# COGNITIVE DISSONANCE AND CHOICE*

Larry G. Epstein  Igor Kopylov

November 10, 2005

**Abstract**

People like to feel good about past decisions. This paper models self-justification of past decisions. The model is axiomatic: axioms are defined on preference over ex ante actions (modeled formally by menus). The representation of preference admits the interpretation that the agent adjusts beliefs after taking an action so as to be more optimistic about its possible consequences. In particular, the ex post choice of beliefs is part of the representation of preference and not a primitive assumption. Behavioral characterizations are given to the comparisons "1 exhibits more dissonance than 2" and "1 is more self-justifying than 2."

## 1. INTRODUCTION

### 1.1. Objective

There is considerable evidence in psychology that people like to view themselves as being smart, and in particular, as having made correct decisions in the past. Thus they may change beliefs *after* taking an action and become more optimistic about its possible consequences, in order to feel better about having chosen it. Such behavior is a special case of an affinity for cognitive consistency - for example, an affinity for consistency among beliefs or opinions (Festinger [12]). Here the two cognitions are "I have taken an action that could lead to unfavorable outcomes"

---

*Epstein is at Department of Economics, University of Rochester, Rochester, NY 14627, lepn@troi.cc.rochester.edu; Kopylov is at Department of Economics, UC Irvine, Irvine, CA 92697, ikopylov@uci.edu. We would like to acknowledge helpful conversations with Emanuela Cardia, Faruk Gul, Jawwad Noor, Wolfgang Pesendorfer and especially Massimo Marinacci.

and "I am a smart person who would not make poor choices"; adopting a more optimistic belief about future outcomes serves to reduce this dissonance. Though the term cognitive dissonance is often used more broadly, we use it here to refer to ex post self-justification of past actions. Our objective is to model an agent who exhibits such cognitive dissonance.

Models of cognitive dissonance in economics treat beliefs as choice variables, on a par with other more standard choice variables, such as consumption and savings. Thus Akerlof and Dickens [1, p. 307] propose as basic propositions of their model of cognitive dissonance that preference is defined over beliefs and that beliefs are subject to choice. While a more optimistic outlook makes one feel better about the past decision, the agent recognizes that adopting more optimistic beliefs would take her further from the "truth" and thus would lead to suboptimal choices in decisions still to be made. The optimal belief is determined by making this trade-off. Similarly, in Rabin [26], utility depends directly on beliefs. This modeling approach is nonstandard in economics and may make one uncomfortable for a number of reasons. First, it begs the question "what is the feasible set from which beliefs are chosen?" Unlike other choice variables for which the market determines feasible (budget) sets, the feasible set of beliefs is presumably subjective (in the mind of the agent) and thus invariably requires an ad hoc specification. A possibly more important concern is the observability of chosen beliefs and hence testability of the model. While in psychology it is standard to take beliefs as observable through interviews or questionnaires, many economists adhere to the choice-theoretic approach to beliefs, initiated by Savage, whereby beliefs are observable only indirectly through choices among actions.

In this paper, we propose a choice-theoretic and axiomatic model of cognitive dissonance. Preferences are defined over actions (modeled formally by menus) and axioms are imposed on these preferences. Thus empirical testability relies only on the ranking of actions being observable. The functional form for utility admits an interpretation whereby the agent behaves *as if* she chooses beliefs ex post in the manner described above, but this is a *result* - part of the representation of preferences over actions. Finally, the above question about the feasible set of beliefs is answered automatically by the representation.

We emphasize that our agent is not boundedly rational or myopic. Rather she is sophisticated and forward-looking - when choosing an action ex ante she is fully aware that she will later experience cognitive dissonance and that this will affect her later decisions. She has this sophistication in common with agents in most economic models, but one may wonder whether individuals outside those models are typically self-aware to this degree. We are not familiar with definitive evidence on this question and in its absence, we are inclined to feel that full self-awareness is a plausible working hypothesis. Even where the opposite extreme of complete naivete seems descriptively more accurate, our model may help to clarify which economic consequences are due to cognitive dissonance per se and which are due to naivete. In addition, the assumption of sophistication is vital for a choice-theoretic approach: because she anticipates her cognitive dissonance, it affects her current choice of actions. This makes it possible to infer cognitive dissonance from her (in principle observable) choice of actions, consistent with the choice-theoretic tradition of Savage. Thus sophistication seems justifiable also on the methodological grounds of permitting the exploration of modest departures from standard models.

### 1.2. Model Outline

As described above, cognitive dissonance implies changing beliefs, hence changing preferences, which poses difficulties for modeling behavior. One possible modeling route is to specify dynamically inconsistent preferences and then to tackle the questions of to what degree the agent anticipates future changes in preference and how intrapersonal conflicts are resolved. These are the issues familiar from Strotz [31]. We follow instead the route advocated by Gul and Pesendorfer [16] (henceforth GP) whereby behavior that indicates changing preferences over underlying alternatives can alternatively be viewed as coming from stable preferences over menus of these alternatives.

A brief outline is as follows: uncertainty is represented by the (finite) state space $S$. Time varies over three periods. The true state is realized and payoffs are received at the terminal time. The intermediate time is called the *ex post* stage. Physical actions chosen then are identified with Anscombe-Aumann acts, maps from $S$ into lotteries over consumption. A physical action is chosen also at the initial *ex ante* stage. Each such action is modeled by a menu of acts - the idea is that any action taken ex ante limits options ex post. The agent understands when choosing a menu that ex post she will choose an act from that menu. She also knows, when ranking menus, that her beliefs about $S$ will change ex post so as to make the previously chosen menu seem more attractive. Thus she knows that she will be tempted to behave differently at the ex post stage than she would if she were able to commit ex ante. She may be able to resist the temptation, but regardless, temptation is costly, and this affects her ranking of menus. Thus the latter reveals her expected change in beliefs, or her cognitive dissonance.

As a concrete illustration of the relevance of choice of menus and the behavioral manifestation of cognitive dissonance, consider a job choice model along the lines of Akerlof and Dickens [1]. Ex ante the worker chooses a job, either in a hazardous industry or in a safe one. If he chooses the hazardous industry, then ex post he can select between two kinds of safety equipment (high quality $h$ and low quality $\ell$). Each kind affects the likelihood of an accident but may not eliminate the risk entirely. Thus $h$ and $\ell$ each imply a random payoff, net of cost of the equipment, that depends on the exogenous state of the world. In other words, each can be viewed as an act and the job corresponds to the menu $\{h, \ell\}$. For the safe industry, there are no choices to be made ex post and the ultimate payoff is certain and given by $c$ (a constant act). Therefore, the safe industry corresponds to the singleton menu $\{c\}$ and the choice of job corresponds to the choice between the menus $\{h, \ell\}$ and $\{c\}$.

If the worker can commit to safety equipment at the same time that he chooses the job, then ex ante beliefs are such that he would prefer the high quality equipment, that is,

$$\{h\} \succ \{\ell\}. \tag{1.1}$$

In the standard model, menus are valued according to the best alternative that they contain, and thus the worker would also exhibit the indifference

$$\{h\} \sim \{h, \ell\}.$$

However, an agent who exhibits cognitive dissonance, and knows this ex ante, may exhibit the ranking

$$\{h\} \succ \{h, \ell\}.$$

The intuition is as follows: after accepting the job in the hazardous industry, the worker faces the two cognitions - "my job is dangerous" and "I am a smart person and would not choose a precarious job". He relieves this dissonance, and reduces doubts about his job choice, by changing his prior beliefs, as reflected in the ex ante ranking (1.1), and believing instead that the job is not so dangerous after all. This creates the temptation to choose $\ell$ rather than $h$. The worker anticipates this temptation. Accordingly, if he dislikes temptation, he would rank $\{h, \ell\}$ as strictly worse than $\{h\}$.

If commitment to high quality equipment can be made simultaneously with choice of the hazardous job, the worker would so commit. Because that would leave no decisions left to be made ex post, cognitive dissonance would not be behaviorally relevant. Assume that such commitment is not possible (Akerlof and Dickens give reasons why commitment may not be possible). Then there remains the question of whether given the menu $\{h, \ell\}$ ex post, he yields to the temptation and chooses $\ell$. He feels that his prior beliefs were "correct" and thus "should" be used to guide decisions - in other words, $h$ is the correct choice. The balance between what he ought to do and the tempting alternative depends on the worker's self-control. With high self-control, he may resist the temptation and choose $h$. Following GP, we suppose that this case (or rather the ex ante expectation thereof) is captured by the ranking

$$\{h, \ell\} \succ \{\ell\}.$$

The expectation of yielding and choosing $\ell$ is captured by the ranking

$$\{h, \ell\} \sim \{\ell\}. \tag{1.2}$$

Rational expectations about cognitive dissonance may lead to choice of the safe industry. However, if

$$\{h, \ell\} \succ \{c\},$$

then the worker chooses the hazardous industry and, assuming (1.2), later adopts the poor safety equipment corresponding to $\ell$. To an outsider, or from the perspective of (1.1), the worker may appear careless or overconfident.

To this point, we have suggested that cognitive dissonance could explain the ranking

$$\{h\} \succ \{h, \ell\} \succeq \{\ell\}. \tag{1.3}$$

This ranking is a special case of GP's central axiom of Set-Betweenness. While GP argue that such rankings reveal the presence of temptation and self-control problems, the reason for temptation is unspecified. Put another way, the ranking under commitment may conflict with choice behavior out of the menu available ex post, but the reason for this difference is not clear given only (1.3). For example, (1.3) could be due to underlying preferences (taste or risk aversion) changing with the passage of time, rather than beliefs changing in order to justify the previous choice of menu. But there is other behavior, described via axioms in our formal model, that would exclude such interpretations and support an interpretation in terms of cognitive dissonance.

