$w_1 \succeq^u w_2 \text{ if } u(p) > u(q), w_2(p) \geq w_2(q) \Rightarrow w_1(p) \geq w_1(q)$$

The relation $w_1 \succeq^u w_2$ (to be read as "w_1 is closer to u than w_2") means that w_1 prescribes the same choice as u among pairs of lotteries at least as often as w_2.

Definition 2.5. Consider a Random Strotz representation (u, μ) and a Random Expected Utility representation (ν, τ). (i) (u, μ, ν, τ) is *sophisticated* if $\mu = \nu$; (ii) (u, μ, ν, τ) is *naive* if for any $w \in \mathcal{W}$, $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$ with strict inequality for some w.

The definition of naivete applies a notion of first-order stochastic dominance along the order $\geq^u$: it says that *ex ante* beliefs systematically overestimate the probability with which *ex post* choices agree with the long-term preference u.

2.4.2 Continuous-intensity Random Strotz model

This subsection provides conditions under which the measure μ admits an intuitive decomposition. Let $\mathcal{V} = \{v \in \mathcal{W}, u.v = 0\}$. Basic linear algebra results show that any element of $\mathcal{W} \backslash \{u, -u\}$ can be written under the form $au + \sqrt{1-a^2}v$, where $v \in \mathcal{V}$ and $a \in (-1, 1)$, a and v being unique.

The following result, due to Dekel and Lipman (2012), characterizes the sets of utility functions that are closed under $\geq^u$.

Lemma 1. *$w_1 \geq^u w_2$ if and only if there exists $v \in \mathcal{V}$ and coefficients $a_1 \geq a_2$ such that $w_1 = a_1 u + \sqrt{1-a_1^2}v$ and $w_2 = a_2 u + \sqrt{1-a_2^2}v$.*

Hence, fixing $v \in \mathcal{V}$, the set $\{au + \sqrt{1-a^2}v, a \in [-1,1])\}$ can be completely ordered according to $\geq^u$, the ranking being given by the coefficients a. Conversely, two elements of $\mathcal{W}$ can be ranked if and only if they belong to such a set. If $w = au + \sqrt{1-a^2}v$, v denotes the *direction* of the temptation w, while a measures its *intensity* (higher values of a correspond to a lower intensity).

Given u, for any $v \in \mathcal{V}$ we write $\mathcal{C}_v = \{au + \sqrt{1-a^2}v, a \in (-1,1)\}$, and $\overline{\mathcal{C}}_v = \mathcal{C}_v \cup \{u, -u\}$ the closure of $\mathcal{C}_v$.

The set $\mathcal{W}$ can be written $\mathcal{W} = \bigcup_{v\in\mathcal{V}} \overline{\mathcal{C}}_v$. Each set $\overline{\mathcal{C}}_v$ identifies the direction of temptation v, and can be considered as a line parameterized by the intensity of temptation in that direction. I impose two additional conditions on $\succeq$: continuity of the representation and finiteness of the set of relevant directions. Dekel and Lipman (2012) show that these properties characterize the subclass of Random Strotz models that satisfy Stovall (2010)'s axioms.

Definition 2.6. (u, μ) is a *finite continuous-intensity Random Strotz representation* if: (i) there exists a collection of lower semi-continuous densities $\{\mu_v\}_{v\in\mathcal{V}}$ defined over $[-1,1]$ such that for any measurable E, we have

$$\mu(E) = \int_{v\in\mathcal{V}} \mu_v(\{a \in [-1,1] | au + \sqrt{1-a^2}v \in E\})\mu_{\mathcal{V}}(dv)$$

(ii) There exists a finite collection $\mathcal{F} \subseteq \mathcal{V}$ such that $\mu(\bigcup_{v\in\mathcal{F}} \overline{\mathcal{C}}_v) = 1$.

2.5 Unidimensional setting

A unidimensional setting is a particular case of a Random Strotz model in which $\{v \in \mathcal{V} | \mu(\mathcal{C}_v) > 0\}$ is a singleton. This refers to a situation where the agent knows the (unique) nature of her temptation but might be uncertain about its strength. This case plays a particular role in the subsequent analysis, since the behavioral axioms will take an appealing and intuitive form in a unidimensional model. This section provides a characterization of this property.

Definition 2.7. A Random Strotz representation (u, μ) is *unidimensional* if there exists $v \in \mathcal{V}$ such that $\mu(\overline{\mathcal{C}_v}) = 1$.

Axiom 2.1 (Ordered Temptations).
If $\{p\} \succ \{q_1\} \sim \{q_2\}$ for any $p \in x \cup y$, then $x \cup \{q_1\} \succ x \cup \{q_2\} \Rightarrow y \cup \{q_1\} \succeq y \cup \{q_2\}$

Axiom 2.1 means that any pair of normatively equivalent temptations $\{q_1, q_2\}$ can be ranked in terms of their desirability *ex post* independently of the menu on which they exert a temptation. Intuitively, this corresponds to a situation where all temptations are appealing *ex post* for the same reason, which is likely to be the case in most experimental settings. Appendix B contains a proof of the following representation theorem.

Theorem 2.1. *Suppose that $\succeq$ has a Random Strotz representation (u, μ). (u, μ) is unidimensional if and only if $\succeq$ satisfies axiom 2.1.*

Example To highlight the behavioral content of axiom 2.1, I provide here an example where it is violated. Suppose that a decision-maker has the option to commit to a schedule for her next working day, splitting her time into three activities: working, exercising and leisure. She anticipates two subjective states: one in which she is lazy to work (but enjoys exercising), and one in which she is lazy to exercise (but enjoys working). Denoting (a, b) an option that consists in working a hours and exercising during b hours, she might display the following preferences:

- According to the long-term preference, $(9, 1) \sim (5, 2) \succ (5, 0) \sim (5, 1)$.
- In state 1, $(5, 2) \succ (0, 1) \succ (9, 1) \succ (5, 0)$: the agent enjoys exercising but dislikes working. She maximizes the number of hours spent exercising, and then minimizes the number of hours spent working among the remaining options.
- In state 2, $(9, 1) \succ (5, 0) \succ (5, 2) \succ (0, 1)$: the agent enjoys working but dislikes exercising, and has lexicographic preferences that mirror state 1.

Her preferences satisfy $\{(9, 1)\} \succ \{(9, 1), (0, 1)\}$ but $\{(9, 1)\} \sim \{(9, 1), (5, 0)\}$ since $(5, 0)$ is never chosen against $(9, 1)$. Similarly, $\{(5, 2)\} \succ \{(5, 2), (5, 0)\}$ but $\{(5, 2)\} \sim \{(5, 2), (0, 1)\}$. These conditions together violate axiom 2.1. Intuitively, $(5, 0)$ is tempting with respect to $(5, 2)$ because of the laziness to exercise in state 2, while $(0, 1)$ is tempting with respect to $(9, 1)$ because of the laziness to work in state 1.

3 Naivete in the Random Strotz model

3.1 Representation result

Consider a lottery p and a menu y such that $\{p\} \succ \{q_1\} \sim \{q_2\}$ for all $q_1, q_2 \in y$. All the options in y have the same *ex ante* valuation and p is strictly preferred to any of them. The decision-maker's *ex ante* valuation associated with the menu $y \cup \{p\}$ depends on her subjective probability of choosing $\{p\}$ or an element of y. The subjective probability of choosing p equals $\mu(\{w \in \mathcal{W} | w(p) \geq \max_{q \in y} w(q)\})$. Let us write it $\alpha^{y \cup \{p\}}(p)$. Considering any $q \in y$, we have

$$V(y \cup \{p\}) = \alpha^{y \cup \{p\}}(p)u(p) + \alpha^{y \cup \{p\}}(y)u(q) \tag{3.1}$$

Consider now the singleton $\{\kappa p + (1-\kappa)q\}$, where $\kappa \in [0, 1]$. Since the decision-maker has no choice to make inside the menu, her corresponding valuation equals

$$V(\{\kappa p + (1 - \kappa)q\}) = \kappa u(p) + (1 - \kappa)u(q) \tag{3.2}$$

Suppose now that the decision-maker faces the choice between the whole menu, $y \cup \{p\}$, and the singleton $\{\kappa p + (1-\kappa)q\}$. Comparing equations 3.1 and 3.2 shows that her choice depends on the relative position of $\alpha^{y \cup \{p\}}(p)$ and κ: she prefers the singleton menu as soon as $\kappa > \alpha^{y \cup \{p\}}(p)$, i.e. as soon as the exogenous probability of delivering p exceeds her anticipated probability of choosing p from the whole menu. Varying the exogenous probability κ hence identifies how the decision-maker forecasts her future choices, and this information can be compared with her observed *ex post* choice $\lambda^{y \cup \{p\}}(p)$. Our first axiom exploits this idea to detect pessimistic forecasts.

Definition 3.1. The menu y is *homogeneous* if $\{p\} \sim \{q\}$ for any $p, q \in y$.

Axiom 3.1 (No Commitment to Inferior Lotteries).
If $\{p\} \succ \{q\}$, $q \in y$ and y is homogeneous, $\kappa < \lambda^{y \cup \{p\}}(p) \Rightarrow \{\kappa p + (1 - \kappa)q\} \prec y \cup \{p\}$

$\kappa p + (1 - \kappa)q$ is an *inferior lottery* as soon as $\kappa < \lambda^{y \cup \{p\}}(p)$. A rational decision-maker anticipates that $\{\kappa p + (1 - \kappa)q\}$ is inferior to the expected value that she derives from the whole menu, since κ is lower than her own probability of choosing p. The fact that the decision-maker discards all inferior lotteries rules out pessimistic expectations. This axiom is satisfied by sophisticated agents but also by naive individuals who hold optimistic beliefs.

Our second axiom complements axiom 3.1 and detects optimistic anticipations.

Axiom 3.2 (Commitment to Superior Lotteries).
If $\{p\} \succ \{q\}$, $p \in x$ and x is homogeneous, $\kappa > \lambda^{x\cup\{q\}}(x) \Rightarrow \{\kappa p + (1-\kappa)q\} \succ x \cup \{q\}$

$\kappa p + (1-\kappa)q$ is a *superior lottery* as soon as $\kappa > \lambda^{x\cup\{q\}}(x)$. This allows to discriminate between sophisticated and naive agents. A naive agent underestimates her propensity to self-indulge *ex post* and fails to accept superior commitments, misguided by the wrong belief that her *ex post* choice will outperform the proposed option.

Theorem 3.1 shows that axiom 3.1 and 3.2 characterize naivete and sophistication.

Theorem 3.1. *Suppose that $\succeq$ has a finite continuous-intensity Random Strotz representation (u, μ), and that λ has a Random Expected Utility representation (ν, τ). Then (i) (u, μ, ν, τ) is sophisticated if and only if $(\succeq, \lambda)$ satisfies axiom 3.1 and 3.2; (ii) (u, μ, ν, τ) is naive if and only if $(\succeq, \lambda)$ satisfies axiom 3.1 and violates axiom 3.2.*

Remark. The conditions in axioms 3.1 and 3.2 only need to be checked for menus x and y of size K, where K is the number of relevant directions in the support of μ. This property is exploited in subsection 3.4 in the unidimensional setting, i.e. in the case where $K = 1$.

3.2 Discussion

3.2.1 *Ex ante* realism vs optimism

This subsection shows how axioms 3.1 and 3.2 relate to more immediate but less easily testable behavioral definitions of naivete and sophistication.

Definition 3.2. Consider a menu x. $\{\succeq, \lambda\}$ is: (i) *realistic* at x if $x \sim \{\sum_{p\in x} \lambda^x(p)p\}$; (ii) *optimistic* at x if $x \succ \{\sum_{p\in x} \lambda^x(p)p\}$.

The lottery $\{\sum_{p\in x} \lambda^x(p)p\}$ can be interpreted as a certain equivalent of the menu x revealed by the actual choices made by the decision-maker from x. A *realistic* agent is indifferent between a menu x and its certain equivalent: she correctly anticipates that both menus deliver the same distribution over lotteries at the consumption stage. In contrast, an *optimistic* agent weakly prefers any menu to its certain equivalent since she believes that her choices from x will be better aligned with her *ex ante* preference than they actually are.

Proposition 3.1 states that realism and optimism characterize, respectively, sophistication and naivete. Together with theorem 3.1, this result shows that an experimenter can restrict attention to tests of the form given by axioms 3.1 and 3.2 to investigate the extent to which the decision-maker's expectations are optimistic.

Proposition 3.1. *(u, μ, ν, τ) is sophisticated if and only if $(\succeq, \lambda)$ is realistic at any menu. (u, μ, ν, τ) is naive if and only if $(\succeq, \lambda)$ is realistic or optimistic at any menu (and optimistic for at least one menu).*

3.2.2 Links with Ahn and Sarver (2013)

Ahn and Sarver (2013) provide a characterization of sophisticated behavior for a decision-maker who values flexibility and not commitment. They also compare preferences over menus, and stochastic choices inside menus, and they assume that the *ex ante* preference admits a Dekel et al. (2001) representation: the decision-maker anticipates a stochastic taste contingency, described by a set of subjective states, but in contrast to the present model there is no conflict of preference between *ex ante* and *ex post* choices, which implies that larger menus are always preferred. Ahn and Sarver (2013) prove that two conditions characterize the correspondence between anticipated and actual taste contingencies. The first one, a variant of *Consequentialism*, requires that options that are never chosen *ex post* are irrelevant *ex ante*: adding an option p to a menu x should not change the valuation of x if p is never chosen from $x \cup \{p\}$ *ex post*. Their second axiom, *Foreseen Contingencies* is the converse condition and ensures that options that are chosen *ex post* are relevant in the *ex ante* valuation.

Variants of these axioms are also necessary to obtain sophistication in the present model. For instance, sophistication implies

Axiom 3.3 (Consequentialism). $\lambda^{x \cup \{p\}}(p) = 0 \Rightarrow x \cup \{p\} \sim x$.

Ahn and Sarver (2013) show that *Consequentialism* and *Foreseen Contingencies* are necessary and sufficient conditions to find a sophisticated representation since subjective probabilities are not unique in Dekel et al. (2001)'s framework. In contrast, in the Random Strotz model more stringent conditions are needed to identify the probability associated with each subjective state, which allows us to test how the decision-maker *ex ante* expectations compare with her actual *ex post* choices.