### 1.3. Related Literature

It has been argued that a moderate degree of (optimistic) illusion can be psychologically beneficial even net of the loss in efficacy of decisions; see Taylor and Brown [33], Taylor [32] and Baumeister [4], for example.

The psychological theory of cognitive dissonance is due to Festinger [12]. Dissonance originates with an action and the subsequent evaluation of that action. Where there exists dissonance between having taken that action and subsequent beliefs, the theory posits that those beliefs will be changed to match or justify the past action. Aronson [3] is an excellent textbook treatment and overview of the supporting evidence from psychology. Some of this evidence is strongly suggestive that cognitive dissonance has economic consequences; for example, the efficacy of the "foot-in-the-door-technique", whereby a small commitment by individuals makes it easier to persuade them later to commit further in that direction, suggests the efficacy of two-stage mechanisms, possibly including an entry fee at the first stage. Several other potential applications have been developed in formal economic models as we describe below.

Akerlof and Dickens suggest that cognitive dissonance can play a role in explaining some economic phenomena that are arguably puzzling from the perspective of more standard models. These include the existence of safety regulation (based on the job-choice model sketched above), why noninformational advertising is effective (it gives external justification for an individual to believe she is making a smart decision in buying the product), and why persons often fail to purchase actuarially favorable disaster (flood or earthquake) insurance. The story here is analogous to that concerning safety equipment and fits naturally into our modeling approach: after choosing a house (or menu), it reduces dissonance to believe that a flood is so unlikely as to not justify buying insurance (the choice of a particular act), even though she would have bought insurance simultaneously with the house purchase. Similarly, cognitive dissonance can explain why researchers or investors may appear "overly optimistic" in their pursuit of a previously chosen project (a menu). It feels good to believe that the research project previously embarked on is a promising one and thus ongoing efforts may be guided by otherwise unwarranted optimism.

Rabin [26] models the choice of an enjoyable but immoral activity in light of dissonance between one's beliefs about what is moral and the chosen level of activity. Haagsma and Koning [17] show how cognitive dissonance can generate barriers to exiting an unproductive industry. Smith [30] shows how cognitive dissonance can explain why wages tend to rise faster than productivity. The worker justifies his job situation ex post by adjusting his beliefs about the cost of effort needed to fulfill his duties - the need for self-justification, and the adjustment in beliefs, are greater the lower is his past wage. The employer can exploit this by offering a contract with an increasing wage profile. Goetzmann and Peles [15] argue that cognitive dissonance leads investors to justify remaining in mutual funds that consistently perform poorly; and that such inertia can help to explain why money flows in more rapidly to mutual funds that have performed well than flows out from those that have performed poorly.

With regard to modeling, we have already acknowledged our debt to GP. Their model does not apply directly, however. One difference is that while they study preferences over menus of lotteries, it is important for our story that menus consist of (Anscombe-Aumann) acts. Kopylov [19] has extended the GP theorem from (menus of) lotteries to abstract mixture spaces, including, in particular, the space of Anscombe-Aumann acts. Moreover, his alternative proof strategy is adapted in proving our main result. A second and more important formal difference from GP, and also from Kopylov's extension, and the major source of technical difficulty in our model, is that we relax the Independence axiom - the latter is not intuitive given cognitive dissonance. Finally, we note that Dekel, Lipman and Rustichini [7] generalize GP's model of temptation. However, their motivation is much different than ours - in particular, they assume Independence.[1]

There are two final but important connections to the literature. The more optimistic beliefs held ex post by our agent come about in our model because she uses a (nonsingleton) *set* of probability measures, and when evaluating a prospect, she chooses the measure that maximizes its utility. This recalls Dreze's [8] model of choice between Anscombe-Aumann acts under moral hazard. It recalls also Gilboa and Schmeidler [14] - they model agents who are averse to ambiguity, in the sense illustrated by the Ellsberg Paradox, by assuming that they minimize (rather than maximize) over a set of priors, but their model has an obvious counterpart for ambiguity loving. Our model differs from both of these primarily through its focus on the time-varying nature of beliefs and the corresponding value of commitment.

---

[1] In their concluding remarks about possible directions for further research, they mention that accommodating guilt may be a reason for relaxing Independence when modeling temptation. This rationale is obviously much different than ours. Fudenberg and Levine [13] do not impose Independence in the context of their non-axiomatic model involving temptation. We are aware of only one other representation result in the menus-of-lotteries/acts setting that does not rely on Independence. Epstein and Marinacci [10] study an agent who is not subject to temptation, but rather values flexibility because she is uncertain about the future. She violates Independence because her conception of the future is coarse.

However, one could reinterpret these changes in terms of changes in ambiguity (loving), for example.

## 2. UTILITY

The model has the following primitives:

- time $t = 0, 1, 2$
- finite state space $S$
- $\mathcal{C}$: set of (Borel) probability measures over a compact metric space
  refer to $c \in \mathcal{C}$ as a lottery over consumption, or more briefly as consumption
  $\mathcal{C}$ is compact metric under the weak convergence topology
- $\mathcal{H}$: set of acts $h : S \longrightarrow \mathcal{C}$, with the usual mixture operation
- compact sets of acts are called *menus* and denoted $A,\ B,\ \ldots$
  $\mathcal{K}(\mathcal{H})$ is the set of all menus
  it is compact metric under the Hausdorff metric[2]
- preference $\succeq$ defined on $\mathcal{K}(\mathcal{H})$

The interpretation is that a menu $A$ is chosen *ex ante* (at time 0) according to $\succeq$. This choice is made with the understanding that at the unmodeled *ex post* stage (time 1), the agent will choose an act from $A$. Uncertainty is resolved and consumption is realized in the terminal period $t = 2$. Cognitive dissonance and choice behavior at time 1 are anticipated ex ante and underlie the ranking $\succeq$.

Menus are natural objects of choice.[3] The consequence of a physical action taken at time 0 is that it determines a feasible set of physical actions at time 1, and these actions can be modeled by acts in the usual way. Thus each physical action at time 0 corresponds to a menu of acts.

---

[2] See [2, Theorem 3.58], for example.

[3] Kreps [21, 23] was the first to propose menus as a way to model physical actions in an ex ante stage.

Our model of utility has the form[4]

$$\mathcal{U}(A) = \max_{h \in A} \left[ (1-\kappa) U(h) + \kappa V(h) \right] - \kappa \max_{h' \in A} V(h'), \tag{2.1}$$

where

$$U(h) = \ p \cdot u(h), \text{ and} \tag{2.2}$$

$$V(h) = \max_{q \in Q} \ q \cdot u(h). \tag{2.3}$$

Here $0 \leq \kappa \leq 1$, $p$ is a probability measure on $S$, $Q$ is a closed and convex set of probability measures on $S$ containing $p$, and $u : \mathcal{C} \longrightarrow \mathbb{R}^1$ is mixture linear and continuous.

The standard model of subjective expected utility maximization is the special case where $\kappa = 0$ or $Q = \{p\}$. More generally, the functional form can be interpreted along the lines suggested by GP. When restricted to singletons, $\mathcal{U}$ coincides (ordinally) with $U$; thus expected utility with prior $p$ represents preference over consumption lotteries when the agent can commit ex ante. When she does not commit, then the new (temptation) utility function $V$ over lotteries becomes relevant. Temptation utility is computed by maximizing over probability measures in the set $Q$. Since $p \in Q$, $V$ imputes higher expected utility to the menu at hand than was the case ex ante using $p$, corresponding to cognitive dissonance. She is tempted to maximize $V$ ex post. Though she recognizes that $p$ is "correct", there is a self-control cost of resisting the temptation given by

$$\kappa \left( V(h) - \max_{h' \in A} V(h') \right) \leq 0.$$

Thus a compromise is struck between maximizing $U$ and maximizing $V$ - choice out of $A$ is described by maximization of the weighted sum, or by solving

$$\max_{h \in A} \ \max_{q \in (1-\kappa)\{p\} + \kappa Q} q \cdot u(h). \tag{2.4}$$

which balances ex ante realism and ex post cognitive dissonance. The nature of the compromise is further illustrated by the fact that

$$p \in (1-\kappa)\{p\} + \kappa Q \subset Q,$$

---

[4] For any real-valued random variable $x$ on $S$, and probability measure $q$, $q \cdot x$ is short-hand for the expected value $\int_S x dq$.

so that the set of beliefs underlying the choice of an act ex post lies "between" the prior view $p$ and the optimistic view represented by $Q$.[5]

Since ex post choice out of the menu maximizes the utility function $\max_{q\in(1-\kappa)\{p\}+\kappa Q} q\cdot u(h)$, which does not depend on the menu, one may wonder whether the model captures beliefs that adjust to make the previously chosen menu attractive. To see a sense in which this is true, note that (by reversing the order of the maximizations),

$$\max_{h\in A} \max_{q\in(1-\kappa)\{p\}+\kappa Q} q\cdot u(h) = \max_{h\in A} q_A^* \cdot u(h),$$

for any $q_A^*$ that solves $\max_{q\in(1-\kappa)\{p\}+\kappa Q} \max_{h\in A} q\cdot u(h)$. Thus ex post choice conforms with SEU and probabilistic beliefs given by $q_A^*$. Evidently, $q_A^*$ depends on the menu $A$ and is chosen to make the value of the menu, given by $\max_{h\in A} q\cdot u(h)$, as large as possible.

We can say something about the qualitative difference between ex post choice and the "correct" choice. Given any menu $A$ ex post, the choice out of $A$ is determined by maximizing $\max_{q\in(1-\kappa)\{p\}+\kappa Q} q\cdot u(h)$, while the "correct" choice would maximize $p\cdot u(h)$. Suppose for concreteness that consumption is real-valued ($\mathcal{C}$ consists of lotteries over $[a,b]\subset\mathbb{R}^1$) - typically, one assumes that $u(\cdot)$ is concave on $[a,b]$ corresponding to risk aversion. On the other hand, the maximization over $q$ in the ex post utility function introduces some convexity. Thus it may not be concave in $h$ and may even exhibit risk loving. (For example, if one restricts attention to Savage acts $h: S\to[a,b]$, then ex post utility is a convex function of $h$ if $u$ is linear.) Consequently, ex post choice may appear extreme - for example, it may correspond more to boundary optima.

A final comment is that both subjective and objective probabilities are present in the model - the latter underlie consumption lotteries - but they are treated differently: while the agent chooses new beliefs ex post about her subjective uncertainty (the state space $S$), she does not distort or modify objective probabilities. For example, both $U$ and $V$ agree about the ranking of lotteries in that, for every lottery $c$, $U(c)=V(c)=u(c)$, the vNM expected utility of $c$. Because an objective probability law is based on undeniable fact, distorting it to a more favorable one, is folly or ignorance that would not be undertaken by the sophisticated individuals that we model. But where facts alone do not pin down beliefs uniquely,

---

[5] As is familiar from GP-style models, this interpretation in terms of ex post choice is suggested by the functional form, and by intuition for the underlying axioms, but ex post choice lies outside the scope of our formal model. See Noor [25] and Epstein, Noor and Sandroni [11] for models of temptation where ex post choice (or preference) is part of the primitives.

an agent is free to choose beliefs and feeling good about oneself is one possible consideration in doing so. As an illustration of the difference, note that Knox and Inkster [18] report that persons leaving the betting window after placing bets at a race-track are more optimistic about "their horse" than persons about to place bets.[6] On the other hand, it is more difficult to imagine someone being similarly optimistic about a coin, which is known to be unbiased, after choosing that coin for a game of chance.

## 3. AXIOMS

The first two axioms require no discussion.

**Axiom 1 (Order).** $\succeq$ *is complete and transitive.*

**Axiom 2 (Continuity).** $\succeq$ *is continuous.*

Menus can be mixed via

$$\alpha A + (1-\alpha) B = \{\alpha f + (1-\alpha) g : f \in A,\ g \in B\} .$$

Formally, the indicated mixture of $A$ and $B$ is another menu and thus when the agent contemplates that menu ex ante, she anticipates choosing out of $\alpha A + (1-\alpha) B$ ex post. It follows that one should think of the randomization corresponding to the $\alpha$ and $(1-\alpha)$ weights as taking place at the end - after she has chosen some mixed act $\alpha f + (1-\alpha) g$ out of the menu.

The above mixture operation permits one to state the Independence axiom, and it is adopted by GP. However, Independence is *not* intuitive under cognitive dissonance. Because our agent anticipates that she will adjust her beliefs ex post to the menu at hand, she will violate Independence. One way to see this is to recognize that because beliefs are chosen before the lotteries are played out, lotteries constitute "temporal risks". As explained by Machina [24], for example, preferences over temporal risks typically violate Independence even at a normative level. We now elaborate on this argument against Independence.

The familiar normative argument for Independence involves two steps. The first is to consider in place of the mixtures shown, rather the two-stage objects $\alpha \circ A' + (1-\alpha) \circ B$ and $\alpha \circ A + (1-\alpha) \circ B$, where the latter indicates that the agent

---

[6] Aronson cites this experiment.

receives menu $A$ with probability $\alpha$ and $B$ otherwise. (The randomization corresponding to $\alpha$ is completed before she needs to choose from the realized menu.) When interpreted in terms of these two-stage menus, the (modified) Independence axiom has compelling intuition. The key, therefore, is to understand why the agent should (or should not) be indifferent between the set mixture $\alpha A + (1-\alpha) B$ and the two-stage menu $\alpha \circ A + (1-\alpha) \circ B$.

Suppose that $A = \{f, g\}$ and $B = \{\overline{h}\}$.[7] When facing $\alpha \circ A + (1-\alpha) \circ B$, she knows that whatever the outcome of the randomization, she will choose her preferred act from the realized menu. Suppose that if $A$ is realized, she finds it optimal to choose $f$. Then randomization over menus effectively gives her $\alpha \circ \{f\} + (1-\alpha) \circ \{\overline{h}\}$. Suppose the latter is indifferent to the mixed act $\alpha f + (1-\alpha)\overline{h}$; if not, then the case against Independence is even stronger. On the other hand, when evaluating the menu $\alpha A + (1-\alpha) B$, she anticipates that she will choose either $\alpha f + (1-\alpha)\overline{h}$ or $\alpha g + (1-\alpha)\overline{h}$. But under cognitive dissonance, she may choose $\alpha g + (1-\alpha)\overline{h}$ over $\alpha f + (1-\alpha)\overline{h}$, even while choosing $f$ from $A = \{f, g\}$. This is because she adjusts beliefs ex post to the menu at hand, so that she may use different beliefs when holding the menu $A$ than when holding $\alpha A + (1-\alpha) B$. Assume this is in fact the case. Then the two-stage menu leads (up to indifference) to $\alpha f + (1-\alpha)\overline{h}$, while the mixture $\alpha A + (1-\alpha) B$ leads to $\alpha g + (1-\alpha)\overline{h}$. The agent anticipates this. She also anticipates the temptation and self-control costs involved in each case. These also affect utility, but there is no reason to expect them to perfectly offset the utility difference between the two ultimately chosen acts. Thus she will in general not be indifferent between $\alpha \circ A + (1-\alpha) \circ B$ and $\alpha A + (1-\alpha)B$. As a result, she will generally violate Independence.

However, suitable relaxations of Independence are intuitive. First, when ranking singleton menus, there is no choice to be made ex post. Thus cognitive dissonance has no effect on the ranking of singletons, and there is no reason for Independence to be violated.

**Axiom 3 (Commitment Independence).** *For all* $f, g, h \in \mathcal{H}$ *and* $\alpha \in [0,1]$,

$$\{f\} \succeq \{g\} \Longrightarrow \{\alpha f + (1-\alpha) h\} \succeq \{\alpha g + (1-\alpha) h\}.$$

The next axiom, which also relaxes Independence, uses the following notion: say that the acts $h'$ and $h$ are *collinear* if[8]

$$h' = t' f + (1-t') c' \text{ and } h = t f + (1-t) c,$$

---

[7] We adapt the example used in [6] to motivate Independence in their setting.

[8] It is easy to see that for every $s'$ and $s$, $u(h'(s')) > u(h'(s)) \implies u(h(s')) \geq u(h(s))$, that is, the real-valued functions $u(h(\cdot))$ and $u(h'(\cdot))$ are comonotonic. Collinearity implies

for some $f \in \mathcal{H}$, $c', c \in \mathcal{C}$ and $t', t \in [0,1]$. Then for any mixture linear $u$ and for all probability measures $q'$ and $q$,

$$q' \cdot u(h') > q \cdot u(h') \quad \Longrightarrow \quad q' \cdot u(h) \geq q \cdot u(h).$$

Thus if $h'$ is strictly more attractive (has strictly larger expected utility) under $q'$ than under $q$, then $q'$ renders $h$ at least as attractive as does $q$. It follows that even if the agent chooses beliefs (over some subjective set) to fit the act, there is an optimistic measure that is common to both $h'$ and $h$.

Say that menus $A$ and $B$ are *collinear* if $h'$ and $h$ are collinear for any selections $h' \in A'$ and $h \in A$. Then it is intuitive that there exists an optimistic probability measure that is optimal for both $A'$ and $A$: the decision-maker ex ante thinks ahead and imagines choosing out of $A'$. She knows that she will choose an act from the menu that maximizes expected utility using the (optimistic) belief that makes $A'$ as attractive as possible. Let $h'$ be any act that is chosen ultimately. Then an optimistic belief is any belief that maximizes the expected utility of $h'$. Similarly for $A$ and a chosen act $h$. However, $h'$ and $h$ are collinear, which completes the intuition.

The existence of a common optimistic measure for collinear menus invalidates the reason given above for violating Independence and motivates:

**Axiom 4 (Collinear Independence).** *If $A', A$ and $B$ are menus such that $B$ is collinear with both $A'$ and $A$, then for all* $0 < \alpha < 1$,

$$A' \succeq A \Longleftrightarrow \alpha A' + (1-\alpha) B \succeq \alpha A + (1-\alpha) B.$$

Any act is collinear with any constant act, and hence any menu is collinear with any menu of constant acts. Thus the axiom has the following important implication: if $A', A$ and $C$ are binary menus and $C \subset \mathcal{C}$, then for all $0 < \alpha < 1$,

$$A' \succeq A \Longleftrightarrow \alpha A' + (1-\alpha) C \succeq \alpha A + (1-\alpha) C. \tag{3.1}$$

In the standard model, a menu is as good as the best alternative that it contains. Then

$$A \succeq B \Longrightarrow A \sim A \cup B,$$

---

the stronger restriction $t\,(u(h'(s')) - u(h'(s))) = t'\,(u(h(s')) - u(h(s)))$. Thus collinearity can be viewed as a cardinal counterpart of comonotonicity.

a property called *strategic rationality* by Kreps [22]. Such a model excludes temptations. GP show that temptation and self-control can be modeled by relaxing strategic rationality to Set-Betweenness. Temptation is an integral part of cognitive dissonance because the agent changes beliefs to make the menu at hand look attractive and then is tempted to make subsequent choices accordingly (see the discussion of utility in Section 2). Thus we adopt:

**Axiom 5 (Set-Betweenness).** *For all* $A, B \in \mathcal{K}(\mathcal{H})$,

$$A \succeq B \Longrightarrow A \succeq A \cup B \succeq B.$$

Since the axiom excludes some forms of temptation, we need to examine its appeal in the context of our story about why temptation arises. GP identify hypotheses about temptation that imply Set-Betweenness. These include: (H1) alternatives (acts in our case, lotteries in theirs) can be ranked according to how tempting they are; and (H2) the temptation caused by a menu equals the temptation caused by the most tempting alternative it contains.[9] To understand these better in our setting, we elaborate on what we have in mind by "choosing beliefs to make the menu attractive."