3.2.3 Non-instrumental concerns

The Random Strotz model assumes that the decision-maker's preferences over menus only reflect her preferences over final consumption goods. It therefore rules out other phenomena that might influence her willingness to commit. First, individuals might have intrinsic preferences over the decision process itself. People might value the ability to make a decision themselves, irrespective of the outcomes obtained, as shown in the experiment by Bartling et al. (2014). In contrast, other authors postulate that making decisions is undesirable, because thinking is costly (Ortoleva, 2013; Ergin and Sarver, 2010) or because controlling one's impulses is unpleasant (see Baumeister et al., 2007; Gul and Pesendorfer, 2001, and section 4). Similarly, common sense suggests that self-esteem and reputation management (Bénabou and Tirole, 2004) might prevent people from choosing commitment options, since this decision reveals the existence of their self-control issues. However, one may also argue that failures to exert self-control at the *ex post* stage entails a large reputation cost, which might increase the willingness to commit. For instance, Exley and Naecker (2015) report the results of a field experiment in which the demand for commitment is higher when the choice is made in public rather than in private, which suggests that individuals signal something positive about themselves by restricting their options. All in all, it is therefore unclear whether non-instrumental concerns increase or decrease the desire to commit, and I leave these interesting questions for future research.

3.3 Absolute and relative measures of naivete

This subsection explores some properties of naive representations, defines a cardinal index of naivete that measures the gap between expected and realized choices as well as two comparative notions of naivete.

3.3.1 A local index of naivete

Definition 3.3. Consider a menu x, and suppose that $(\succeq, \lambda)$ is realistic or optimistic at x. Consider $\Delta(x)$ the set of lotteries defined over x, and the subset of $\Delta(x)$ defined by

$$\mathcal{N}^{\succeq,\lambda}(x) = \{\kappa \in \Delta(x) | \{\sum_{p \in x} \lambda^x(p)p\} \prec \{\sum_{p \in x} \kappa_p p\} \prec x\}$$

The *index of naivete* of $(\succeq, \lambda)$ at x is the (normalized) volume of $\mathcal{N}^{\succeq,\lambda}(x)$:

$$N^{\succeq,\lambda}(x) = \frac{\forall(\mathcal{N}^{\succeq,\lambda}(x))}{\forall(\Delta(x))}$$

A lottery κ belongs to $\mathcal{N}^{\succeq,\lambda}(x)$ if the agent should objectively commit to κ instead of x but naively refuses to do so. $N^{\succeq,\lambda}(x) \in [0,1]$ measures the disagreement between *ex ante* and *ex post* probabilities of choice in x. $N^{\succeq,\lambda}(x) = 0$ if $(\succeq, \lambda)$ is realistic at x, $N^{\succeq,\lambda}(x) > 0$ if $(\succeq, \lambda)$ is optimistic at x, and $N^{\succeq,\lambda}(x) = 1$ if the decision-maker does not anticipate preference reversals but always chooses the worst element of the set according to the *ex ante* preferences.

Remark. If $\{\succeq, \lambda\}$ is pessimistic at x, an index of pessimism can be defined in a similar way by reversing the inequalities in the definition.

Example Suppose that $K = 2$, and denote $p = (1,0)$ and $q = (0,1)$ the two degenerate lotteries over prizes. Suppose that $\{p\} \succ \{p,q\}$. For any lotteries z_1, z_2 such that $z_1(1) > z_2(1)$, we write $\alpha^{\{z_1,z_2\}}(z_1)$ the probability with which the decision-maker anticipates choosing z_1. We obtain $\alpha^{\{z_1,z_2\}}(z_1) = \alpha^{\{p,q\}}(p)$ and $\lambda^{\{z_1,z_2\}}(z_1) = \lambda^{\{p,q\}}(p)$. Hence, $N^{\succeq,\lambda}(\{z_1, z_2\})$ is constant over all pairs $\{z_1, z_2\}$ such that $z_1 \neq z_2$, and it equals $N^{\succeq,\lambda}(\{p,q\}) = \alpha^{\{p,q\}}(p) - \lambda^{\{p,q\}}(p)$. It measures the gap between the *ex ante* and *ex post* probabilities of choosing p, the normatively superior prize, over q. $N^{\succeq,\lambda}(\{p,q\}) = 0$ for a sophisticated agent, and $N(\{p,q\}) = 1$ at the limit when $\alpha^{\{p,q\}}(p) = 1$ and $\lambda^{\{p,q\}}(p) = 0$.

3.3.2 Comparative measures of naivete

This paragraph compares the accuracy of the anticipations held by two agents. To focus on how differences in metacognitions are related to differences in behavior, let us first restrict attention to pairs of agents who have the same *ex post* behavior λ but whose commitment preferences $\succeq_1$ and $\succeq_2$ are different. We will say that agent 1 is more naive than agent 2 if agent 1 is less willing to commit to singleton menus than agent 2, reflecting the fact that individual 2 has a greater awareness of her tendency to being tempted.

Definition 3.4. Suppose that $(\succeq_1, \lambda)$ and $(\succeq_2, \lambda)$ have naive representations. $(\succeq_1, \lambda)$ is *more naive* than $(\succeq_2, \lambda)$ if

$$\{p\} \succ_1 x \Rightarrow \{p\} \succ_2 x$$

This definition is equivalent to Dekel and Lipman (2012)'s definition of agent 2 being more temptation-averse than agent 1. With the additional assumption that both joint representations are naive, a higher aversion to temptation is naturally interpreted as a greater degree of sophistication in the present case. Our next proposition characterizes how this notion is reflected in terms of representation. A first immediate observation is that $\succeq_1$ and $\succeq_2$ have the same preference over singletons, which implies that $u_1 = u_2$. Moreover, as Dekel and Lipman (2012) show, agent 2's greater demand for commitment is equivalent to the fact that the beliefs held by agent 2 over the intensity of temptation first-order stochastically dominate the beliefs held by agent 1, in each direction.

These two equivalent properties can also be related to the absolute measure of naivete defined in 3.3. If agent 1 is more naive than agent 2, then the index of naivete of agent 1 exceeds the index of naivete of agent 2 at any menu x, reflecting the fact that individual 1 is uniformly more optimistic than individual 2. The reverse implication also holds under the assumption $u_1 = u_2$ (which is not guaranteed by the inequality $N^{\succeq_1,\lambda}(x) \geq N^{\succeq_2,\lambda}(x)$ for all x).

Proposition 3.2. *Suppose that $\succeq_1$ has a Random Strotz representation (u_1, μ_1), and $\succeq_2$ has a Random Strotz representation (u_2, μ_2).*

The following statements are equivalent: (i) $(\succeq_1, \lambda)$ is more naive than $(\succeq_2, \lambda)$; (ii) $u_1 = u_2$, and $\mu_1(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succ_u w\}) \geq \mu_2(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succ_u w\}$ for any $w \in \mathcal{W}$. In addition, assuming that $u_1 = u_2$, (i) and (ii) are equivalent to: (iii) for any menu x, $N^{\succeq_1,\lambda}(x) \geq N^{\succeq_2,\lambda}(x)$.

Another possibility is to compare pairs of agents who have the same *ex ante* preference $\succeq$, but whose *ex post* choices might differ. We will say that agent 1 is more naive than agent 2 if agent 1 chooses the tempting options as least as often as agent 2 does, the definition of a tempting object being given by the common preference $\succeq$. Since agent 1 and agent 2 have the same beliefs *ex ante*, this comparative property suggests that agent 1's beliefs over future taste contingencies are less accurate than agent 2's perception.

Definition 3.5. Suppose that $(\succeq, \lambda_1)$ and $(\succeq, \lambda_2)$ have naive representations. $(\succeq, \lambda_1)$ is *more naive* than $(\succeq, \lambda_2)$ if

$$\{p\} \succ \{q\} \text{ for all } p \in x, q \in y \Rightarrow \lambda_1^{x \cup y}(x) \leq \lambda_2^{x \cup y}(y)$$

Agent 1's greater tendency to self-indulge *ex post* is equivalent to the first-order stochastic dominance of the distribution of realized tastes ν_1 over ν_2 on the intensity scale, in each direction. These two properties are also equivalent to the uniform ranking of N_1 and N_2 over all menus.

Proposition 3.3. *Suppose that λ_1 and λ_2 have respective Random Expected Utility representations (ν_1, τ_1) and (ν_2, τ_2) and satisfy assumption 1.*

The following statements are equivalent: (i) $(\succeq, \lambda_1)$ is more naive than $(\succeq, \lambda_2)$; (ii) For any $w \in \mathcal{W}, \nu_1(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq_u w\}) \leq \nu_2(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq_u w\}$; (iii) for any menu x, $N^{\succeq,\lambda_1}(x) \geq N^{\succeq,\lambda_2}(x)$.

3.4 Example: a unidimensional setting

3.4.1 Measuring naivete

The previous results take a simple and intuitive form in a unidimensional framework. This property is useful since, as discussed above, most experimental settings are likely to belong to this class. Axioms 3.1 and 3.2 can be summarized by:

Axiom 3.4 (Sophistication for pairs).
If $\{p\} \succ \{q\}$, $\kappa < \lambda^{\{p,q\}}(p) \Leftrightarrow \{\kappa p + (1 - \kappa)q\} \prec \{p, q\}$

If $\succeq$ is unidimensional, axiom 3.4 is a necessary and sufficient condition for the sophistication of the pair $(\succeq, \lambda)$. Hence, the experimenter can restrict attention to pairs of lotteries instead of considering larger menus. Suppose that p and q are such that $\{p\} \succ \{p, q\}$. Since $V(\{p, q\}) = \alpha^{\{p,q\}}(p)u(p) + \alpha^{\{p,q\}}(q)u(q)$, the value κ such that the decision-maker is indifferent between $\{\kappa p + (1 - \kappa)q\}$ and $\{p, q\}$ identifies $\alpha^{\{p,q\}}(p)$. This value can be elicited by an adapted Becker-DeGroot-Marschak mechanism. Comparing this threshold with $\lambda^{\{p,q\}}(p)$ reveals the nature of *ex ante* beliefs at $\{p, q\}$: naive if $\alpha^{\{p,q\}}(p) > \lambda^{p,q}(p)$, pessimistic if $\alpha^{\{p,q\}}(p) < \lambda^{p,q}(p)$, and sophisticated if $\alpha^{\{p,q\}}(p) = \lambda^{p,q}(p)$.

3.4.2 Quasi-hyperbolic discounting

A particular example of a unidimensional setting is given by the (β, δ) framework (Laibson, 1997). Consider the case where the prizes are consumption streams over an infinite horizon $\{0, 1, \cdots, t, \cdots\}$. A prize c is characterized by an infinite sequence $\{c_0, \cdots, c_t, \cdots\}$ of consumption levels. Suppose that the preference over singletons u can be represented by the standard discounted-utility model, and each *ex post* taste contingency belongs to the quasi-hyperbolic discounting class, as axiomatized by Olea and Strzalecki (2014):

- There exists $\delta \in [0,1]$ and a function $w : \mathbb{R} \to \mathbb{R}$ such that $u(c) = \sum_{t=1}^{+\infty} \delta^{t-1} w(c_t)$.
- There exists a measure $\mu : [0,1] \to \mathbb{R}$ such that

$$V(x) = \int_{\beta=0}^{1} \mu(\beta) \max_{c \in \mathcal{M}_{v_\beta}(x)} u(c) d\beta$$

where $v_\beta(c) = w(c_1) + \beta \sum_{t=1}^{+\infty} \delta^{t-1} w(c_t)$.

- There exists a measure $\nu : [0,1] \to \mathbb{R}$ and a tie-breaker τ such that

$$\forall y \subseteq x, \lambda^x(y) = \int_{\beta=0}^{1} \tau_\beta(\{\tilde{\beta} \in [0,1] | \mathcal{M}_{v_{\tilde{\beta}}}(\mathcal{M}_{v_\beta}(x)) \in y\}) d\nu(\beta)$$

This framework is a particular case of a unidimensional setting in which β parameterizes the intensity of temptation. The definitions and results provided above admit the following forms:

(i) Sophistication is equivalent to the identity $\mu = \nu$. A naive joint representation is such that μ strictly dominates ν at the first-order. For instance, if μ and ν are Dirac distributions respectively on $\hat{\beta}$ and β, sophistication is equivalent to $\hat{\beta} = \beta$, while naivete is equivalent to $\hat{\beta} > \beta$.

(ii) If $\succeq_1$ is represented by μ_1 and $\succeq_2$ by μ_2, $(\succeq_1, \lambda)$ is more naive than $(\succeq_2, \lambda)$ if μ_1 dominates μ_2. If λ_1 is represented by ν_1 and λ_2 by ν_2, $(\succeq, \lambda_1)$ is more naive than $(\succeq, \lambda_2)$ if ν_2 dominates ν_1.

(iii) If $\{c\} \succ \{c, c'\}$, the index of naivete at the set $\{c, c'\}$ equals $\mu([\beta^*, 1]) - \nu([\beta^*, 1])$, where β^* is defined by $v_{\beta^*}(c) = v_{\beta^*}(c')$. Thus, the index of naivete at a pair measures the distance between the cumulative distribution functions over β measured at the switching point between the two elements of the pair.

4 Naivete in the Random Gul-Pesendorfer model

The Random Strotz model provides a possible interpretation of the behavior of a decision-maker who values smaller menus. Nevertheless, other representations of the desire for commitment are conceivable. This section studies how the results of section 3 are modified if $\succeq$ is rationalized by a Random Gul-Pesendorfer model[2].

4.1 Random Gul-Pesendorfer model

Definition 4.1. The preference relation $\succeq$ admits a *Random Gul-Pesendorfer representation* (Stovall, 2010) if there exists a nontrivial expected utility u, and a nontrivial measure η on $\mathcal{U}$ such that $\succeq$ is represented by the functional

$$V(x) = \int_{w\in\mathcal{U}} [\max_{p\in x} (u(p) + w(p)) - \max_{q\in x} w(q)]\eta(dw) \tag{4.1}$$

If η is a degenerate lottery, this definition comes down to the Gul-Pesendorfer model of temptation-driven preferences (Gul and Pesendorfer, 2001). In the subjective state w, the decision-maker trades off her long-term preference u against her short-term temptation w, choosing the element that maximizes $u+w$ in x, and incurring a self-control cost equal to $\max_{q\in x} w(q)$. In contrast with the Random Strotz model, a tempting option can lower the *ex ante* valuation of a menu even if it is never chosen, provided that its presence in the menu inflicts a self-control cost to the decision-maker. Equation 4.1 adds some uncertainty by considering that temptations are drawn according to a measure η (Stovall, 2010). In contrast to the Random Strotz representation, the Random Gul-Pesendorfer functional does not identify η: several functions of the form of equation 4.1 can rationalize the same preference (see Dekel and Lipman, 2012).