Think of the agent as having ex post preferences over acts that conform to SEU, with some vNM index $u$, and with beliefs that she chooses. Given a menu $A$, she chooses those beliefs $q^*$ that maximize the best expected utility achievable in $A$, that is, $q^*$ maximizes $\max_{h \in A} q \cdot u(h)$. These beliefs cause the temptation to choose $h^* \in \arg\max_{h \in A} q^* \cdot u(h)$ ex post, and the temptation caused by the menu can be measured by $q^* \cdot u(h^*)$. Then the temptation caused by any single act $h$ can be represented by the maximum expected utility of $h$ as one varies over all possible beliefs $q$ - thus temptation can be ranked as in H1. To this point we have been silent on the issue of the 'feasible set' from which she chooses beliefs. This set, denoted $Q$, can be any fixed subset of $\Delta(S)$, where 'fixed' means independent of the menu under consideration. Then the preceding description implies H2 immediately (while both H2 and Set-Betweenness are suspect if $Q$ can vary with the menu). Does it make sense to think of $Q$ as fixed across menus? It does if one thinks of $Q$ as describing a "psychologically feasible set" of beliefs. Then if she can choose $q$ given menu $A$, why would she not be able to choose it also given menu $B$? On the other hand, H2 is undoubtedly restrictive, as GP were well aware

---

[9] Another underlying hypothesis is that unchosen alternatives cannot increase utility. This excludes a utility gain from feeling virtuous at having resisted a temptation, but we find this to be an acceptable exclusion.

and as demonstrated further by the examples in Dekel, Lipman and Rustichini [7]. However, we do not find it to be especially restrictive in our setting and thus suppress it in the sequel.[10]

Set-Betweenness captures only partially our (GP-based) underlying notion of temptation; for example, GP show that Set-Betweenness is implied by the hypotheses described above, but these may also have other implications for preference. Thus we adopt an additional axiom about the nature of temptation implicit in preference. The axiom is inspired by Kreps' [21] central axiom. Translated into our setting, his axiom states: for all menus $A$ and $A'$,

$$A \sim A \cup A' \quad \Rightarrow \quad A \cup B \sim (A \cup B) \cup A', \text{ for all } B.$$

In Kreps' model, the agent has some subjective uncertainty about the future and thus values flexibility. If $A \sim A \cup A'$, then the flexibility gained by adding $A'$ to $A$ is of no value. But then $A'$ cannot provide any useful flexibility when added to $A \cup B$, that is, one would expect $A \cup B \sim (A \cup B) \cup A'$. This implication is not intuitive, however, if the reason for violating strategic rationality is the anticipation of temptation rather than the presence of subjective uncertainty. For example, if $\{f\} \sim \{f'\}$, then an agent satisfying Set-Betweenness would also exhibit $\{f\} \sim \{f, f'\}$. Nevertheless, one would not expect her necessarily to be indifferent between $\{f, g\}$ and $\{f, g, f'\}$ for any third act $g$. For example, if $g$ is ranked highest under commitment, then $\{f, g, f'\}$ could be ranked strictly lower than $\{f, g\}$ if $f'$ is strictly more tempting than $f$ , which is not ruled out by the hypothesis that $\{f\} \sim \{f'\}$. In fact, the latter indifference says nothing at all about how tempting are $f$ and $f'$, which is the reason that Kreps' axiom is not intuitive in our model. However, we adopt the following weakening:

**Axiom 6 (Ex Post Nash-Chernoff (NC)).** *For all menus $A, A', B, \in \mathcal{A}$,*

$$A \sim A \cup A' \not\sim A' \quad \Rightarrow \quad A \cup B \sim (A \cup A') \cup B.$$

Continuing for concreteness to consider the case where all three menus are singletons, suppose that $\{f\} \sim \{f, f'\} \succ \{f'\}$. Then $f'$ does not tempt $f$ and given the menu $\{f, f'\}$ at the ex post stage, the agent would choose (only) the normatively best alternative $f$. Add a third hypothesis about the nature of temptation (H3): *given that she would not choose $f'$ out of one menu containing $f$* (say

---

[10] See both Dekel, Lipman and Rustichini [7] and Kopylov [19] for representation results using a weakening of Set-Betweenness.

$\{f, f'\}$), *then neither would she choose* $f'$ *out of any larger menu* (for example, out of $\{f, f', g\}$). Readers will recognize this condition as the Nash-Chernoff condition (or Sen's property $\alpha$) for ex post choice behavior, which partially explains the name chosen for the axiom.[11] Suppose that the agent has rejected $f'$ in favor of $f$ at the ex post stage when deciding on which act to choose out of the given menu $A$. Adding other acts to $A$ should not affect the normative appeal of $f$ and $f'$, but it may change the self-control costs associated with choosing them. H3 is implied if the change in self-control costs balance one another so that the agent would prefer to choose $f$ rather than $f'$ also out of the larger menu.

Now we can explain indifference between $\{f, g\}$ and $\{f, g, f'\}$ as follows: by H1, alternatives can be ranked according to how tempting they are. In the present case, $f$ is at least as tempting as $f'$. Therefore, $f'$ does not affect the temptation involved in choosing out of $\{f, g, f'\}$ since, by H2, that temptation depends only on the most tempting alternative, either $f$ or $g$. By H3, neither would $f'$ be chosen out of $\{f, g, f'\}$. It follows that the presence of $f'$ in $\{f, g, f'\}$ is of no significance, and thus one would expect $\{f, g\}$ and $\{f, g, f'\}$ to be indifferent.

Suppose next that $\{f'\} \succ \{f, f'\} \sim \{f\}$, that is, $f'$ is better under commitment, but $f$ is more tempting and the agent yields to the temptation and chooses (only) $f$ out of $\{f', f\}$. Then since $f'$ would not be chosen out of $\{f, g, f'\}$, and since $f'$ is not the most tempting alternative in $\{f, g, f'\}$, the desired indifference follows.

As for the relation between NC and Set-Betweenness, the hypotheses H1 and H2 underlie them both but H3 is invoked only in order to derive (or explicate) NC. In that limited sense, H3 is the new element in NC. Note, however, that NC incorporates elements of both H1 and H2, as well as H3.

If Independence is assumed, then NC is implied by the remaining axioms as in GP [16] and Kopylov [19]. We conjecture (but have not yet proven) that NC is not redundant in our model.[12] It plays a big part in the proof of our representation

---

[11] Sen's condition $\beta$ states that if $f$ and $f'$ are both chosen in $A$ and $f$ is chosen in $B \supset A$, then $f'$ is also chosen in $B$. This condition is not intuitive for the same reason that Kreps' original axiom is problematic; for example, consider the case $\{f\} \sim A = \{f, f'\} \sim \{f'\}$. The weak axiom of revealed preference (WARP) requires both conditions $\alpha$ and $\beta$. In his model where choice out of menus, rather than ex ante preference, is the primitive, Noor [25] adopts WARP as a formal axiom. He also provides an example to illustrate why condition $\alpha$, and hence also WARP, may be problematic in a model of temptation.

[12] Set-Betweenness alone does not imply NC. This is most easily demonstrated in a modified setting when the domain of preference consists of all subsets of a finite set $X$ of alternatives. For example, let $X = \{x, y, z\}$ and define preference by: $\{x\} \sim \{x, y\} \succ \{y\} \sim \{y, z\} \sim$

result. In the two cited papers, Independence is used heavily in the proofs, in part in order to extend globally representations for preference that apply on limited subdomains. NC plays a similar role here (see particularly Lemma A.4).

To proceed, say that the pair of acts $(h, g)$ is *interior* if

$$\{h\} \succ \{h, g\} \succ \{g\},$$

that is, following GP, if $g$ tempts $h$ ($\{h\} \succ \{h, g\}$) but the temptation is resisted and $h$ is chosen ($\{h, g\} \succ \{g\}$).[13] Say that $f$ *dominates* $g$ if

$$\{f(s)\} \succeq \{g(s)\} \text{ for every } s.$$

**Axiom 7 (Monotonicity).** *If $f$ dominates $g$, then: (i) $\{f\} \succeq \{g\}$; and (ii) $\{c, g\} \succeq \{c, f\}$ for any constant act $c$ such that both $(c, f)$ and $(c, g)$ are interior.*

A dominating act is preferred under commitment - this is part (i). But dominance works in the opposite direction in (ii). Since $(c, g)$ is interior, $g$ is tempting but is not chosen ex post. Thus the utility of $\{c, g\}$ depends on $g$ only through the temptation it provides; similarly for $\{c, f\}$. If $f$ dominates $g$, then $f$ is more tempting and hence $\{c, g\} \succeq \{c, f\}$.

Our axioms thus far have for the most part been concerned with modeling temptation in general, that is, not tied specifically to cognitive dissonance. A partial exception is Collinear Independence, the intuition for which did rely on the assumption that temptation arises because of an ex post choice of beliefs to "fit the menu" in hand. However, Collinear Independence would be satisfied even if the agent becomes more pessimistic ex post and adopts beliefs that make the menu less attractive ex post. The final two axioms build in ex post optimism and hence cognitive dissonance.