A notion of unidimensionality can be defined analogously to the Random Strotz model. The representation (u, η) is *unidimensional* if there exists $v \in \mathcal{U}$ such that $u.v = 0$ and the support of ν is included in $\{\alpha u+\beta v, \beta \geq 0\}$. Notice that definition 4.1 does not impose any normalization condition on u nor on the temptations v. This explains why η is defined over $\mathcal{U}$ and not $\mathcal{W}$ and why the equivalence classes of $\succeq^u$ now take the form $\{\alpha u + \beta v, \beta \geq 0\}$.

4.2 Naivete

Defining naivete associated with a preference $\succeq$ requires a more robust definition than in section 3 to deal with the non-uniqueness of the representation. A set of equivalent Random Gul-Pesendorfer representations is classified as naive with

[2] Dekel and Lipman (2012) show that any preference with a Random Gul-Pesendorfer representation also admits a continuous-intensity Random Strotz representation, and vice versa.

respect to some *ex post* preference if all the representations belonging to this set are naive.

Definition 4.2. Consider a Random Gul-Pesendorfer representation (u, η) and a Random Expected Utility representation (ν, τ). (u, η, ν, τ) is *naive* if for any $\tilde{w} \in \mathcal{U}$, $\eta(\{w \in \mathcal{U} | (u + w) \succeq^u \tilde{w}\}) \geq \nu(\{w \in \mathcal{U} | w \succeq^u \tilde{w}\})$. Consider a set $\mathcal{E}$ of equivalent Random Gul-Pesendorfer models. $(\mathcal{E}, \nu, \tau)$ is *naive* if for any (u, η) belonging to $\mathcal{E}$, (u, η, ν, τ) is naive.

Theorem 4.1 is adapted from Dekel and Lipman (2012) and states that the Random Strotz representation of a preference $\succeq$ is less pessimistic about the intensity of future temptations than any of its equivalent Random Gul-Pesendorfer representations.

Theorem 4.1. *Suppose that the preference $\succeq$ admits a Random Strotz representation (u, μ) and a Random Gul-Pesendorfer (u, η). For any $\tilde{w} \in \mathcal{W}$*

$$\mu(\{w \in \mathcal{W} | w \succeq^u \tilde{w}\}) \leq \eta(\{w \in \mathcal{U} | (u + w) \succeq^u \tilde{w}\})$$

As a consequence, if (u, μ, ν, τ) is naive then the set of Random-Gul Pesendorfer models $\mathcal{E}$ associated with $\succeq$ is such that $(\mathcal{E}, \nu, \tau)$ is naive. The same conclusion holds if (u, μ, ν, τ) is sophisticated and $\succeq$ is temptation-averse.

An immediate implication is that the degree of naivete of a naive Random Strotz model is a lower bound of the degree of naivete of all corresponding Random Gul-Pesendorfer models. Intuitively, anticipating self-control costs *ex ante* reinforces the desire to commit; therefore commitment choices that appear too optimistic in light of *ex post* choices cannot be rationalized by assuming that the decison-maker was expecting decision costs. If the behavior of a decision-maker satisfies axiom 3.1 and violates axiom 3.2, theorem 4.1 shows that this finding is sufficient to conclude that all Random Gul-Pesendorfer representations of her behavior are also identified as naive. In that case, the Random Strotz interpretation of behavior is the conservative hypothesis regarding the degree of naivete attributed to the agent's behavior. In addition, Dekel and Lipman (2012) show that, except in the trivial case where temptation is not a concern, the Random Strotz model prescribes choices that are strictly more aligned with the long-term preference than any of its equivalent Random Gul-Pesendorfer models. Therefore, a pattern of choice that is sophisticated under the Random Strotz model cannot be rationalized by a Random Gul-Pesendorfer representation.

4.3 Sophistication

Suppose now that the observed choices satisfy axiom 3.2 and violate 3.1, in which case the Random Strotz representation is classified as pessimistic. This subsection studies under which conditions it is possible to rationalize the joint preferences by a Random Gul-Pesendorfer model of commitment preferences consistent with the hypothesis of sophistication. For the sake of simplicity, we restrict attention to the simplest possible setting with only two goods. I assume that η and ν have finite support[3] written $\{v_s\}_{s\in S}$ and $\{w'_s\}_{s'\in S'}$. Since only two subjective states are possible, ν is simply characterized by two values $\nu(\{u\})$ and $\nu(\{-u\})$ and two tie-breakers τ_u and τ_{-u}.

The definition of sophistication must be adapted to take into account the lack of normalization in equation 4.1: a sophisticated joint representation associates any *ex ante* subjective state with a unique *ex post* taste contingency that represents the same preference and occurs with the same probability.

Definition 4.3. Consider a finite Random Gul-Pesendorfer representation $(u, \{\eta_s, v_s\}_{s\in S})$ and a finite Random Expected Utility representation $\{\nu_{s'}, w_{s'}, \tau_{s'}\}_{s'\in S'}$. The pair $((u, \{\eta_s, v_s\}), \{\nu_{s'}, w_{s'}, \tau_{s'}\})$ is *sophisticated* if there exists a bijection $\phi : S \to S'$ such that for any s, $\eta_s = \nu_{\phi(s)}$ and $u + v_s$ and $w_{\phi(s)}$ represent the same preference.

As discussed above, consequentialism is not implied by sophistication, but an asymmetric version of this axiom is necessary. Axiom 4.1 states that options that are normatively preferred and chosen with positive probability are valued at the *ex ante* stage. It rules out the extreme pessimism of a decision-maker who incorrectly believes that she never chooses the normatively superior option *ex post*.

Axiom 4.1 (Uphill Consequentialism). If $\{p\} \succ \{q\}$ and $\lambda^{\{p,q\}}(p) > 0$, then $\{p, q\} \succ \{q\}$.

Provided that $(\succeq, \lambda)$ satisfies axiom 3.2, axioms 4.1 is the only revealed preference implication of sophistication. Proposition 4.1 states that any intermediate level of *ex ante* pessimism can be attributed to the presence of self-control costs that are not observed *ex post*.

Proposition 4.1. *Suppose that $|\mathcal{Z}| = 2$, that $\succeq$ admits a finite Random Gul-Pesendorfer representation and that λ admits a finite Random Expected Utility. $(\succeq, \lambda)$ admits a sophisticated Random Strotz or Random Gul-Pesendorfer representation if and only if $(\succeq, \lambda)$ satisfies axioms 3.2 and 4.1.*

[3] These properties are axiomatized in Stovall (2010) and Ahn and Sarver (2013) respectively.

Consider a simple experiment with two goods p and q. Suppose that the experimenter elicits the indifference threshold $\alpha^{\{p,q\}}(p)$ and the actual choice probability $\lambda^{\{p,q\}}(p)$ of a subject. As figure 1 shows, four cases arise: (i) if $\alpha^{\{p,q\}}(p) = \lambda^{\{p,q\}}(p)$, the joint behavior is rationalizable by a Random Strotz model; (ii) if $\alpha^{\{p,q\}}(p) > \lambda^{\{p,q\}}(p)$, the choices are naive under any interpretation; (iii) if $0 < \alpha^{\{p,q\}}(p) < \lambda^{\{p,q\}}(p)$, the joint preferences can be rationalized by a Random Gul-Pesendorfer interpretation; (iv) if $\alpha^{\{p,q\}}(p) = 0$, every interpretation concludes that the subject has pessimistic beliefs. In general, pessimism is therefore almost impossible to detect, and a representation that includes unobservable self-control costs can rationalize virtually any pattern of choice that would be considered as pessimistic under the consequential interpretation of commitment choices.

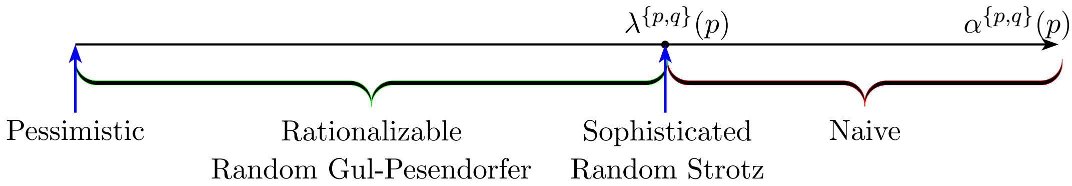

Figure 1: Classification of behavior in the two-goods case

5 Experiment

This section describes an experimental design that builds on the theoretical analysis to elicit naivete at the individual level. The experimental procedure involves a task where preference reversals are not due to temptation but to memory failures.

5.1 Experimental procedure

5.1.1 Task

Participants have the choice between earning a monetary prize $p =$ "X dollars" or earning nothing, $q =$ "0 dollars". Preference reversals do not arise because subjects value p and q differently at different points in time, since we can expect them to always rank p above q. However, the *ex post* choice between p and q is made in the future and participants receive the prize p if they make an active choice at a given date, otherwise they receive the default q. A subject who forgets to claim the monetary prize p therefore behaves as if she ranked q above p at the time of the choice.

At the initial stage of the experiment, subjects are offered the choice between the whole choice set $\{p, q\}$ or a commitment device of the form $\{\kappa p + (1 - \kappa)q\}$, where κ takes values from 0 to 1. It is reasonable to expect participants to exhibit the following preferences:

$$\{p\} \succeq \{p, q\} \succeq \{q\} \tag{5.2}$$

The commitment device $\{p\}$ always delivers the monetary prize, while the singleton $\{q\}$ delivers zero dollars with certainty. In between, the decision-maker's valuation of the choice set $\{p, q\}$ depends on her subjective probability of remembering to claim p in the future. Equation 5.2 shows that, while self-control problems and memory issues correspond to different psychological phenomena, the Random Strotz interpretation of behavior is equally suitable to both situations.

The experimental test consists in eliciting for each participant the value $\kappa^*(p, q)$ such that $\{p, q\}$ and $\{\kappa^*(p, q)p + (1 - \kappa^*(p, q))q\}$ are valued similarly at the initial stage. $\kappa^*(p, q)$ is interpreted as the participant's subjective probability of remembering to claim the prize p at the *ex post* stage. Comparing $\kappa^*(p, q)$ with $\lambda^{\{p,q\}}(p)$, the participant's actual probability of remembering to choose the monetary reward, identifies whether the subject behaves in a sophisticated, naive or pessimistic manner.

5.1.2 Recruitment and instructions

Subjects were recruited on Mechanical Turk, an online labor platform where individuals perform Human Intelligence Tasks on their personal computer in exchange of monetary rewards. Requesters can propose tasks with a fixed payment and award bonuses depending on the quality of the worker's answers. The identity of the workers is entirely anonymous, since they are identified with a personal ID given by the website.

The lack of control over the conditions in which the subjects answer the questions might be problematic to interpret the data. For this reason, participants were asked two questions aimed at verifying their understanding of the protocol. Correctly answering both questions was necessary to receive the baseline participation fee and to be allowed to proceed with the experiment. Overall, 95% of participants correctly answered both questions, suggesting that understanding issues represent a minor problem. Recent research using Mechanical Turk has shown that the quality of answers gathered on online labor markets is not significantly different from traditional laboratory experiments (Horton et al., 2011).

The task was described to the workers as an economics experiment on intertemporal decision-making. Workers who chose to participate received a link to an external website containing the experimental instructions and the answer forms. Participants were identified by means of their Mechanical Turk ID and received a personal code randomly generated to validate their participation on the Mechanical Turk platform. All payments were made on the platform by the intermediary of Amazon services.

Initial session After providing informed consent, participants were informed that they would have the opportunity to earn a monetary reward every day during 10 consecutive days. Each session consisted in a 24 hours-slot during which subjects had the possibility to add the monetary prize p to their earnings by simply signing in the experimental website with their Mechanical Turk identifier. Participants were informed that they would not receive any reminder from the experimenter, and nothing was said about the use of artificial devices. They were invited to write down the URL of the website. They also had the possibility to contact the experimenter at any moment through the Mechanical Turk platform to ask for the URL.

Commitment choices Subjects were offered the possibility to modify the payment rules for one of the ten sessions, the other nine sessions remaining unchanged. They were asked to report their preference between: (i) choosing later, that is receiving the prize only conditional on signing in; (ii) being paid with probability κ, irrespective of their behavior that day. The parameter κ took 21 values for all the multiples of 5 from 0% to 100%. For each of these values, the participant had to choose between options (i) and (ii). One of these rows was randomly selected and the corresponding choice was implemented, thereby ensuring the incentive-compatibility of the elicitation method (Azrieli et al., 2015).

The date at which the commitment choice was relevant was selected randomly among the 10 possible session dates, and participants did not learn which date had been chosen until the corresponding day. This procedure eliminates the effect of any private information that subjects might have regarding the evolution of their probability of remembering over time: for instance, a sophisticated participant who always remembers to log in to the website on the first day, but who always forgets to do so after that, would strictly prefer $\{p, q\}$ to $\{0.95p + 0.05q\}$ if this choice were implemented on the first day, even though her actual frequency of visit across all sessions only equals 10%. Implementing the payment rule at a random date makes sure that the participants should report their subjective probability of remembering to choose p across the 10 sessions, which is the value that is estimated from their subsequent behavior.

Attention questions and simple checks Before submitting their choices, participants were required to answer two basic questions to verify that they read and understood the instructions. Questions were based on hypothetical scenarios: participants were asked how much they would earn depending on the row selected, their choice between the two menus, and their behavior that day. Subjects were informed that any wrong or missing answer would prevent them from proceeding with the experiment. In contrast, subjects who provided correct answers received the baseline participation fee and were allowed to proceed with the 10 regular sessions.

A simple test of understanding and rationality can also be performed by observing the commitment choices: as κ goes down, each participant should have at most one switching point from $\{\kappa p + (1-\kappa)q\}$ to $\{p, q\}$. For each participant who satisfies this criterion, we record the values $(\underline{\kappa}, \overline{\kappa})$ such that $\{\overline{\kappa}p + (1-\overline{\kappa})q\} \succeq \{p, q\} \succeq \{\underline{\kappa}p + (1-\underline{\kappa})q\}$: the subjective belief of the agent is partially identified in the interval $[\underline{\kappa}, \overline{\kappa}]$.