**Axiom 8 (Constants-Do-Not-Tempt).** *For all $A, B, C \in \mathcal{K}(\mathcal{H})$ with $C \subset \mathcal{C}$,*

$$A \succeq C \Longrightarrow A \sim A \cup C \succeq C.$$

---

$\{x, y, z\} \succ \{x, z\} \succ \{z\}$. In our framework, Set-Betweenness but not NC, is satisfied if we generalize (2.1) by allowing $\kappa$ to vary suitably with the menu via $\kappa(A) = \widehat{\kappa}(\max_{h \in A} V(h))$, where $\widehat{\kappa}(\cdot)$ is increasing; we are grateful to Jawwad Noor for this observation.

[13] GP say that there is self-control at $\{h, g\}$ if either $\{h\} \succ \{h, g\} \succ \{g\}$ or $\{g\} \succ \{h, g\} \succ \{h\}$. We introduce new terminology so as to more easily specify which of these rankings is relevant.

Temptation is due to a change in beliefs (as opposed to a change in risk aversion, for example), which leaves the evaluation of constant acts unaffected. In addition, the noted change is always to become more optimistic ex post about the available menu, rendering it even more attractive relative to $C$ than it was ex ante. Therefore, menus of constant acts cannot tempt. Note that, in contrast, $C \succ A \cup C \succeq A$ is both permitted by the axiom and intuitive given our story.

Suppose now that $(c, f)$ and $(c, g)$ are both interior, where $c$ is a constant act; in particular, both $f$ and $g$ tempt $c$. Suppose also that

$$\{c, f\} \sim \{c, g\}. \tag{3.2}$$

Since $(c, f)$ is interior, $c$ is chosen ex post. This means that $f$ affects the utility of $\{c, f\}$ only through the temptation it provides. Therefore, (3.2) indicates that $f$ and $g$ are equally tempting. Because ex post beliefs are chosen to make the menu attractive, and because the expected utility of $c$ does not depend on beliefs, we can further interpret (3.2) as follows: the act $f$, when matched with the beliefs that make it attractive, is equally tempting as $g$, given the adoption of beliefs that make it attractive. Consider now the menu $\{c, \alpha f + (1 - \alpha)g\}$. Beliefs to render this menu attractive are chosen ex post (time 1), before the randomization is completed (which, as noted earlier, occurs only at the terminal time after the true state in $S$ is realized). Since *the beliefs that make $f$ attractive may differ from those that make $g$ attractive*, matching beliefs with the mixed act is more difficult. Therefore, one would expect the mixed act either not to tempt $c$, that is, $\{c\} \sim \{c, \alpha f + (1 - \alpha)g\}$, or else to be tempting but less so than $f$ and $g$. Thus we assume:

**Axiom 9 (Convex Temptation).** *For all interior pairs $(c, f)$ and $(c, g)$,*

$$\{c, f\} \sim \{c, g\} \Longrightarrow \{c, \alpha f + (1 - \alpha)g\} \succeq \{c, f\}.$$

## 4. REPRESENTATION RESULT

Our main result is that the preceding axioms characterize the functional form described in Section 2.

**Theorem 4.1.** *The binary relation $\succeq$ on $\mathcal{K}(\mathcal{H})$ may be represented as in (2.1)-(2.3) if and only if it satisfies Axioms 1-9. Moreover, $u$ is unique up to a positive linear transformation, and if $\succeq$ is not strategically rational, then $p$, $Q$ and $\kappa$ are unique.*

Convex Temptation is used only at the very end of the sufficiency proof in order to prove that $V$ has the form given in (2.3). If the axiom is deleted, then the remaining axioms characterize the representation (2.1)-(2.2), for some $V:\mathcal{H}\to\mathbb{R}^1$ that is continuous, monotone ($V(f)\geq V(g)$ if $f$ dominates $g$), satisfies certainty additivity ($V(\alpha f+(1-\alpha)c)=\alpha V(f)+(1-\alpha)V(c)$ for all $c$ in $\mathcal{C}$), and that satisfies $V(f)\geq p\cdot u(f)$ for all $f$ with equality if $f$ is constant. (See Example 3 below.)

We present some examples to demonstrate the tightness of the characterization in the theorem. Each example satisfies Order, Continuity, Commitment Independence, Set-Betweenness, Ex Post Nash-Chernoff and Monotonicity, and violates precisely one of the axioms that relate specifically to cognitive dissonance - Collinear Independence, Constants-Do-Not-Tempt, and Convex Temptation.

*Example 1*: Let

$$\mathcal{U}(A)=\frac{\max_{h\in A}[U(h)V(h)]}{\max_{h'\in A}V(h')},$$

where $U$ and $V$ are as in (2.2)-(2.3), and where $u>0$. Then $\succeq$ violates only Collinear Independence. (See Appendix C for details.)

There exist simpler examples violating only Collinear Independence - these retain (2.1)-(2.2) but modify the specification of $V$. However, because the above ratio form deviates from the GP functional form, we find it more revealing about the power of Collinear Independence.[14]

*Example 2*: Assume (2.1)-(2.2), but take

$$V(h)=\ q\cdot u(h),$$

for some probability measure $q\neq p$. Then $\succeq$ violates only Constants-Do-Not-Tempt.

*Example 3*: Modify Example 2 by taking

$$V(h)=\max\left\{p\cdot u(f),\ \int u(f)\,d\nu\right\},$$

---

[14] The example is inspired by weighted utility theory [5], a model of risk preference in which the utility function over lotteries equals a ratio of expected utility functions. Readers familiar with the 'non-expected utility' literature will not be surprised by the observation that $\succeq$ satisfies the following alternative relaxation of Independence: $A\sim B\Longrightarrow\ \alpha A+(1-\alpha)B\sim A$.

where $\nu$ is a capacity on $S$ and the integral $\int u(f)\,d\nu$ is in the sense of Choquet (see Schmeidler [29]). Then $\succeq$ violates only Convex Temptation.

A tuple $(u, p, Q, \kappa)$ as in the theorem is said to represent $\succeq$. The representing tuple is unique (up to cardinal equivalence for $u$) if the degenerate case of strategic rationality is excluded. Thus it is meaningful to ask about behavioral interpretations of its components. We have already noted those of $u$ and $p$: $u$ ranks lotteries (constant acts) and $p$ is the "commitment prior" - it underlies the ranking of singleton menus. Turn to $Q$ and $\kappa$. In what follows, we adopt variants of GP's comparative notions "greater preference for commitment" and "greater self-control", renamed so as to reflect better the psychological motives we have in mind.

Say that $\succeq^*$ *has greater dissonance than* $\succeq$ if for all acts $f$ and $g$,

$$\{f\} \succ \{f, g\} \Longrightarrow \{f\} \succ^* \{f, g\}. \tag{4.1}$$

The ranking $\{f\} \succ \{f, g\}$ indicates that though $f$ is better than $g$ ex ante, $g$ is better ex post when holding the menu $\{f, g\}$. Then there is dissonance for the agent with preference $\succeq$ between the ex ante ranking under commitment (or the underlying beliefs) and the distinct ex post ranking (or ex post beliefs). If $\succeq^*$ has greater dissonance, then she should strictly prefer $\{f\}$ to $\{f, g\}$.

**Theorem 4.2.** *Suppose that both $\succeq$ and $\succeq^*$ have utility representations (2.1)-(2.3), with components $(u, p, Q, \kappa)$ and $(u^*, p^*, Q^*, \kappa^*)$ respectively, and that neither is strategically rational. Then $\succeq^*$ has greater dissonance than $\succeq$ if and only if*

$$(u, p) = (au^* + b, p^*) \text{ for some } a > 0 \text{ and some } b, \text{ and} \tag{4.2}$$

$$Q = (1 - \epsilon)\{p\} + \epsilon Q^*, \text{ for some } 0 < \epsilon \leq 1. \tag{4.3}$$

The characterizing conditions assert both that the commitment rankings induced by $\succeq$ and $\succeq^*$ coincide (this is (4.2)) and that $Q$ is "closer to $p$" than is $Q^*$ in the sense of an epsilon contamination (this is (4.3)). Note that if $\succeq$ is strategically rational, then any $\succeq^*$ has greater dissonance - the defining condition is satisfied vacuously - and no restrictions on commitment preferences are implied. If $\succeq^*$ is strategically rational, then (4.1) is satisfied if and only if $\succeq$ is also strategically rational, and again, condition (4.2) is not implied.

We are interested not only in how much dissonance an agent experiences (or expects to experience), but also in what she does about it, or more precisely, in the extent to which ex post choices are distorted by dissonance. Say that $\succeq^*$ *is more self-justifying than* $\succeq$ if it has more dissonance than $\succeq$ and

$$\{f\} \succ \{f,g\} \sim \{g\} \implies \{f\} \succ^* \{f,g\} \sim^* \{g\}.$$

The hypothesized rankings for $\succeq$ indicate not only that there is dissonance but also that given $\{f,g\}$ at the ex post stage, the agent succumbs and chooses $g$, even though $f$ was optimal ex ante under commitment. She does this because the choice of $g$ better justifies her previous choice of $\{f,g\}$. If $\succeq^*$ is more self-justifying, then she should also choose $g$ out of $\{f,g\}$.

**Theorem 4.3.** *Suppose that both* $\succeq$ *and* $\succeq^*$ *have utility representations (2.1)-(2.3), with components* $(u,p,Q,\kappa)$ *and* $(u^*,p^*,Q^*,\kappa^*)$ *respectively, and that neither is strategically rational. Then* $\succeq^*$ *is more self-justifying than* $\succeq$ *if and only if* $(u^*,p^*,Q^*,\kappa^*)$ *and* $(u,p,Q,\kappa)$ *satisfy (4.2), (4.3) and* $\kappa^* \geq \epsilon\kappa$.