Regular sessions The 10 regular sessions took place on the 10 days that followed the initial stage. If a participant signed in to the website during a session, a confirmation message displayed: her earnings for the session, explaining the outcome of the payment rule if it was implemented that day; her total earnings so far; the dates of the remaining sessions. For each of the 10 sessions, a dummy variable records the agent's behavior, and takes value 1 if she logged in during the session and 0 otherwise. The sum of these 10 variables λ^* yields the actual frequency with which the subject remembered to participate in the session.

Payment The baseline participation fee of \$1 was paid in the 12 hours that followed the initial session. In one condition, the monetary prize was equal to \$0.25; in the other treatment, it was equal to \$0.4. These values might appear very small, but they yield approximate hourly wages of respectively \$14 and \$20 dollars (\$3.5 or \$5 for 15 minutes of participation in total) for a subject who would pass the initial test and log in to the website every day. This amount is substantially higher than wages usually proposed on the platform (\$1.38 per hour as reported by Mason and Suri (2012)).

The earnings corresponding to the 10 regular sessions were paid the day after the last session under the form of a bonus on the Mechanical Turk platform. Daily payments were not provided regularly because participants who had forgotten about a session would have receive a payment if their commitment choice had been successful, which would have played the role of a reminder.

5.2 Results

5.2.1 Sample

A total of 143 subjects participated in the initial stage of the experiment in June 2015. 7 participants failed the attention tests and were not allowed to proceed. In addition, 3 participants provided multiple switch points and are excluded from the analysis. The final sample includes 133 participants. Average earnings amount to $2.73 per person including the participation fee.

5.2.2 Raw data

This subsection provides some preliminary data comparing commitment choices with frequencies of visit.

Aggregate level The average anticipated probability of remembering to visit the website lies in the range 0.84-0.88 (std=0.20), compared with an average frequency of visit of 0.47 (std=0.45) averaged across participants and sessions. The distributions of *ex ante* and *ex post* probabilities are displayed in figure 2.

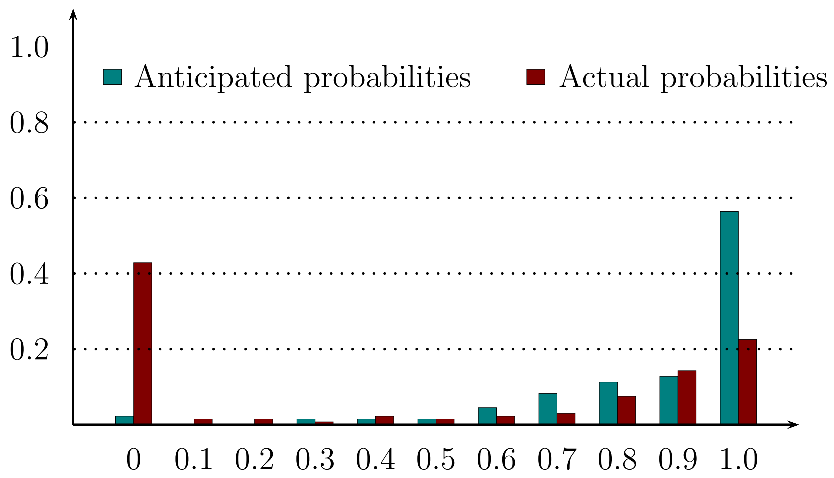

Figure 2: Distributions of anticipated and actual probabilities of visit

Individual level Each participant is characterized by a set of admissible *ex ante* beliefs $[\underline{\kappa}, \overline{\kappa}]$ and a frequency of visit λ^*. A subject's choices are classified as: naive if $\underline{\kappa} > \lambda^*$; pessimistic if $\overline{\kappa} < \lambda^*$; sophisticated otherwise. Table 1 reports the number and the fraction of participants in each category as a function of the payment. Overall, 66% of the subjects made naive choices, 22% made sophisticated decisions while 12% only choose inferior commitment devices. Among the subjects who made sophisticated choices, 79% exhibited extreme values of λ^* ($\lambda^* = 0$ or $\lambda^* = 1$) and correctly anticipated this.

	Naive choices	**Sophisticated choices**	**Pessimistic choices**	**Total**
$\boldsymbol{p}$ **= \$0.25**	24 (0.69)	6 (0.17)	5 (0.14)	35
$\boldsymbol{p}$ **= \$0.4**	64 (0.65)	23 (0.23)	11 (0.11)	98
Total	88 (0.66)	29 (0.22)	16 (0.12)	133

Table 1: Number of choices per category
(in parentheses, the fraction per category)

Measure of naivete The individual index of naivete equals $\underline{\kappa} - \lambda^*$ if the individual is naive. The average index among agents who made naive choices equals 0.62 (std=0.35), which means that these subjects on average overestimated by 62 percentage points their probability of visiting the website. This overconfidence has substantial economic consequences: the probability of receiving the prize for agents with similar memory issues but who would make the right *ex ante* decisions is 25 percentage points higher.

5.2.3 Statistical procedure

Independent events To provide a more elaborate elicitation of naivete, I first assume that participants view their future choices during the 10 sessions as independent variables drawn from the same distribution: they believe that their probability of visiting the website at any given day is independent of the day and of their behavior so far. In that case, the 10 dummy variables are independent realizations of a Bernoulli random variable. If the individual made a naive choice, the null hypothesis is that the parameter of the Bernoulli distribution equals $\underline{\kappa}$. A one-tailed binomial test consists in computing the probability with which, under this null hypothesis of sophistication, the decision-maker's frequency of choosing p

	H_0 **rejected (naivete)**	H_0 **not rejected**	H_0 **rejected (pessimism)**	**Total**
$p = \$0.25$	19 (0.54)	13 (0.34)	3 (0.11)	35
$p = \$0.4$	46 (0.47)	50 (0.51)	2 (0.02)	98
Total	65 (0.49)	63 (0.47)	5 (0.05)	133

Table 2: Test at the individual level
(in parentheses, the fraction per category)

is smaller than or equal to λ^*. Similarly, for an individual who made a pessimistic choice, the test consists in computing the probability with which, under the null hypothesis of correct beliefs equal to $\overline{\kappa}$, the individual visits the website with a frequency greater than or equal to λ^*.

The results of the test are reported in table 2. Overall, the data rules out sophistication for a large fraction of agents: at the 1% significance level, the hypothesis is rejected for 65 optimistic agents (49% of the sample).

Correlated events A possible caveat with the above test is that the 10 sessions might not appear independent to the participants. For instance, a decision-maker might believe that she will either participate to all sessions or forget entirely about the experiment, both events happening with the same probability. In that case, her *ex ante* subjective probability of visiting the website at a random date equals 0.5, but she would half of the time exhibit $\lambda^* = 0$ and make a naive choice. This issue brings the general problem of estimating individual choice probabilities, which requires to observe a subject making repeated decisions, in which case interdependencies between choices are likely to create confounds.

If the hypothesis of independent events is relaxed, nothing can be said at the individual level except for extreme values of $(\underline{\kappa}, \overline{\kappa})$ and λ^*: for instance, choices given by $\{p\} \sim \{p, q\}$ and $\lambda^* = 0$ indicate naive anticipations, but this pattern of choice only represents 13% of the data (17 participants). However, an aggregate test can be performed by observing the proportion of agents who made naive choices in the population. To obtain a conservative estimate for the proportion of naive subjects, I assume that all agents who made a sophisticated or a pessimistic choice are not naive. For each individual who made a naive choice that can be rationalized by a correlated beliefs structure, I compute the beliefs that rationalizes her joint behavior and that maximizes the probability with which she makes choices that appear naive *ex post*. According to this procedure, each subject is characterized by a probability of making naive, sophisticated or pessimistic choices: for instance, the individual described above exhibits $\lambda^* = 0$ half of the time (naive choice) and $\lambda^* = 1$ half of the time (pessimistic choice). More precisely, for each individual such that $\lambda^* < \underline{\kappa}$ I assume that her *ex ante* beliefs are given by

$$\mathbb{P}(\lambda = 1) = \frac{\underline{\kappa} - \lambda^*}{1 - \lambda^*} \text{ and } \mathbb{P}(\lambda = \lambda^*) = \frac{1 - \underline{\kappa}}{1 - \lambda^*}$$

This beliefs structure yields a subjective probability of visit equal to $\underline{\kappa}$ and a positive probability of signing in with frequency λ^*. It also suggests that the individual makes a naive choice with probability $\frac{1-\underline{\kappa}}{1-\lambda^*}$ and a pessimistic choice with probability $\frac{\underline{\kappa}-\lambda^*}{1-\lambda^*}$.

Under the hypothesis of sophistication, each individual is characterized by a probability of displaying a naive joint behavior. The number of naive choices in the population follows a Poisson-Binomial distribution whose vector of parameters is given by the individual probabilities. Given the number of naive choices in the data (88 out of 133), all hypotheses of the form "x% of the population is naive" are rejected at the 1% significance level for any x lower than 56%. These results suggest a large prevalence of naive anticipations in the population.

Appendix A: proofs of section 3 and 4

Notation

In the following, we denote by $\mathbb{1} = (\frac{1}{|\mathcal{Z}|}, ..., \frac{1}{|\mathcal{Z}}|)$ the (scaled) unit vector. For any subset (a, b) of $[-1, 1]$ and any $v \in \mathcal{V}$, we write $\mathcal{C}_v(a, b) = \{cu + \sqrt{1 - c^2}v | a < c < b\}$ and $\overline{\mathcal{C}}_v(a, b) = \{cu + \sqrt{1 - c^2}v | a \leq c \leq b\}$. Let us also define $\mathcal{C}(a, b) = \bigcup_{v \in \mathcal{V}} \mathcal{C}_v(a, b)$ and $\overline{\mathcal{C}}(a, b) = \bigcup_{v \in \mathcal{V}} \overline{\mathcal{C}}_v(a, b)$.

In all this section, $\succeq$ has a finite continuous Random Strotz representation (u, μ) and λ has a Random Expected Utility representation (ν, τ). In addition, assumption 1 is satisfied.

1 Proof of theorem 3.1

1.1 Preliminary results

Lemma A.2. *$(\succeq, \lambda)$ satisfies axiom 3.1 if and only if for any $w \in \mathcal{W}$, $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$.*

Proof. Let us start with the "if" part. Take p and a homogeneous menu y such that $\{p\} \succ \{q\}$ for all $q \in y$.

Define, for $v \in \mathcal{V}$, $a(v) = \sup\{a \in [-1, 1] | au(p) + \sqrt{1 - a^2}v(p) \geq \max_{q \in y} au(q) + \sqrt{1 - a^2}v(q)\}$ and $w(v) = a(v)u + \sqrt{1 - a(v)^2}v$. If $w = au + \sqrt{1 - a^2}v$, $w(p) \geq \max_{q \in y} w(q)$ is equivalent to $a \geq a(v)$, i.e. to $w \succeq^u w(v)$. Thus, we have

$$V(y \cup \{p\}) = u(p) \int_{v \in \mathcal{V}} \mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w(v)\}) dv + u(q) \int_{v \in \mathcal{V}} \mu(\{\tilde{w} \in \mathcal{W} | w(v) \succ^u \tilde{w}\}) dv$$

where q is any element of y, while

$$\lambda^{y \cup \{p\}}(p) = \int_{v \in \mathcal{V}} \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w(v)\}) dv \leq \int_{v \in \mathcal{V}} \mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w(v)\}) dv$$

and hence $V(\{y \cup \{p\}\}) \geq \lambda^{y \cup \{p\}}(p)u(p) + \lambda^{y \cup \{p\}}(y)u(q)$

For $\kappa < \lambda^{y \cup \{p\}}(p)$, we obtain

$$\{\kappa p + (1 - \kappa)q\}) \prec \{\lambda^{y \cup \{p\}}(p)p + \lambda^{y \cup \{p\}}(y)q\} \preceq y \cup \{p\}$$

which proves axiom 3.1.

We prove the "only if" part by contradiction. Suppose that there exists $w \in \mathcal{W}$ such that $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) < \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$. Write $w = au + \sqrt{1 - a^2}v$, where $a \in [-1,1]$. This condition is equivalent to $\mu(\overline{\mathcal{C}}_v(a,1)) < \nu(\overline{\mathcal{C}}_v(a,1))$. Therefore, by the continuity of μ and by theorem 10.2 of Billingsley (2012) there exists $a^* \in (a,1)$ such that $\mu(\overline{\mathcal{C}}_v(a,1)) + \mu(\overline{\mathcal{C}}(a^*,1)) < \nu(\overline{\mathcal{C}}_v(a,1))$.

Consider $\alpha > 0$, $\beta > 0$ and $\phi > 0$. Take $\tilde{w} \in \mathcal{V} \setminus \{v\}$. Since $v \neq \tilde{w}$, by the Cauchy-Schwarz inequality we have $|v.\tilde{w}| < 1$, thus it is possible to find $\gamma_{\tilde{w}}$ such that $\beta > \gamma_{\tilde{w}} v.\tilde{w}$, and $\beta v.\tilde{w} < \gamma_{\tilde{w}}$.

Define now: (i) $p = \mathbb{1} + \phi(\alpha u + \beta v)$; (ii) $q_v = \mathbb{1} + \phi\gamma_v v$ where γ_v is chosen to satisfy $au(p) + \sqrt{1-a^2}v(p) = au(q_v) + \sqrt{1-a^2}v(q_v)$; (iii) for all $\tilde{w} \in \mathcal{V}$ such that $\mu(\mathcal{C}_{\tilde{w}}) > 0$ and $\tilde{w} \neq v$, $q_{\tilde{w}} = \mathbb{1} + \phi\gamma_{\tilde{w}}\tilde{w}$. Consider finally the menu $y = \{q_v\} \cup \bigcup_{\tilde{w}\neq v}\{q_{\tilde{w}}\}$. y is finite since μ has a finite number of directions. Notice that $u(q_{\tilde{w}}) = u(q_v) = 0$ for all $\tilde{w}$, and that $u(p) > 0$.

We have $w(p) = w(q_v)$ by definition of γ_v. Moreover,

$$\begin{cases} w(p) = \phi(\alpha a + \sqrt{1-a^2}\beta) \\ w(q_{\tilde{w}}) = \phi\sqrt{1-a^2}\gamma_{\tilde{w}} v.\tilde{w} \text{ for } \tilde{w} \neq v \end{cases}$$

Since $\beta > \gamma_{\tilde{w}} v.\tilde{w}$, it is possible to pick α low enough to ensure that $w(p) \geq w(q_{\tilde{w}})$. In that case, on the set $\mathcal{C}_v$, p is chosen in the set $y \cup \{p\}$ if and only if the intensity of temptation lies in $[a,1]$[4]. Therefore $\lambda^{y\cup\{p\}}(p) \geq \nu(\overline{\mathcal{C}}_v(a,1))$.

Finally, suppose that $\hat{w} = \hat{a}u + \sqrt{1-(\hat{a})^2}\tilde{w}$, where $\tilde{w} \neq v$. We have

$$\begin{cases} \hat{w}(p) = \phi(\alpha\hat{a} + \sqrt{1-(\hat{a})^2}\beta v.\tilde{w}) \\ \hat{w}(q_{\tilde{w}}) = \phi\sqrt{1-(\hat{a})^2}\gamma_{\tilde{w}} \end{cases}$$

Since $\beta v.\tilde{w} < \gamma_{\tilde{w}}$, it is possible to pick α low enough to obtain $\hat{w}(p) < \hat{w}(q_{\tilde{w}})$ as soon as $\hat{a} < a^*$. This proves that p is chosen in $y \cup \{p\}$ at most on $\overline{\mathcal{C}}_v(a,1) \cup \overline{\mathcal{C}}(a^*,1)$.

Therefore

$$\begin{aligned} V(y \cup \{p\}) &\leq [\mu(\overline{\mathcal{C}}_v(a,1)) + \mu(\overline{\mathcal{C}}(a^*,1))]u(p) + [1 - \mu(\overline{\mathcal{C}}_v(a,1)) - \mu(\overline{\mathcal{C}}(a^*,1))]u(q) \\ &< \nu(\overline{\mathcal{C}}_v(a,1))u(p) + (1 - \nu(\overline{\mathcal{C}}_v(a,1)))u(q) \\ &\leq \lambda^{y\cup\{p\}}(p)u(p) + \lambda^{y\cup\{p\}}(y)u(q) \end{aligned}$$

It is sufficient to take $\kappa \in (\mu(\overline{\mathcal{C}}_v(a,1)) + \mu(\overline{\mathcal{C}}(a^*,1)), \nu(\overline{\mathcal{C}}_v(a,1)))$ to obtain a violation of axiom 3.1. □

Lemma A.3. $(\succeq, \lambda)$ *satisfies axiom 3.2 if and only if for any $w \in \mathcal{W}$,*

[4] Remember that λ picks a maximizer of u in case of indifference, and p is the unique maximizer of u in $y \cup \{p\}$.

$\mu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\})$.

Proof. We skip the proof. The arguments are similar to the demonstration of lemma A.2. $\square$

1.2 Proof of necessity

Suppose that (u, μ, ν, τ) is sophisticated, i.e. that $\mu = \nu$. The conditions of lemmas A.3 and A.2 are trivially true. Therefore, axioms 3.1 and 3.2 are satisfied.

Suppose now that (u, μ, ν, τ) is naive. The condition of lemma A.2 is satisfied, therefore axiom 3.1 is valid. In addition, the continuity of $\succeq$ yields $\nu(\{u\}) = 0$.

Suppose that (u, μ, ν, τ) is strictly naive, i.e. that there exists $w \in \mathcal{W}$ such that

$$\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) > \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) \tag{A.1}$$

Our next step is to show that this particular w satisfies

$$\mu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\}) < \nu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\})$$

and to use lemma A.3 to conclude that axiom 3.2 is violated.

Suppose, in contrast, that

$$\mu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\}) \tag{A.2}$$

Summing A.1 and A.2 yields

$$\mu(\mathcal{C}_v) > \nu(\{-u\}) + \nu(\mathcal{C}_v) \tag{A.3}$$

Take $\tilde{w} \in \mathcal{V} \setminus \{v\}$. Since (u, μ, ν, τ) is naive, $\mu(\overline{\mathcal{C}}_{\tilde{w}}(b, 1)) \geq \nu(\overline{\mathcal{C}}_{\tilde{w}}(b, 1))$ for all $b > -1$. Taking the limit when b tends to -1 shows that

$$\mu(\mathcal{C}_{\tilde{w}}) \geq \nu(\mathcal{C}_{\tilde{w}}) \tag{A.4}$$

Integrating A.4 and summing with A.3 yields

$$\begin{aligned} \int_{\hat{w} \in \mathcal{W}} d\mu(\hat{w}) &= \mu(\mathcal{C}_v) + \int_{\tilde{w} \in \mathcal{V} \setminus \{v\}} \mu(\mathcal{C}_{\tilde{w}}) d\tilde{w} \\ &> \nu(\{-u\}) + \nu(\mathcal{C}_v) + \int_{\tilde{w} \in \mathcal{V} \setminus \{v\}} \nu(\mathcal{C}_{\tilde{w}}) d\tilde{w} = \int_{\hat{w} \in \mathcal{W}} d\nu(\hat{w}) \end{aligned}$$

which is impossible since $\int_{\hat{w} \in \mathcal{W}} d\mu(\hat{w}) = \int_{\hat{w} \in \mathcal{W}} d\nu(\hat{w}) = 1$.

1.3 Proof of sufficiency

Suppose that $(\succeq, \lambda)$ satisfies axioms 3.1 and 3.2.

By lemma A.2 we have $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$ for any $w \in \mathcal{W}$, and by A.3 $\mu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\}) \geq \mu(\{\tilde{w} \in \mathcal{W} | w \succ^u \tilde{w}\})$. We further obtain $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) = \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$ using the same argument as in the proof of necessity above. Hence, μ and ν coincide on all sets that are closed and closed by $\succeq^u$. Dekel and Lipman (2012) show that this is a sufficient condition for the measures μ and ν to coincide on all Borel sets (see their proof of theorem 1). This proves that $\mu = \nu$.

Suppose now that $(\succeq, \lambda)$ satisfies axiom 3.1 but violates axiom 3.2. We have $\mu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\}) \geq \nu(\{\tilde{w} \in \mathcal{W} | \tilde{w} \succeq^u w\})$ for every w. If this holds with equality for every w, this implies $\mu = \nu$, which contradicts the fact that $(\succeq, \lambda)$ violates axiom 3.2. Therefore there exists w such that the inequality is strict, and the representation is naive.

2 Remaining proofs of section 3

2.1 Proof of proposition 3.1

We only prove the equivalence between realism and sophistication, the second part of the proposition being proved similarly. Suppose first that $\{\succeq, \lambda)$ is realistic. Take a homogeneous menu y, $q \in y$, p such that $\{p\} \succ \{q\}$ and $\kappa < \lambda^{y \cup \{p\}}$. We have

$$\begin{aligned} V(y \cup \{p\}) &= V(\{\lambda^{y \cup \{p\}}(p)p + \sum_{\tilde{q} \in y} \lambda^{y \cup \{p\}}(\tilde{q})\tilde{q}\}) \text{ since } (\succeq, \lambda) \text{ is realistic at } y \cup \{p\} \\ &= \lambda^{y \cup \{p\}}(p)u(p) + \lambda^{y \cup \{p\}}(y)u(q) \text{ since } y \text{ is homogenous} \\ &> \kappa u(p) + (1 - \kappa)q = V(\{\kappa p + (1 - \kappa)q\}) \end{aligned}$$

Hence, $\{\succeq, \lambda\}$ satisfies axiom 3.1. A similar proof shows that $\{\succeq, \lambda\}$ satisfies axiom 3.2. By theorem 3.1, $\{\succ, \lambda\}$ is sophisticated.

Suppose now that $\{\succeq, \lambda\}$ is sophisticated, and consider a finite menu x. x can be decomposed in k disjoint non-empty equivalence classes $\mathcal{E}_i, i = 1, \cdots, k$ such that $p \in \mathcal{E}_i, q \in \mathcal{E}_j$ imply $u(p) > u(q)$ if and only if $i < j$, and $u(p) = u(q)$ if $i = j$. It is therefore possible to define $\alpha^x(\mathcal{E}_i) = \mu(\{w \in \mathcal{W} | \mathcal{M}_u(\mathcal{M}_w(x)) \in \mathcal{E}_i\})$ the anticipated probability attached to the class $\mathcal{E}_i$[5]. Since $\mu = \nu$ and both measures

[5]The representation does not specify how the choice is made inside a class $\mathcal{E}_i$ between two

break ties in favor of u, it is easy to see that $\alpha^x(\mathcal{E}_i) = \lambda^x(\mathcal{E}_i)$ for any i. Writing p_i for any element of the class $\mathcal{E}_i$, we obtain

$$V(x) = \sum_{i=1}^{k} \alpha^x(\mathcal{E}_i)u(p_i) = \sum_{i=1}^{k} \lambda^x(\mathcal{E}_i)u(p_i) = V(\{\sum_{x\in p} \lambda^x(p)p\})$$

which proves that $x \sim \{\sum_{p\in x} \lambda^x(p)p\}$. Hence, $\{\succeq, \lambda\}$ is realistic.

2.2 Proof of proposition 3.2

$(i) \Leftrightarrow (ii)$ This result is proved by Dekel and Lipman (2012) in their Theorem 4, p. 1284.

$(i) \Rightarrow (iii)$ Consider a menu x, and a lottery $\kappa \in \Delta(x)$. Since $(\succeq_1, \lambda)$ is more naive than $(\succeq_2, \lambda)$,

$$\{\sum_{p\in x} \kappa_p p\} \prec_1 x \Rightarrow \{\sum_{p\in x} \kappa_p p\} \prec_2 x$$

Therefore $\mathcal{N}^{\succeq_2,\lambda}(x) \subseteq \mathcal{N}^{\succeq_1,\lambda}(x)$, which implies $N^{\succeq_2,\lambda}(x) \leq N^{\succeq_1,\lambda}(x)$.

$(iii) \Rightarrow (i)$ For $i = 1, 2$, we write V_i the functional associated with $\succeq_i$ and u the (common) normative utility function. Consider a menu x. A lottery κ on $\Delta(x)$ belongs to $\mathcal{N}^{\succeq_i,\lambda}(x)$ if and only if

$$\sum_{p\in x} \lambda^x(p)u(p) < \sum_{p\in x} \kappa_p u(p) < V_i(x)$$

Hence, from $N^{\succeq_1,\lambda}(x) \geq N^{\succeq_2,\lambda}(x)$ we obtain $V_1(x) \geq V_2(x)$. Now, for any lottery p we have

$$\{p\} \succ_1 x \Leftrightarrow u(p) > V_1(x) \Rightarrow u(p) > V_2(x) \Leftrightarrow \{p\} \succ_2 x$$

which proves that $(\succeq_1, \lambda)$ is more naive than $(\succeq_2, \lambda)$.

2.2.1 Proof of proposition 3.3

We skip the proof, which relies on the same arguments as proposition 3.2.

options that are equally valued at the *ex post* stage, but this choice is irrelevant here.

3 Proofs of section 4

3.1 Proof of theorem 4.1

The proof is given by Dekel and Lipman (2012) (see their theorem 5 p. 1286).

3.2 Proof of proposition 4.1

In the two goods case, only two generic utilities $u \in \mathcal{W}$ exist: $u = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ and $-u$. Therefore each subjective utility can be written $v_s = \alpha_s u$ where $\alpha_s \neq 0$. A Random Gul-Pesendorfer representation is simply written

$$V(x) = \sum_{s \in S} \eta_s [\max_{p \in x} (1 + \alpha_s) u(p) - \max_{q \in x} \alpha_s u(q)] \tag{A.5}$$

The decision utility $(1+\alpha_s)u$ coincides with u if and only if $\alpha_s > -1$. A Random Expected Utility model ν simply specifies the weights ν_u and ν_{-u} attached to the states u and $-u$ respectively.

Proof of necessity If$(\succeq, \lambda)$ admits a sophisticated Random Strotz representation, necessity of axioms 3.2 and 4.1 are established in section 3. Suppose that $(\succeq, \lambda)$ admits a sophisticated Random Gul-Pesendorfer representation, i.e. that there exists a representation A.5 of $\succeq$ such that $\sum_{s \in S, \alpha_s > -1} \eta_s = \nu_u$. Equation A.5 yields

$$V(\{p,q\}) = \sum_{s \in S, \alpha_s > 0} \eta_s u(p) + \sum_{s \in S, -1 < \alpha_s < 0} \eta_s [u(p) + \alpha_s (u(p) - u(q))] + \sum_{s \in S, \alpha_s < -1} \eta_s u(q) \tag{A.6}$$

If there exists p, q such that $\{p\} \succ \{q\}$ and $\lambda^{\{p,q\}}(p) > 0$ then $\nu_u > 0$. Therefore $\sum_{s \in S, \alpha_s > -1} \eta_s > 0$, and by equation A.6, $V(\{p,q\}) > V(\{q\})$, which proves that $(\succeq, \lambda)$ satisfies axiom 4.1.

In addition, since $u(p) + \alpha_s(u(p) - u(q)) < u(p)$ when $\alpha_s < 0$, equation A.6 yields

$$\begin{aligned} V(\{p,q\}) &\leq \sum_{s \in S, \alpha_s > -1} \eta_s u(p) + \sum_{s \in S, \alpha_s < -1} \eta_s u(q) \\ &\leq \nu_u u(p) + \nu_{-u} u(q) = V(\{\lambda^{\{p,q\}}(p) p + \lambda^{\{p,q\}}(q) q\}) \end{aligned}$$

which proves that axiom 3.2 holds.

Proof of sufficiency Suppose first that $\nu_u = 0$. In that case, for any p, q such that $\{p\} \succ \{q\}$ we have $\lambda^{\{p,q\}}(p) = 0$, which by axiom 3.2 implies $\{p, q\} \sim \{q\}$. This identifies the Random Strotz representation associated with $\succeq$, given by $V = \min u$. This function attaches a weight equal to 1 to the state $-u$, therefore the joint representation is sophisticated.