It follows that a change from $\kappa$ to $\kappa^* > \kappa$, keeping other components of the functional form fixed, renders $\succeq^*$ more self-justifying than $\succeq$ but leaves the two preference orders equally dissonant (each has greater dissonance than the other).

## 5. SPECIALIZATIONS AND EXTENSIONS

To conclude, we consider briefly one specialization of the above model and two generalizations.

### 5.1. Distorting Probabilities

Suppose that $Q$ satisfies

$$sup_{q\in Q}\, q(\cdot) = \psi(p(\cdot)) \tag{5.1}$$

for some distortion $\psi : [0,1] \longrightarrow [0,1]$ of $p$.[15] An implication of this specialization is that when choosing out of a menu of bets on different events but with common stakes, the agent chooses to bet on the event having the largest probability according to $p$, just as she would ex ante under commitment. Consequently, she

---

[15] This is true, for example, if $Q = \{q \in \Delta(S) : q(\cdot) \leq \psi(p(\cdot))\}$, where $\psi : [0,1] \longrightarrow [0,1]$ is strictly increasing, concave and $\psi(0) = 0$, $\psi(1) = 1$.

satisfies the requirement that $[A \succeq B \Longrightarrow A \sim A \cup B]$ for all menus $A$ and $B$ consisting exclusively of bets with common stakes. However, dissonance may still affect the ranking of nonbinary acts.

The idea is that the distortions of beliefs involved in ex post optimism are cardinal rather than ordinal. That is, thinking of $p$ as the truth, optimism ex post does not change the ranking of events from their true likelihoods, but takes the form of shifting probability mass to "good" events. Though restrictive, this added structure has the benefit of distinguishing our model from ambiguity loving. Since both ex ante and ex post, the agent ranks bets according to the probability measure $p$, she cannot exhibit either the intuitive Ellsberg-style choices or the counterintuitive (ambiguity loving) opposite choices at either stage.

### 5.2. Effort or Intermediate Consumption

Consumption occurs only at the terminal time in our model. Here we outline an extension of the model (specifically, of the functional form for utility) that permits consumption also in the initial period. Such an extension permits incorporation of the 'cost' of any action chosen ex ante, via foregone initial consumption. As a result, one can address an intuitive prediction of dissonance theory, that cognitive dissonance is more pronounced when past actions are "difficult" (see [3, pp. 175-8] and [1, p. 310] for discussion and references to supporting experimental evidence).

Modify the time line described in Section 2 only by supposing that the choice to made at the ex ante stage is of a pair $(c_0, A)$, where $c_0$ denotes initial consumption (or lottery) and $A$ is, as before, a menu of Anscombe-Aumann acts one of which will be chosen in the following period. Ex ante choices are assumed to maximize preference $\succeq$, which is defined on $\mathcal{C} \times \mathcal{K}(\mathcal{H})$.

Let utility have the form

$$\mathcal{U}(c_0, A) = \max_{h \in A} \left[ U(c_0, h) + \tfrac{\kappa(c_0)}{1-\kappa(c_0)} \, V(c_0, h) \right] - \tfrac{\kappa(c_0)}{1-\kappa(c_0)} \, \max_{h' \in A} V(c_0, h'), \qquad (5.2)$$

where

$$U(c_0, h) = u(c_0) + \delta \; p \cdot u(h), \text{ and} \qquad (5.3)$$

$$V(c_0, h) = u(c_0) + \delta \max_{q \in Q} \; q \cdot u(h). \qquad (5.4)$$

Here $0 \leq \kappa(c_0) < 1$, $p$ is a probability measure on $S$, $Q$ is a convex and compact set of probability measures on $S$ containing $p$, $u : \mathcal{C} \longrightarrow \mathbb{R}^1$ is mixture linear and continuous, and $0 < \delta < 1$ is a discount factor. When restricted to singletons,

$$\mathcal{U}(c_0, \{h\}) = u(c_0) + \delta p \cdot u(h).$$

For nonsingletons, ex post choice out of $A$ solves

$$\max_{h\in A}\ \max_{q\in(1-\kappa(c_0))\{p\}+\kappa(c_0)Q} q\cdot u(h),$$

which depends on $c_0$ via $\kappa(\cdot)$.

Suppose that

$$\kappa(c_0)=\widehat{\kappa}(u(c_0)),$$

where $\widehat{\kappa}(\cdot)$ is decreasing. Then a fall in $u(c_0)$ renders the agent more self-justifying, but leaves the level of dissonance unchanged.[16] More generally, we could also specify $Q$ as a function of $u(c_0)$, for example,

$$Q=(1-\epsilon(u(c_0)))\{p\}+\epsilon(u(c_0))\Delta(S).$$

If $\widehat{\kappa}(\cdot)\epsilon(\cdot)$ is decreasing, then a fall in $u(c_0)$ implies both greater dissonance and greater self-justification.[17]

### 5.3. Response to Information

Justification of a past decision involves also the reaction to information - dissonance theory predicts that information is interpreted in a way that is favorable to past choices. By adding a signal realized at time 1 and building on Epstein [9], we can extend our model to capture also the response to information.

An outline of the model is as follows: let $S_1$ denote the (finite) space of signals, one of which is realized at time 1. Ex ante, the agent chooses a *contingent menu* - a mapping $F$ from signals into menus of Anscombe-Aumann acts. At time 1, she observes the realized signal, updates her beliefs about $S$, and then chooses an act from the realized menu $F(s_1)$. Denote by $p$ prior beliefs on $S_1\times S$, by $p_1$ its first marginal, and, for each signal $s_1$, let $Q_{s_1}$ be a (closed and convex) set of probability measures on $S$ containing $p(\cdot\mid s_1)$, the Bayesian update of $p$. Then the utility of any contingent menu $F$ is given by

$$\mathcal{W}(F)=\int_{S_1}\mathcal{U}(F(s_1);s_1)\ dp_1(s_1),$$

---

[16] We are using the formal comparative notions defined in the preceding section applied to the preferences on $\mathcal{K}(\mathcal{H})$ induced by $\succeq$ and the two levels of consumption.

[17] If $Q_i=(1-\epsilon_i)\{p\}+\epsilon_i\Delta(S)$, $i=1,2$, with $\epsilon_1\geq\epsilon_2$, then $Q_2=(1-\epsilon)\{p\}+\epsilon Q_1$ with $\epsilon=\epsilon_2/\epsilon_1$. Thus Theorem 4.3 implies that preference 1 is more self-justifying (and has greater dissonance) than preference 2 if $\epsilon_1\kappa_1\geq\epsilon_2\kappa_2$.

where, for any menu $A$,

$$\mathcal{U}(A; s_1) = \max_{h \in A} \left[ (1-\kappa) U(h; s_1) + \kappa V(h; s_1) \right] - \kappa \max_{h' \in A} V(h'; s_1),$$

$$U(h; s_1) = \ p(\cdot \mid s_1) \cdot u(h), \text{ and}$$

$$V(h) = \max_{q \in Q_{s_1}} \ q \cdot u(h).$$

The interpretation is clear given the parallel with our model (2.1)-(2.3). The key is that at the ex post stage, the agent does not rely simply on the Bayesian update $p(\cdot \mid s_1)$ of her prior beliefs, but rather behaves as though she adjusts the latter in a direction that renders the realized menu $F(s_1)$ attractive, as indicated by the maximization over $Q_{s_1}$. As a result the signal is interpreted so as to justify the past choice of an action (that is, $F$).[18]

# A. Appendix: Proof of the Representation Theorem

For necessity, verification of the axioms is straightforward.

The proof of sufficiency proceeds roughly as follows: apply the Anscombe-Aumann Theorem to derive an expected utility function $U : \mathcal{H} \rightarrow \mathbb{R}^1$ for preference restricted to singleton menus. This delivers a linear utility index $u : \mathcal{C} \rightarrow \mathbb{R}^1$ and a prior $p$ on $S$, such that $U(f) = p \cdot u(f)$. Next, for any $f \in \mathcal{H}$, let $\mathcal{H}_f = \{\alpha c + (1-\alpha) f : \alpha \in [0,1], c \in \mathcal{C}\}$, and $\mathcal{A}_f$ be the class of menus in $\mathcal{H}_f$. Then $\mathcal{H}_f$ is a compact mixture space, and $\succeq$ restricted to $\mathcal{A}_f$ satisfies Independence (because $\succeq$ satisfies Collinear Independence) and Set-Betweenness. Thus, by Kopylov's [19, Theorem 2.1] extension of GP's theorem to mixture spaces, one obtains a continuous and linear function $V_f : \mathcal{H}_f \rightarrow \mathbb{R}$ such that

$$\mathcal{U}(A) = \max_{h \in A} (U(h) + V_f(h)) - \max_{h \in A} V_f(h),$$

represents $\succeq$ on $\mathcal{A}_f$. The next step is to show that these local representations on the various $\mathcal{A}_f$'s can be tied together to provide a representation on the set of all menus; NC is critical here. A global representation of the form (2.1) is derived using $U$ and the temptation function $V : \mathcal{H} \rightarrow \mathbb{R}^1$, where

$$V(f) = V_f(f),$$

for all acts $f$ that are "potentially tempting", which means that $\{c\} \succ \{c, h\}$ for some $c \in \mathcal{C}$ and $h \in \mathcal{H}_f$. (See the proof for the definition of $V(f)$ for acts that are not potentially tempting.) The remaining step is to show that $V$ has the form (2.3) for some $Q$, which is done by verifying that $V$ satisfies the Gilboa and Schmeidler [14] axioms (suitably modified for the maxmax model rather than maxmin).