If $\nu_u > 0$, let us define $\eta_u = \sum_{s\in S,\alpha_s>0} \eta_s + \sum_{s\in S,0<\alpha_s<1} \eta_s\alpha_s$ and rewrite equation A.6 under the form

$$V(x) = \eta_u \max_{p\in x} u(p) + (1 - \eta_u) \min_{q\in x} u(q)$$

Axiom 4.1 yields $\eta_u > 0$. If axiom 3.1 is satisfied, a direct adaptation of theorem 3.1 proves the existence of a sophisticated joint representation under the Random Strotz interpretation. If axiom 3.1 is violated, we obtain $\eta_u < \nu_u$. It is therefore possible to rewrite

$$V(x) = \nu_u[\max_{p\in x} \frac{\eta_u}{\nu_u} u(p) - \max_{q\in x} (\frac{\eta_u}{\nu_u} - 1)u(q)] + \nu_{-u} \min_{q\in x} u(q) \qquad \text{(A.7)}$$

Equation A.7 defines a sophisticated representation of $\succeq$, since the function V attaches weights equal to (ν_u, ν_{-u}) to the states $(u, -u)$ respectively.

References

Ahn, David S, Ryota Iijima, and Todd Sarver (2015), "Comparative measures of naiveté." Working paper.

Ahn, David S and Todd Sarver (2013), "Preference for flexibility and random choice." *Econometrica*, 81, 341–361.

Azrieli, Yaron, Christopher P Chambers, and Paul J Healy (2015), "Incentives in experiments: a theoretical analysis." Working paper.

Bartling, Björn, Ernst Fehr, and Holger Herz (2014), "The intrinsic value of decision rights." *Econometrica*, 82, 2005–2039.

Baumeister, Roy F, Kathleen D Vohs, and Dianne M Tice (2007), "The strength model of self-control." *Current directions in psychological science*, 16, 351–55.

Bénabou, Roland and Jean Tirole (2004), "Willpower and personal rules." *Journal of Political Economy*, 112, 848–886.

Billingsley, Patrick (2012), *Probability and Measure*. Wiley.

Dean, Mark and John McNeill (2015), "Preference for flexibility and random choice: an experimental analysis." Working paper.

Dekel, Eddie and Barton Lipman (2012), "Costly self-control and random self-indulgence." *Econometrica*, 80, 1271–1302.

Dekel, Eddie, Barton Lipman, and Aldo Rustichini (2001), "Representing preferences with a unique subjective state space." *Econometrica*, 69, 891–934.

Eliaz, Kfir and Ran Spiegler (2006), "Contracting with diversely naive agents." *Review of Economic Studies*, 73, 689–714.

Ergin, Haluk and Todd Sarver (2010), "A unique costly contemplation representation." *Econometrica*, 78, 1285–1339.

Exley, Christine and Jeffrey K Naecker (2015), "Observability increases the demand for commitment devices." Working paper.

Gul, Faruk and Wolfgang Pesendorfer (2001), "Temptation and self-control." *Econometrica*, 69, 1403–1435.

Gul, Faruk and Wolfgang Pesendorfer (2006), "Random expected utility." *Econometrica*, 74, 121–146.

Heidhues, Paul and Botond Köszegi (2010), "Exploiting naivete about self-control in the credit market." *American Economic Review*, 100, 2279–2303.

Horton, John J, David G Rand, and Richard J Zeckhauser (2011), "The online laboratory: conducting experiments in a real labor market." *Experimental economics*, 14, 399–425.

Kopylov, Igor (2012), "Perfectionism and choice." *Econometrica*, 80, 1819–1843.

Kopylov, Igor and Jawwad Noor (2015), "Self-deception and choice." Working paper.

Kreps, David M (1979), "A preference for flexibility." *Econometrica*, 47, 565–576.

Köszegi, Botond (2014), "Behavioral contract theory." *Journal of Economic Literature*, 52, 1075–1118.

Laibson, David (1997), "Golden eggs and hyperbolic discounting." *Quarterly Journal of Economics*, 112, 443–478.

Mason, Winter and Siddharth Suri (2012), "Conducting behavioral research on amazon's mechanical turk." *Behavioral Research Methods*, 44, 1–23.

Noor, Jawwad (2007), "Commitment and self-control." *Journal of Economic Theory*, 135, 1–34.

Noor, Jawwad and Norio Takeoka (2010), "Uphill self-control." *Theoretical economics*, 5, 127–158.

Olea, Jose Luis Montiel and Tomasz Strzalecki (2014), "Axiomatization and measurement of quasi-hyperbolic discounting." *Quarterly Journal of Economics*, 129, 1449–1499.

Ortoleva, Pietro (2013), "The price of flexibility: towards a theory of thinking aversion." *Journal of Economic Theory*, 148, 903–934.

Rabin, Matthew and Ted O'Donoghue (1999), "Doing it now or later." *American Economic Review*, 89, 103–124.

Stovall, John (2010), "Multiple temptations." *Econometrica*, 78, 349–376.

Appendix B: Unidimensional Random Strotz models

This appendix proves theorem 2.1, and provides another characterization of unidimensional Random Strotz representations. Unlike the rest of the paper, this part does not assume any conditions on μ except non-triviality: continuity and finiteness are not required. The preference $\succeq$ is here defined over compact menus, as in Dekel and Lipman (2012).

1 Proof of Theorem 2.1

1.1 Necessity of axiom 2.1

Lemma B.4. *Suppose that $\succeq$ has a unidimensional Random Strotz representation (u, μ) with direction $v \in \mathcal{V}$ such that $\mu(\mathcal{C}_v) > 0$. Consider two lotteries q_1 and q_2 verifying $\{q_1\} \sim \{q_2\}$, and suppose that there exists a lottery z such that $\{z\} \succ \{q_1\} \sim \{q_2\}$. The two following statements are equivalent: (i) There exists a menu x such that $\{p\} \succ \{q_1\} \sim \{q_2\}$ for any $p \in x$ and $x \cup \{q_1\} \succ x \cup \{q_2\}$; (ii) $v(q_1) < v(q_2)$.*

Proof. We prove $(i) \Rightarrow (ii)$ by contrapositive. $\{q_1\} \sim \{q_2\}$ implies $u(q_1) = u(q_2)$. Suppose that $v(q_1) \geq v(q_2)$, and take any x such that $u(p) > u(q_1)$. If $w \in \overline{\mathcal{C}}_v$, we can write $w = au + \sqrt{1-a^2}v$, where $|a| \leq 1$, and obtain $w(q_1) = au(q_1) + \sqrt{1-a^2}v(q_1) \geq au(q_2) + \sqrt{1-a^2}v(q_2) = w(q_2)$. Since the support of μ is included in $\overline{\mathcal{C}}_v$, q_1 dominates q_2 on all the possible subjective states. For $i = 1, 2$, we write $\Omega^{x\cup\{q_i\}}(q_i) = \{w \in \overline{\mathcal{C}_v} | w(q_i) > \max_{p\in x} w(p)\}$ the list of subjective states on which q_i is chosen. The observation above yields $\Omega^{x\cup\{q_2\}}(q_2) \subseteq \Omega^{x\cup\{q_1\}}(q_1)$, and hence

$$
\begin{aligned}
V(x \cup \{q_1\}) &= \mu(\Omega^{x\cup\{q_1\}}(q_1))u(q_1) + \int_{w \notin \Omega^{x\cup\{q_1\}}(q_1)} \max_{p \in \mathcal{M}_w(x)} u(p)\mu(dw) \\
&\leq \mu(\Omega^{x\cup\{q_2\}}(q_2))u(q_2) + \int_{w \notin \Omega^{x\cup\{q_2\}}(q_2)} \max_{p \in \mathcal{M}_w(x)} u(p)\mu(dw) \\
&\leq V(x \cup \{q_2\})
\end{aligned}
$$

This proves that $x \cup \{q_2\} \succeq x \cup \{q_1\}$.

$(ii) \Rightarrow (i)$. Suppose that $u(q_1) = u(q_2)$ and $v(q_1) < v(q_2)$. Consider an increasing sequence $0 < a_n < 1$ of limit 1, and the increasing sequence of sets $\mathcal{C}_v(-a_n, a_n)$. This sequence has limit $\mathcal{C}_v$ which has positive measure, therefore by theorem 10.2 of Billingsley (2012) we have $\mu(\mathcal{C}_v(-a_n, a_n)) > 0$ for n large enough. Define $a = a_n$.

Consider now a number γ such that: $0 < \gamma < \frac{\sqrt{1-a^2}}{2a}[v(q_2) - v(q_1)]$. Such a number exists since $a \in (0,1)$ and $v(q_2) > v(q_1)$. Suppose first that we can find γ small enough to ensure that $p = \frac{1}{2}q_1 + \frac{1}{2}q_2 + \gamma u$ is a legitimate lottery. Notice that $u(p) > u(q_1), u(q_2)$ and that

$$\begin{aligned} au(p) + \sqrt{1-a^2}v(p) &= a\gamma + au(q_2) + \sqrt{1-a^2}[\frac{1}{2}v(q_1) + \frac{1}{2}v(q_2)] \\ &< au(q_2) + \sqrt{1-a^2}v(q_2) \end{aligned}$$

Therefore q_2 dominates p over the set $\overline{\mathcal{C}}_v(-1, a)$. Similarly,

$$\begin{aligned} -au(p) + \sqrt{1-a^2}v(p) &= -a\gamma - au(q_1) + \sqrt{1-a^2}[\frac{1}{2}v(q_1) + \frac{1}{2}v(q_2)] \\ &> -au(q_1) + \sqrt{1-a^2}v(q_1) \end{aligned}$$

And hence p dominates q_1 over the set $\overline{\mathcal{C}}_v(-a, 1)$.

Thus we obtain $\Omega^{\{p,q_1\}}(q_1) \subset \Omega^{\{p,q_2\}}(q_2)$, and $\mathcal{C}_v(-a, a) \subseteq \Omega^{\{p,q_2\}}(q_2) \backslash \Omega^{\{p,q_1\}}(q_1)$. Since $\mu(\mathcal{C}_v(-a, a)) > 0$, this yields $\mu(\Omega^{\{p,q_1\}}(q_1)) < \mu(\Omega^{\{p,q_2\}}(q_2))$.

Hence

$$\begin{aligned} V(\{p, q_1\}) &= \mu(\Omega^{\{p,q_1\}}(q_1))u(q_1) + (1 - \mu(\Omega^{\{p,q_1\}}(q_1)))u(p) \\ &> \mu(\Omega^{\{p,q_2\}}(q_2))u(q_2) + (1 - \mu(\Omega^{\{p,q_2\}}(q_2)))u(p) = V(\{p, q_2\}) \end{aligned}$$

Therefore the triple $(x = \{p\}, q_1, q_2)$ satisfies $x \cup \{q_1\} \succ x \cup \{q_2\}$.

If $\frac{1}{2}q_1 + \frac{1}{2}q_2 + \gamma u$ is not a lottery for any $\gamma > 0$, since $\frac{1}{2}q_1 + \frac{1}{2}q_2$ is not a maximizer of u among $\Delta(\mathcal{Z})$, a standard separation argument shows that it is possible to find $\tilde{u} \in \mathcal{W}$ that satisfies $u.\tilde{u} > 0$ and such that $\frac{1}{2}q_1 + \frac{1}{2}q_2 + \gamma\tilde{u}$ is a lottery. The same construction holds with a large enough to satisfy $au.\tilde{u} > \sqrt{1-a^2}$ and γ such that $0 < \gamma[au.\tilde{u} \pm \sqrt{1-a^2}] < \dfrac{\sqrt{1-a^2}}{2}[v(q_2) - v(q_1)]$.

□

To prove the necessity of axiom 2.1, consider a unidimensional Random Strotz (u, μ) with direction v. If $\mu(\mathcal{C}_v) = 0$, then $\mu(\{u, -u\}) = 1$, thus for any (x, q_1, q_2) such that $\{p\} \succ \{q_1\} \sim \{q_2\}$ for all $p \in x$, we have $x \cup \{q_1\} \sim x \cup \{q_2\}$, and the condition of axiom 2.1 is trivially satisfied.

Suppose now that $\mu(\mathcal{C}_v) > 0$ and take (x, y, q_1, q_2) such that $\{p\} \succ \{q_1\} \sim \{q_2\}$ for all $p \in x$ and $x \cup \{q_1\} \succ x \cup \{q_2\}$. Since q_1 and q_2 do not maximize u on $\Delta(\mathcal{Z})$, the implication $(i) \Rightarrow (ii)$ of lemma B.4 yields $v(q_1) < v(q_2)$, which in turn implies $y \cup \{q_1\} \succeq y \cup \{q_2\}$. This completes the proof of necessity.

1.2 Sufficiency of axiom 2.1

Suppose that the representation (u, μ) of $\succeq$ is not unidimensional. It is clear that $\mu(\mathcal{C}(-1, 1)) > 0$, otherwise we would have $\mu(\{u, -u\}) = 1$ and therefore $\mu(\overline{\mathcal{C}}_v) = 1$ for any $v \in \mathcal{V}$.

Claim B.1. There exists $\xi < 1$ such that $\mu(\mathcal{C}_v) < \mu(\mathcal{C}(-\xi, \xi))$ for all $v \in \mathcal{V}$.

Proof. If this is not the case, there exists an increasing sequence $0 < a_n < 1$ of limit 1, and a sequence $v_n \in \mathcal{V}$ such that

$$\mu(\mathcal{C}_{v_n}) \geq \mu(\mathcal{C}(-a_n, a_n)) \tag{B.1}$$

Take m sufficiently large to guarantee $\mu(\mathcal{C}(-a_m, a_m)) > \dfrac{1}{2}\mu(\mathcal{C}(-1, 1))$. This is possible since $\mathcal{C}(-a_m, a_m)$ is an increasing sequence of limit $\mathcal{C}(-1, 1)$, which has positive measure. Suppose that v_n is not constant for $n \geq m$. We can find $v_1 \neq v_2$ such that $\mu(\mathcal{C}_{v_1}) > \dfrac{1}{2}\mu(\mathcal{C}(-1, 1)$ and $\mu(\mathcal{C}_{v_2}) > \dfrac{1}{2}\mu(\mathcal{C}(-1, 1)$, which implies $\mu(\mathcal{C}_{v_1} \cup \mathcal{C}_{v_2}) > \mu(\mathcal{C}(-1, 1))$. This is a contradiction, since $\mathcal{C}_{v_1} \cup \mathcal{C}_{v_2} \subseteq \mathcal{C}(-1, 1)$. Hence, v_n is constant for n large enough. Denote v^* its limit, and take the limit in B.1. We obtain $\mu(\mathcal{C}_v) \geq \mu(\mathcal{C}(-1, 1))$, which further implies

$$\begin{aligned}\mu(\overline{\mathcal{C}}_v) &= \mu(\{u, -u\}) + \mu(\mathcal{C}_v) \\ &\geq \mu(\{u, -u\}) + \mu(\mathcal{C}(-1, 1)) = 1\end{aligned}$$

And hence, $\mu(\overline{\mathcal{C}}_v) = 1$. This is a contradiction, since (u, μ) is not unidimensional. □

For each $v \in \mathcal{V}$, we define $\mathcal{B}(v, \epsilon) = \{w \in \mathcal{V} | \|w - v\| < \epsilon\}$ the open ball of radius ϵ and center v, restricted to utilities which are orthogonal to u. We also define

$$\mathcal{A}(v, \epsilon) = \bigcup_{w \in \mathcal{B}(v, \epsilon)} \mathcal{C}_w = \{au + \sqrt{1 - a^2}w | -1 < a < 1, w \in \mathcal{V}, \|w - v\| < \epsilon\}$$

$\mathcal{A}(v, \epsilon)$ contains the utilities whose direction lies in the open ball of center v and radius ϵ.

Claim B.2. For any $\epsilon > 0$ low enough, there exists two expected utilities v_1^ϵ and v_2^ϵ such that $\mu(\mathcal{A}(v_1^\epsilon, \epsilon)) > 0$, $\mu(\mathcal{A}(v_2^\epsilon, \epsilon)) > 0$ and $\|v_1^\epsilon - v_2^\epsilon\| > 7\epsilon$.

Proof. Consider ξ defined by claim B.1. ξ can be chosen high enough to guarantee that $\mu(\overline{\mathcal{C}}(-\xi, \xi)) > 0$. The set $\overline{\mathcal{C}}(-\xi, \xi) = \{au + \sqrt{1 - a^2}w, w \in \mathcal{V}, |a| \leq \xi\}$ is compact since it is closed and bounded. Moreover, it is covered by the union of the open sets $\mathcal{A}(v, \epsilon)$ for all $v \in \mathcal{V}$. By the Borel-Lebesgue theorem, there exists a finite family $\mathcal{F} \subset \mathcal{V}$ such that $\{\mathcal{A}(v, \epsilon)\}_{v\in\mathcal{F}}$ covers $\overline{\mathcal{C}}(-\xi, \xi)$. An immediate implication is that $\sum_{v\in\mathcal{F}} \mu(\mathcal{A}(v, \epsilon)) > 0$.

Let us show that, if ϵ is low enough, this subcovering contains at least two sets of positive measure that can be separated as stated in the claim. We proceed by contradiction. Suppose that for any $\eta > 0$, there exists $\epsilon < \eta$, a finite subset $\mathcal{F}_\epsilon$ such that $\{\mathcal{A}(v, \epsilon)\}_{v\in\mathcal{F}}$ covers $\overline{\mathcal{C}}(-\xi, \xi)$, and $v_\epsilon \in \mathcal{F}_\epsilon$ such that for any $v \in \mathcal{F}_\epsilon$, $\mu(\mathcal{A}(v, \epsilon)) > 0 \Rightarrow \|v - v_\epsilon\| < 7\epsilon$. Notice that $\bigcup_{v\in\mathcal{B}(v_\epsilon,7\epsilon)} \mathcal{A}(v, \epsilon) \subseteq \mathcal{A}(v_\epsilon, 8\epsilon)$. Hence, since $\sum_{v\in\mathcal{F}_\epsilon\cap\mathcal{B}(v_\epsilon,7\epsilon)} \mu(\mathcal{A}(v, \epsilon)) \geq \mu(\mathcal{C}(-\xi, \xi))$, we obtain $\mu(\mathcal{A}(v_\epsilon, 8\epsilon)) \geq \mu(\mathcal{C}(-\xi, \xi))$.

Consider a decreasing sequence $\epsilon_n \to 0$. By the Bolzano-Weierstrass theorem, the sequence $v_n = v_{\epsilon_n}$ defined over the compact $\mathcal{V}$ admits a convergent subsequence. To simplify the notation, let us assume that v_n is itself convergent to a value v^*. For any n, we have $\mu(\mathcal{A}(v_n, 8\epsilon_n)) \geq \mu(\mathcal{C}(-\xi, \xi))$. Our next step is to show that we can take the limit in this inequality and obtain $\mu(\mathcal{C}_{v^*}) \geq \mu(\mathcal{C}(-\xi, \xi))$.

Notice that

$$\mu(\mathcal{A}(v^*, 8\epsilon_n)) - \mu(\mathcal{A}(v_n, 8\epsilon_n)) = \mu(\bigcup_{w\in\mathcal{B}(v^*,8\epsilon_n)\setminus\mathcal{B}(v_n,8\epsilon_n)} \mathcal{C}_w) - \mu(\bigcup_{w\in\mathcal{B}(v_n,8\epsilon_n)\setminus\mathcal{B}(v^*,8\epsilon_n)} \mathcal{C}_w) \tag{B.2}$$

Consider the sets

$$\mathcal{G}_n = \bigcup_{m=n}^{+\infty} \bigcup_{w\in\mathcal{B}(v^*,8\epsilon_m)\setminus\mathcal{B}(v_m,8\epsilon_m)} \mathcal{C}_w$$

$\mathcal{G}_m$ is decreasing and has for limit $\lim_{n\to+\infty} \mathcal{G}_n = \emptyset$. Thus, $\lim_{n\to+\infty} \mu(\mathcal{G}_n) = 0$, and since $\bigcup_{w\in\mathcal{B}(v^*,8\epsilon_n)\setminus\mathcal{B}(v_n,8\epsilon_n)} \mathcal{C}_w \subseteq \mathcal{G}_n$, we obtain

$$\lim_{n\to+\infty} \mu(\bigcup_{w\in\mathcal{B}(v^*,8\epsilon_n)\setminus\mathcal{B}(v_n,8\epsilon_n)} \mathcal{C}_w) = 0$$

A similar argument proves that

$$\lim_{n\to+\infty} \mu(\bigcup_{w\in\mathcal{B}(v_n,8\epsilon_n)\setminus\mathcal{B}(v^*,8\epsilon_n)} \mathcal{C}_w) = 0$$

Moreover, the sequence of sets $\{\mathcal{A}(v^*, 8\epsilon_n)\}$ is decreasing and converges to $\mathcal{C}_{v^*}$ when $n \to +\infty$. Thus, taking the limit in equation B.2 shows that $\mu(\mathcal{A}(v_n, 8\epsilon_n)) \to \mu(\mathcal{C}_{v^*})$, which implies $\mu(\mathcal{C}_{v^*}) \geq \mu(\mathcal{C}(-\xi, \xi))$. This latter inequality contradicts the statement in claim B.1. This completes the proof of claim B.2. $\square$

If $v \in \mathcal{V}$, we write $\mathcal{A}^{\epsilon}_{v}(-1, a^*) = \{au + \sqrt{1-a^2}w | -1 < a < a^*, w \in \mathcal{V} \cap \mathcal{B}(v, \epsilon)\}$: it contains the utilities whose direction lies in the open ball of center v and radius ϵ, and whose intensity lies strictly between -1 and by a^*.

Claim B.3. For any $\epsilon > 0$ small enough, there exists $v_1, v_2 \in \mathcal{V}$ and $a^* < 1$ such that
$\|v_1 - v_2\| > 7\epsilon$ and

$$\min(\mu(\mathcal{A}^{\epsilon}_{v_1}(-1, a^*)), \mu(\mathcal{A}^{\epsilon}_{v_2}(-1, a^*))) > \mu(\mathcal{C}(-1, -a^*))$$

Proof. Take v_1 and v_2 given by claim B.2. The proof stems immediately from the fact that $\mu(\mathcal{A}^{\epsilon}_{v_1}(-1, a^*))$ and $\mu(\mathcal{A}^{\epsilon}_{v_2}(-1, a^*))$ converge to positive values when $a^* \to 1$, while $\mu(\mathcal{C}(-1, -a^*))$ tends to zero. $\square$

Claim B.4. For any ϵ small enough, there exists three numbers α, β, γ such that $\alpha > 0$,

$$\begin{cases} \alpha > \beta \\ \alpha \sup_{w \in \mathcal{B}(v_2, 3\epsilon)} w.v_1 < \beta \inf_{w \in \mathcal{B}(v_2, 3\epsilon)} w.v_2 \\ \alpha \sup_{w \in \mathcal{B}(v_1, 3\epsilon)} w.v_2 < \beta \inf_{w \in \mathcal{B}(v_1, 3\epsilon)} w.v_1 \end{cases}$$

and

$$\begin{cases} \alpha < \gamma \\ \alpha \inf_{w \in \mathcal{B}(v_1, \epsilon)} w.v_1 > \gamma \sup_{w \in \mathcal{B}(v_1, \epsilon)} w.v_2 \\ \alpha \inf_{w \in \mathcal{B}(v_2, \epsilon)} w.v_2 > \gamma \sup_{w \in \mathcal{B}(v_2, \epsilon)} w.v_1 \end{cases}$$

Proof. Suppose that $w \in \mathcal{B}(v_1, \epsilon)$. We have

$$\begin{aligned} w.v_1 &= \frac{1}{2}(\|w\|^2 + \|v_1\|^2 - \|w - v_1\|^2) \\ &= 1 - \frac{1}{2}\|w - v_1\|^2 \\ &> 1 - \frac{1}{2}\epsilon^2 \end{aligned}$$

And since $\|w - v_2\| \geq \|v_1 - v_2\| - \|w - v_1\| \geq 6\epsilon$, we also obtain $w.v_2 < 1 - \frac{1}{2}(6\epsilon)^2$. Similarly, $\inf_{w \in \mathcal{B}(v_2, \epsilon)} w.v_2 > 1 - \frac{1}{2}\epsilon^2$, and $\sup_{w \in \mathcal{B}(v_2, \epsilon)} w.v_1 < 1 - \frac{1}{2}(6\epsilon)^2$. Hence, if

ϵ is small enough for the denominators to be positive, the ratios $\dfrac{\sup_{w\in\mathcal{B}(v_2,\epsilon)} w.v_1}{\inf_{w\in\mathcal{B}(v_2,\epsilon)} v.w_2}$ and $\dfrac{\sup_{w\in\mathcal{B}(v_1,\epsilon)} w.v_2}{\inf_{w\in\mathcal{B}(v_1,\epsilon)} w.v_1}$ are bounded above away from 1. It is thus easy to find α and γ that satisfy the conditions. A similar reasoning can be applied to find β. $\square$

To complete the proof, take $\phi > 0$, ϵ low enough to ensure that $w.v_1 > 0$ as soon as $w_1 \in \mathcal{B}(v_1, 3\epsilon)$, and $w.v_2 > 0$ as soon as $w_2 \in \mathcal{B}(v_2, 3\epsilon)$, and take $a^*, v_1, v_2, \alpha, \beta, \gamma$ defined by claims B.3 and B.4. Define $p_1 = \mathbb{1} + \epsilon(\phi u + \gamma v_1)$, $p_2 = \mathbb{1} + \epsilon(\phi u + \beta v_2)$, $r_1 = \mathbb{1} + \epsilon(\phi u + \beta v_1)$ and $r_2 = \mathbb{1} + \epsilon(\phi u + \gamma v_2)$, and for any $w \in \mathcal{V}$ such that $w \notin \mathcal{B}(v_1, 3\epsilon) \cup \mathcal{B}(v_2, 3\epsilon)$, $z_w = \mathbb{1} + \epsilon(\phi u + \alpha w)$. u is indifferent between all these elements, with valuation $\epsilon\phi$. Consider the menu x containing p_1, p_2 and all the z_w, and the menu y containing r_1, r_2 and all the z_w.

Define now $q_1 = \mathbb{1} + \epsilon\alpha v_1$ and $q_2 = \mathbb{1} + \epsilon\alpha v_2$. We have $u(q_1) = u(q_2) = 0 < \epsilon\phi$, hence q_1 and q_2 are normatively inferior to all elements of x and y. Our next step is to show that q_1 is more tempting than q_2 with respect to the menu y, while q_2 is more tempting than q_1 with respect to the menu y.

Consider the *ex post* choice in $x \cup \{q_1\}$. It is clear that q_1 is chosen by $-u$ and not chosen by u. Consider $w \notin \{u, -u\}$, and write $w = au + \sqrt{1-a^2}\overline{w}$, where $-a^* \leq a \leq 1$. We have

$$w(q_1) = \epsilon\sqrt{1-a^2}\alpha\overline{w}.v_1 \tag{B.3}$$

Suppose first, that $\overline{w} \in \mathcal{B}(v_1, 3\epsilon)$. We have

$$w(p_1) = a\epsilon\phi + \epsilon\sqrt{1-a^2}\gamma\overline{w}.v_1 \tag{B.4}$$

Compare B.3 with B.4. Since $\gamma > \alpha$ and $a^* < 1$ one can pick ϕ low enough to impose the inequality $w(p_1) > w(q_1)$ for all values of $a \geq a^*$. Given this choice, p_1 dominates q_1 if $\overline{w} \in \mathcal{B}(v_1, 3\epsilon)$ and $a \geq -a^*$.

Suppose now that $\overline{w} \in \mathcal{B}(v_2, 3\epsilon)$. We have

$$w(p_2) = a\epsilon\phi + \epsilon\sqrt{1-a^2}\beta\overline{w}.v_2 \tag{B.5}$$

Compare B.3 with B.5, and notice that $\beta \inf \overline{w}.v_2 > \alpha \sup \overline{w}.v_1$ by claim B.4. Hence, again, if ϕ is low enough, p_2 dominates q_1 if $\overline{w} \in \mathcal{B}(v_2, 3\epsilon)$ and $a \geq -a^*$.

Finally, suppose that $\overline{w} \notin \mathcal{B}(v_1, 3\epsilon) \cup \mathcal{B}(v_2, 3\epsilon)$. Notice that

$$w(z_{\overline{w}}) = a\epsilon\phi + \epsilon\sqrt{1-a^2}\alpha \tag{B.6}$$

Compare B.3 and B.6. $v_1.\overline{w}$ is uniformly bounded away from 1 since $\|\overline{w} - v_1\| \geq 3\epsilon$.

Hence, by the same argument as above, if ϕ is small enough, q_1 is dominated by $z_{\overline{w}}$ if $\overline{w} \notin \mathcal{B}(v_1, 3\epsilon) \cup \mathcal{B}(v_2, 3\epsilon)$ and $a \geq -a^*$.