---

[18] A closely related bias, called confirmatory bias, states that people tend to interpret evidence in ways that confirm prior beliefs, as opposed to past actions (see [27], for example).

Turn to the detailed proof. Throughout abbreviate the domain $\mathcal{K}(\mathcal{H})$ by $\mathcal{A}$, and assume that $\succeq$ is *non-degenerate*, that is, $A \succ B$ for some $A, B \in \mathcal{A}$. (Otherwise, the desired representation holds trivially with $u \equiv 0$.)

**Lemma A.1.** *There exist a continuous function $\mathcal{U} : \mathcal{A} \to \mathbb{R}$, a probability measure $p$ on $S$, and a non-constant expected utility function $u : \mathcal{C} \to \mathbb{R}$ such that $\mathcal{U}$ represents $\succeq$ and*

$$\mathcal{U}(\{f\}) = p \cdot u(f) \quad \text{for all } f \in \mathcal{H}. \tag{A.1}$$

*Such $p$ is unique, and $u$ is unique up to a positive linear transformation.*

*Proof.* By the Anscombe–Aumann Theorem, the axioms of Order, Continuity, Monotonicity, and Commitment Independence imply that the preference $\succeq$ restricted to singleton menus can be represented by $\mathcal{U}(\{f\}) = p \cdot u(f)$, where $p$ is a probability measure on $S$, and $u : \mathcal{C} \to \mathbb{R}$ is a continuous vNM expected utility function. As $\mathcal{C}$ is compact, there exist lotteries $c_+, c_- \in \mathcal{C}$ such that $u(c_+) \geq u(c) \geq u(c_-)$ for all $c \in \mathcal{C}$. Then $\{c_+\} \succeq \{f\} \succeq \{c_-\}$ for all $f \in \mathcal{H}$. By Set-Betweenness, $\{c_+\} \succeq A \succeq \{c_-\}$ for all finite menus $A$; by Continuity, $\{c_+\} \succeq A \succeq \{c_-\}$ for all menus $A \in \mathcal{A}$. As $\succeq$ is non-degenerate, $\{c_+\} \succ \{c_-\}$ and hence, $u$ is non-constant.

By Continuity, for any $A \in \mathcal{A}$, there exists a unique $\alpha \in [0, 1]$ such that $A \sim \alpha c_+ + (1 - \alpha) c_-$. Let

$$\mathcal{U}(A) = u(\alpha c_+ + (1 - \alpha) c_-).$$

Then $\mathcal{U}$ represents $\succeq$ on $\mathcal{A}$ and inherits continuity from $\succeq$. $\square$

Hereafter, fix $c_+, c_- \in \mathcal{C}$ as in the proof of the above lemma, and fix the unique $u$ (and the unique corresponding $\mathcal{U}$) such that $u(c_+) = 1$ and $u(c_-) = -1$. Let $c_0 = \frac{c_+ + c_-}{2}$; then $u(c_0) = 0$.

For every act $f \in \mathcal{H}$, let

- $U(f) = p \cdot u(f)$
- $f + \alpha$ denote the mixture $\alpha c_+ + (1 - \alpha) f$
- $f - \alpha$ denote the mixture $\alpha c_- + (1 - \alpha) f$
- $\mathcal{H}_f = \{\alpha c + (1 - \alpha) f : \alpha \in [0, 1], c \in \mathcal{C}\}$; then $f + \alpha,\ f - \alpha \in \mathcal{H}_f$
- $\mathcal{A}_f$ be the set of menus in $\mathcal{H}_f$
- $e(f) = \frac{1 + U(f)}{2} c_+ + \frac{1 - U(f)}{2} c_-$; then $e(f) \in \mathcal{C}$ and $\{f\} \sim \{e(f)\}$

Say that an act $f \in \mathcal{H}$ is *never tempting* if $\{c, h\} \succeq \{c\}$ for all $c \in \mathcal{C}$ and $h \in \mathcal{H}_f$; otherwise, call $f$ *potentially tempting.*

**Lemma A.2.** *If $f \in \mathcal{H}$ is never tempting, then*

$$\{f\} \cup A \sim \{e(f)\} \cup A \quad \text{for all } A \in \mathcal{A}.$$

*Proof.* Let $f \in \mathcal{H}$ be never tempting. By Commitment Independence and Set-Betweenness, for all $\alpha \in (0,1)$,

$$\{f+\alpha\} \succ \{e(f)-\alpha\},\ \{f+\alpha\} \succeq \{f+\alpha, e(f)-\alpha\} \succeq \{e(f)-\alpha\}.$$

But Constants-Do-Not-Tempt implies that $\{f+\alpha\} \preceq \{f+\alpha, e(f)-\alpha\}$, and hence

$$\{f+\alpha\} \sim \{f+\alpha, e(f)-\alpha\} \succ \{e(f)-\alpha\}.$$

Similarly, Commitment Independence, Set-Betweenness and the hypothesis that $f$ is never tempting imply that, for all $\alpha \in (0,1)$,

$$\{e(f)+\alpha\} \sim \{e(f)+\alpha, f-\alpha\} \succ \{f-\alpha\}.$$

Hence NC implies that, for all $A \in \mathcal{A}$,

$$\begin{aligned}&\{f+\alpha\} \cup A \sim \{f+\alpha, e(f)-\alpha\} \cup A, \text{ and}\\ &\{e(f)+\alpha\} \cup A \sim \{e(f)+\alpha, f-\alpha\} \cup A.\end{aligned}$$

Let $\alpha \to 0$ and deduce that $\{f\} \cup A \sim \{f, e(f)\} \cup A \sim \{e(f)\} \cup A$. $\square$

Lemma A.2 and Constants-Do-Not-Tempt imply that if all acts $f \in \mathcal{H}$ are never tempting, then the representation (2.1)-(2.3) with $\kappa = 0$ obtains for finite menus and hence, by Continuity, for all menus. In this case, $\succeq$ satisfies Strategic Rationality.

Thus, assume wlog that there exist potentially tempting acts.

**Lemma A.3.** *If $f \in \mathcal{H}$ is potentially tempting, then for all $A \in \mathcal{A}_f$,*

$$\mathcal{U}(A) = \max_{g \in A}(U(g) + V_f(g)) - \max_{g \in A} V_f(g), \tag{A.2}$$

*where $V_f : \mathcal{H}_f \to \mathbb{R}$ is continuous, linear, non-constant, and $V_f(c_0) = 0$.*

*Moreover, such $V_f$ is unique and satisfies the following properties:*

- *(i) (Monotonicity) if $h \in \mathcal{H}_f$ dominates $h' \in \mathcal{H}_f$, then $V_f(h) \geq V_f(h')$;*
- *(ii) $V_f(\cdot) = \frac{\kappa_f}{1-\kappa_f} U(\cdot)$ on $\mathcal{C}$, for some $\kappa_f \in (0,1)$;*
- *(iii) $V_f(h) \geq V_f(e(h))$ for all $h \in \mathcal{H}_f$;*
- *(iv) $V_f(f) > V_f(e(f))$.*

*Proof.* Fix a potentially tempting act $f \in \mathcal{H}$. Then by Set-Betweenness, $\{c\} \succ \{c, h^*\} \succeq \{h^*\}$ for some $c \in \mathcal{C}$ and $h^* \in \mathcal{H}_f$. We claim that

$$\{c_f\} \succ \{c_f, h_f\} \succ \{h_f\}, \tag{A.3}$$

for some $c_f \in \mathcal{C}$ and $h_f \in \mathcal{H}_f$: we have

$$\mathcal{U}(\{c+1, h^*-1\}) = 1 \geq \mathcal{U}(\{c\}) > \mathcal{U}(\{c+0, h^*-0\}).$$

By Continuity, there exists $\alpha \in [0,1]$ such that

$$\mathcal{U}(\{c\}) > \mathcal{U}(\{c+\alpha, h^* - \alpha\}) > \mathcal{U}(\{c+0, h^* - 0\}),$$

that is, $\{c+\alpha\} \succeq \{c\} \succ \{c+\alpha, h^* - \alpha\} \succ \{c+0, h^* - 0\} \succeq \{h^*\} \succeq \{h^* - \alpha\}$. Thus $c_f = (c+\alpha)$ and $h_f = (h^* + \alpha)$ satisfy (A.3).

The assertion (A.2) now follows from [19, Theorem 2.1]: $\mathcal{H}_f$ is a compact mixture space satisfying properties M1–M4 in [19] and containing $\mathcal{C}$, and $\succeq$ restricted to $\mathcal{A}_f$ satisfies Order, Continuity, Binary Independence, and Set-Betweenness, the assumptions in the cited theorem. Thus $\succeq$ is represented on $\mathcal{A}_f$ by

$$\mathcal{U}_f(A) = \max_{g \in A}(U_f(g) + V_f(g)) - \max_{g \in A} V_f(g),$$

where $U_f : \mathcal{H}_f \to \mathbb{R}$ and $V_f : \mathcal{H}_f \to \mathbb{R}$ are continuous and linear. Up to a positive linear transformation $U_f(\cdot) = U(\cdot)$ on $\mathcal{H}_f$ and hence, $\mathcal{U}_f(\cdot) = \mathcal{U}(\cdot)$ on $\mathcal{A}_f$. In the presence of the ranking (A.3) and the normalization $V_f(c_0) = 0$, [19, Theorem 2.1] establishes that $V_f$ is non-constant and unique.