To sum up, if ϕ is low enough, q_1 is never chosen in $x \cup \{q_1\}$ by a state whose intensity of temptation is stronger than $-a^*$. Therefore, denoting $\alpha^{x\cup\{q_1\}} = \mu(\{w \in \mathcal{W} | w(q_1) > \max_{p\in x} w(p)\})$ the *ex ante* probability of choosing q_1 we have

$$\alpha^{x\cup\{q_1\}}(q_1) \leq \mu(\{-u\}) + \mu(\mathcal{C}(-1, a^*)) \tag{B.7}$$

Consider now the anticipated choice in the set $y \cup \{q_1\}$ in a state of the form $w = au + \sqrt{1-a^2}\overline{w}$ where $a < a^*$ and $\overline{w} \in \mathcal{B}(v_1, \epsilon)$. Notice that

$$w(r_1) = a\epsilon\phi + \epsilon\sqrt{1-a^2}\beta\overline{w}.v_1 \tag{B.8}$$

Since $\alpha > \beta$, $v_1.\overline{w} > 0$ and $a < a^* < 1$, by equations B.3and B.8 it is possible to take ϕ low enough to ensure that $w(r_1) < w(q_1)$. Observe now that

$$w(r_2) = a\epsilon\phi + \epsilon\sqrt{1-a^2}\gamma\overline{w}.v_2 \tag{B.9}$$

But by claim B.4, we have $\alpha \inf \overline{w}.v_1 > \gamma \sup \overline{w}.v_2$. Hence we can choose ϕ low enough to ensure that $w(r_2) < w(q_1)$.

Finally, for any $\hat{w} \notin \mathcal{B}(v_1, 3\epsilon) \cup \mathcal{B}(v_2, 3\epsilon)$ it is easy to show that $\hat{w}.\overline{w} < \overline{w}.v_1$. Thus, since $w(z_{\hat{w}}) = a\epsilon\phi + \epsilon\alpha\sqrt{1-a^2}\hat{w}.\overline{w}$, we obtain $w(z_{\hat{w}}) < w(q_1)$ for any $\hat{w}$ as soon as $a < a^*$, if ϕ is chosen small enough.

To sum up, q_1 is the single maximizer $y \cup \{q_1\}$ at least on the utilities of the form $au + \sqrt{1-a^2}\overline{w}$, where $\overline{w} \in \mathcal{B}(v_1, \epsilon)$ and $a < a^*$, which means that the probability of choosing it verifies

$$\alpha^{y\cup\{q_1\}}(q_1) \geq \mu(\{-u\}) + \mu(\mathcal{A}^{\epsilon}_{v_1}(-1, a^*)) \tag{B.10}$$

The same arguments prove that the choice of q_2 in x or y satisfies

$$\alpha^{x\cup\{q_2\}}(q_2) \geq \mu(\{-u\}) + \mu(\mathcal{A}^{\epsilon}_{v_2}(-1, a^*)) \tag{B.11}$$

and

$$\alpha^{y\cup\{q_2\}}(q_2) \leq \mu(\{-u\}) + \mu(\mathcal{C}(-1, -a^*)) \tag{B.12}$$

Compare equations B.7 and B.11. By claim B.3, we obtain $\alpha^{x\cup\{q_1\}}(q_1) < \alpha^{x\cup\{q_2\}}(q_2)$, which implies $x \cup \{q_1\} \succ x \cup \{q_2\}$. In contrast, by equations B.10 and B.12, $\alpha^{y\cup\{q_1\}}(q_1) > \alpha^{y\cup\{q_2\}}(q_2)$, which implies $y \cup \{q_1\} \prec y \cup \{q_2\}$. These two properties together violate axiom 2.1.

2 An alternative characterization

2.1 Representation theorem

In this section we show that Unidimensional Random Strotz representations are characterized by another intuitive behavioral property.

Axiom B.1 (Unique Temptation).
If $\{p\} \succ \{q\}$ for any $p \in x \cup y$, then $x \succ x \cup \{q\}, y \succ y \cup \{q\} \Rightarrow x \cup y \succ x \cup y \cup \{q\}$.

Axiom B.1 also characterizes unidimensional models, with one caveat: if the support of μ cannot be bounded away from $-u$, the consequent $x \cup y \succ x \cup y \cup \{q\}$ is always true because all inferior options are tempting. Axiom B.1 has no content in that case. To overcome this issue, we impose another property that guarantees the existence of a higher bound on the intensity of the temptation.

Axiom B.2 (Limited Temptation).
$\exists x \in \mathcal{A}, q \in \Delta(Z)$ such that $\{p\} \succ \{q\}$ for any $p \in x \cup y$ and $x \sim x \cup \{q\}$.

Axiom B.2 is a richness condition: it simply states that some options in the choice set are not tempting. This condition is innocuous, since it is satisfied if an option that appears extremely undesirable both *ex ante* and *ex post* is added to the set of prizes. We will first show that axiom B.2 is equivalent to the existence of a a neighborhood of $\{-u\}$ of measure zero; and then proceed to show that, among the Random Strotz models that satisfy axiom B.2, unidimensional models are characterized by axiom B.1.

Lemma B.5. *Suppose that $\succeq$ has a Random Strotz representation (u, μ). The following statements are equivalent: (i) $\succeq$ satisfies axiom B.2; (ii) there exists $a > -1$ such that $\mu(\overline{\mathcal{C}}(-1, a)) = 0$.*

Proof. $(ii) \Rightarrow (i)$. Suppose that μ satisfies (ii) for some $a > -1$. Consider a pair (ϵ, γ) such that $\epsilon > 0$ and $a\epsilon + \sqrt{1-a^2}\gamma > 0$. Define $q = \mathbb{1}$ and $p_v = \mathbb{1} + \phi(\epsilon u + \gamma v)$ for $v \in \mathcal{V}$, where $\phi > 0$ is taken sufficiently small for p_v to be an interior lottery for all v. We have $u(p_v) = \phi\epsilon > u(q) = 0$ for all v, and

$$au(p_v) + \sqrt{1-a^2}v(p_v) = \phi(a\epsilon + \sqrt{1-a^2}\gamma) > 0 = au(q) + \sqrt{1-a^2}v(q)$$

This shows that $\alpha^{x \cup \{q\}}(q) \leq \mu(\overline{\mathcal{C}}(-1, a)) = 0$, and and hence $x \sim x \cup \{q\}$.

$(i) \Rightarrow (ii)$. Suppose that (ii) does not hold, i.e. that for any $a > -1$, we have $\mu(\overline{\mathcal{C}}(-1,a)) > 0$.

Consider now x and q such that $u(p) > u(q)$ for any $p \in x$. Consider the function $a : \mathcal{V} \times x \Rightarrow (-1,1)$ defined by the equation

$$\frac{a(v,p)}{\sqrt{1-a(v,p)^2}} = \frac{v(q)-v(p)}{u(p)-u(q)} \tag{B.13}$$

equation B.13 uniquely defines a value $a(v,p)$ such that $a(v,p) > -1$. Moreover, a is continuous in the product topology. By Tychonoff's theorem, $\mathcal{V} \times x$ is compact since $\mathcal{V}$ and x are compact. Thus, $\inf_{(v,p)\in\mathcal{V}\times x} a(v,p) > -1$. Take any a such that $-1 < a < \inf_{(v,p)\in\mathcal{V}\times x} a(v,p)$, any $(v,p) \in \mathcal{X} \times a$. By equation B.13 we have $au(p) + \sqrt{1-a^2}v(p) < au(q) + \sqrt{1-a^2}v(q)$. Hence, $\Omega^{x\cup\{q\}} \supseteq \overline{\mathcal{C}}(-1,a)$, which yields $\mu(\Omega^{x\cup\{q\}}) > 0$, and finally $x \cup \{q\} \prec x$. Since this result is obtained for any pair (x,q), the preference $\succeq$ does not satisfy axiom B.2. □

Theorem B.1. *Suppose that $\succeq$ has a Random Strotz representation (u,μ), and that $\succeq$ satisfies axiom B.2. (u,μ) is Unidimensional if and only if $\succ$ satisfies axiom B.1.*

2.2 Necessity of axiom B.1

Suppose that $\succ$ has a Unidimensional Random Strotz representation (u,μ) of direction v. Take a triple (x,y,q) such that $\{p\} \succ \{q\}$ for any $p \in x\cup y$, $x \succ x\cup\{q\}$ and $y \succ y \cup \{q\}$. The same arguments used to find a in the proof of lemma B.5 can be used to obtain

$$a_x = \sup\{a \in [-1,1] | au(q) + \sqrt{1-a^2}v(q) \geq \sup_{p\in x} au(p) + \sqrt{1-a^2}v(p)\}$$

and a_y similarly by substituting y for x. $x \succ x \cup \{q\}$ and $y \succ y \cup \{q\}$ imply $\mu(\{-u\}) + \mu(\mathcal{C}_v(-1,a_x)) > 0$ and $\mu(\{-u\}) + \mu(\{-u\}) + \mu(\mathcal{C}_v(-1,a_y)) > 0$. Define $a = \min(a_x,a_y)$. It is easy to see that q is chosen in the menu $x\cup y\cup\{q\}$ by $-u$ and by all the utilities of the form $\tilde{a}u + \sqrt{1-\tilde{a}^2}v$ where $-1 < \tilde{a} < a$, and that this set has positive measure. As a consequence, $x\cup y \succ x\cup y\cup\{q\}$.

2.3 Sufficiency of axiom B.1

We prove the sufficiency of the axiom by contrapositive. Suppose that the Random Strotz representation (u,μ) of $\succ$ is not unidimensional, and that $\succ$ satisfies axiom B.2.

Claim B.5. For any $\epsilon > 0$ low enough, there exists $v_1, v_2 \in \mathcal{V}$ and $a^* \in (0,1)$ such that: (i) $\mu(\mathcal{A}^{\epsilon}_{v_1}(-1, a^*)) > 0$, (ii) $\mu(\mathcal{A}^{\epsilon}_{v_2}(-1, a^*)) > 0$, (iii) $\|v_1 - v_2\| > 7\epsilon$, and (iv) $\mu(\overline{\mathcal{C}}(-1, -a^*))) = 0$.

Proof. The first three parts come from claims B.2 and B.3 in the proof of theorem 2.1. Part (iv) comes from lemma B.5. □

Take $\alpha, \phi > 0$ and define now $q = \mathbb{1}$, and: (i) for any $\hat{w} \notin \mathcal{B}(v_1, 3\epsilon)$, $p_{\hat{w}} = \mathbb{1} + \phi u + \alpha(\hat{w} - v_1)$; (ii) for any $\hat{w} \notin \mathcal{B}(v_2, 3\epsilon)$, $r_{\hat{w}} = \mathbb{1} + \phi u + \alpha(\hat{w} - v_2)$. α and ϕ can be taken small enough to make sure that these elements are well-defined lotteries. Define also $x = \{p_{\hat{w}}\}_{\hat{w}\notin\mathcal{B}(v_1,3\epsilon)}$ and $y = \{r_{\hat{w}}\}_{\hat{w}\notin\mathcal{B}(v_2,3\epsilon)}$. We observe that u equals $\phi > u(q) = 0$ on any element of $x \cup y$.

Consider the choice made in $x\cup\{q\}$. Take $w \in \mathcal{W}$, written $w = au+\sqrt{1-a^2}\overline{w}$. Suppose that $\overline{w} \in \mathcal{B}(v_1, \epsilon)$. We have $w(q) = 0$, and for any $\hat{w} \notin \mathcal{B}(v_1, 3\epsilon)$,

$$w(p_{\hat{w}}) = a\phi + \sqrt{1-a^2}\alpha(\hat{w}.\overline{w} - v_1.\overline{w})$$

In addition, we have

$$\begin{aligned}\hat{w}.\overline{w} &= \frac{1}{2}(\|\hat{w}\|^2 + \|\overline{w}\|^2 - \|\overline{w} - \hat{w}\|^2) \\ &= 1 - \frac{1}{2}\|\overline{w} - \hat{w}\|^2 \\ &\leq 1 - 2\epsilon^2\end{aligned}$$

since $\|\overline{w} - \hat{w}\| \geq \|\hat{w} - v_1\| - \|\overline{w} - v_1\| \geq 2\epsilon$.

A similar argument shows that $v_1.\overline{w} \geq 1 - \frac{\epsilon^2}{2}$, which implies $\hat{w}.\overline{w} - v_1.\overline{w} < -\frac{3\epsilon^2}{2} < 0$. Therefore we can choose ϕ small enough such that the inequality $w(p_{\hat{w}}) < 0$ is satisfied provided that $a < a^*$. We obtain $\Omega^{x\cup\{q\}}(q) \supseteq \mathcal{A}^{\epsilon}_{v_1}(-1, a^*)$, which implies $\alpha^{x\cup\{q\}}(q) > 0$, and hence $x \cup \{q\} \prec x$. Similarly, we choose ϕ small enough to obtain $y \cup \{q\} \prec y$.

Suppose now that $\overline{w} \notin \mathcal{B}(v_1, 3\epsilon)$. We have

$$w(p_{\overline{w}}) = a\phi + \sqrt{1-a^2}\alpha(1 - v_1.\overline{w})$$

And since $\|v_1 - \overline{w}\| \geq 3\epsilon$, $v_1.\overline{w} < 1 - \frac{9\epsilon^2}{2}$. Therefore we can choose ϕ low enough to ensure that $w(p_{\overline{w}}) > 0$ is satisfied as soon as $a \geq -a^*$. Similarly, if ϕ is small enough and $\overline{w} \notin \mathcal{B}(v_2, 3\epsilon)$, the inequality $w(r_{\overline{w}}) > 0$ is satisfied if $a \geq -a^*$. Since

$\|v_1 - v_2\| > 7\epsilon$, $(\mathcal{W} \setminus \mathcal{B}(v_1, 3\epsilon)) \cup (\mathcal{W} \setminus \mathcal{B}(v_2, 3\epsilon)) = \mathcal{W}$. This proves that, if $a \geq -a^*$, in every direction $\overline{w}$, q is dominated by an element of $x \cup y$. Therefore $\alpha^{x\cup y\cup\{q\}}(q) \leq \mu(\overline{\mathcal{C}}(-1, -a^*)) = 0$, and thus $x \cup y \cup \{q\} \sim x \cup y$. The triple (x, y, q) violates axiom B.1.