As $\mathcal{H}_f$ and $[0,1]$ are compact, and $\succeq$ and the mixture operation are continuous, there exists $\alpha > 0$ such that for all $h \in \mathcal{H}_f$,

$$\{c_f\} \succ \{c_f, \alpha h + (1-\alpha) h_f\} \succ \{\alpha h + (1-\alpha) h_f\}.$$

Let $W_f(\cdot) = U(\cdot) + V_f(\cdot)$ on $\mathcal{H}_f$. The utility (A.2) of $\{c_f, \alpha h + (1-\alpha) h_f\}$ is

$$W_f(c_f) - V_f(\alpha h + (1-\alpha) h_f) = -\alpha V_f(h) + \gamma,$$

where $\gamma = W_f(c_f) - (1-\alpha) V_f(h_f)$ does not vary with $h$. It follows from Monotonicity that $V_f$ is monotonic on $\mathcal{H}_f$. In particular, for all $c, c' \in \mathcal{C}$, if $\{c\} \succeq \{c'\}$, then $V_f(c) \geq V_f(c')$. As $V_f$ is linear and non-constant, it is a positive linear transformation of $U$ on $\mathcal{C}$. As $u(c_0) = V_f(c_0) = 0$, there exists $\kappa_f$ as asserted in (ii).

Suppose that, contrary to (iii), $V_f(e(h)) > V_f(h)$ for some $h \in \mathcal{H}_f$. Then $W_f(e(h)) > W_f(h)$ and for sufficiently small $\alpha$, $W_f(e(h) - \alpha) > W_f(h + \alpha)$ and $V_f(e(h) - \alpha) > V_f(h + \alpha)$. The representation (A.2) implies that

$$\{h + \alpha, e(h) - \alpha\} \sim \{e(h) - \alpha\}.$$

Yet by Commitment Independence, $\{h+\alpha\} \succ \{e(h) - \alpha\}$, and thus $\{h+\alpha\} \succ \{h + \alpha, e(h) - \alpha\} \sim \{e(h) - \alpha\}$, contradicting Constants-Do-Not-Tempt.

For (iv), $V_f(f) = V_f(e(f)) \Longrightarrow V_f(f) = \frac{\kappa_f}{1-\kappa_f} U(e(f)) = \frac{\kappa_f}{1-\kappa_f} U(f)$. Then

$$V_f(\alpha c + (1-\alpha) f) = \tfrac{\kappa_f}{1-\kappa_f} U(\alpha c + (1-\alpha) f),$$

for all $c$, that is, $V_f(\cdot) = \frac{\kappa_f}{1-\kappa_f} U(\cdot)$ on $\mathcal{H}_f$, contradicting (A.3). □

**Lemma A.4.** *If $f, g \in \mathcal{H}$ are potentially tempting, then: (i) $\kappa_f = \kappa_g$; and* *(ii) there exist potentially tempting acts $f' \in \mathcal{H}_f$ and $g' \in \mathcal{H}_g$ such that $\{f'\} \cup A \sim \{g'\} \cup A$ for all $A \in \mathcal{A}$.*

*Proof.* Fix potentially tempting acts $f, g \in \mathcal{H}$. Take $c_f, c_g \in \mathcal{C}$ such that $u(c_f) = -U(f)$ and $u(c_g) = -U(g)$. Then $\left\{\frac{f+c_f}{2}\right\} \sim \left\{\frac{g+c_g}{2}\right\} \sim \{c_0\}$. By Lemma A.3, $V_f\left(\frac{f+c_f}{2}\right) = \frac{V_f(f)+V_f(c_f)}{2} > \frac{V_f(e(f))+V_f(c_f)}{2} = 0$ and analogously, $V_g\left(\frac{g+c_g}{2}\right) > 0$. Take a sufficiently small $\alpha > 0$ such that $V_f\left(\frac{f+c_f}{2} - \alpha\right) > 0$ and $V_g\left(\frac{g+c_g}{2} - \alpha\right) > 0$; let $f^* = \frac{f+c_f}{2} - \alpha$ and $g^* = \frac{g+c_g}{2} - \alpha$. Then $U(f^*) = U(g^*) = -\alpha < 0$, but $V_f(f^*) > 0$ and $V_g(g^*) > 0$. It follows that

$$\{c_0\} \succ \{c_0, f^*\} \succ \{f^*\} \sim \{g^*\} \quad \text{and} \quad \{c_0\} \succ \{c_0, g^*\} \succ \{g^*\} \sim \{f^*\}.$$

We claim there exist potentially tempting acts $f' \in \mathcal{H}_f$ and $g' \in \mathcal{H}_g$ such that

$$\{c_0\} \succ \{c_0, f'\} \sim \{c_0, f', g'\} \sim \{c_0, g'\} \succ \{f'\} \sim \{g'\}. \tag{A.4}$$

To see this, consider three possible cases:

1. $\{c_0, f^*\} \sim \{c_0, g^*\}$: Let $f' = f^*$ and $g' = g^*$.
2. $\{c_0, f^*\} \succ \{c_0, g^*\}$: Then $\mathcal{U}(\{c_0, g^*\}) < \mathcal{U}(\{c_0, f^*\}) < 0 = \mathcal{U}(\{c_0, e(g^*)\})$. By Continuity, there exists $\gamma \in (0,1)$ such that

   $$\{c_0, \gamma e(g^*) + (1-\gamma)g^*\} \sim \{c_0, f^*\}.$$

   Let $f' = f^*$ and $g' = \gamma e(g^*) + (1-\gamma) g^*$.
3. $\{c_0, g^*\} \succ \{c_0, f^*\}$: As in the previous case, there exists $\gamma \in (0,1)$ such that $\{c_0, \gamma e(f^*) + (1-\gamma) f^*\} \sim \{c_0, g^*\}$. Let $f' = \gamma e(f^*) + (1-\gamma) f^*$ and $g' = g^*$.

From (A.4), $\mathcal{U}(\{c_0, f'\}) = \mathcal{U}(\{c_0, g'\}) < 0$, and by (A.2), $\mathcal{U}(\{c_0, f'\}) = -V_f(f')$ and $\mathcal{U}(\{c_0, g'\}) = -V_g(g')$. Thus

$$V_f(f') = V_g(g') > 0. \tag{A.5}$$

Take $\alpha > 0$ such that $V_f(c_0+\alpha) < V_f(f')$ and $V_g(c_0+\alpha) < V_g(g')$. Since $\{c_0+\alpha\} \sim \{c_0+\alpha, c_0\}$, NC implies

$$\{c_0+\alpha, f'\} \sim \{c_0+\alpha, c_0, f'\} \quad \text{and} \quad \{c_0+\alpha, g'\} \sim \{c_0+\alpha, c_0, g'\}.$$

Thus (A.4) and NC imply that

$$\{c_0+\alpha, f'\} \sim \{c_0+\alpha, c_0, f'\} \sim \{c_0+\alpha, c_0, f', g'\}, \text{ and}$$
$$\{c_0+\alpha, g'\} \sim \{c_0+\alpha, c_0, g'\} \sim \{c_0+\alpha, c_0, f', g'\}.$$

It follows that $\mathcal{U}(\{c_0+\alpha, f'\}) = \mathcal{U}(\{c_0+\alpha, g'\})$ and hence, by Lemma A.3,

$$W_f(c_0+\alpha) - V_f(f') = W_g(c_0+\alpha) - V_g(g'),$$

$\frac{1}{1-\kappa_f} u(c_0+\alpha) = \frac{1}{1-\kappa_g} u(c_0+\alpha)$, and $\kappa_f = \kappa_g$.

Fix $\gamma \in (0,1]$. We claim that

$$\{f' + \gamma\} \sim \{f' + \gamma, g'\} \succ \{g'\} \text{ and } \{g' + \gamma\} \sim \{g' + \gamma, f'\} \succ \{f'\}. \tag{A.6}$$

By Lemma A.3, there exists $c_V \in \mathcal{C}$ such that $V_f(c_V) = V_f(f') > V_f(e(f'))$. Then $u(c_V) > u(e(f'))$ and

$$W_f(c_V) = u(c_V) + V_f(c_V) > u(e(f')) + V_f(f') = W_f(f') > W_f(e(f')).$$

By continuity of $W_f$, there exists $c^* \in \mathcal{C}$ such that

$$\begin{gathered} W_f(c_V) > W_f(c^*) > W_f(f') > W_f(e(f')), \\ W_f(f' + \gamma) > W_f(c^*) > W_f(f'), \\ U(f' + \gamma) \neq U(c^*). \end{gathered}$$

It follows that $V_f(f' + \gamma) \geq V_f(f') = V_f(c_V) > V_f(c^*)$. By (A.2),

$$\{f' + \gamma, c^*\} \sim \{f' + \gamma\} \not\sim \{c^*\},$$

but also $\{c^*\} \succ \{c^*, f'\} \succ \{f'\}$. Moreover, as $V_f(f') = V_g(g')$, $W_f(f') = W_g(g')$, and $\kappa_f = \kappa_g$, then by (A.2), $\mathcal{U}(\{c^*, f'\}) = \mathcal{U}(\{c^*, g'\})$, and hence,

$$\{c^*\} \succ \{c^*, f'\} \sim \{c^*, f', g'\} \sim \{c^*, g'\} \succ \{f'\} \sim \{g'\}.$$

By Lemma A.3, $\{f' + \gamma\} \sim \{f' + \gamma, c^*, f'\}$ because both menus belong to $\mathcal{A}_f$. By NC,

$$\{f' + \gamma, c^*, f'\} \sim \{f' + \gamma, c^*, f', g'\} \sim \{f' + \gamma, c^*, g'\} \sim \{f' + \gamma, g'\}.$$

Commitment Independence completes the proof of the first part of the claim (A.6). The second part is analogous.

By NC, for all $A \in \mathcal{A}$,

$$\begin{gathered} \{f' + \gamma\} \cup A \sim \{f' + \gamma, g'\} \cup A, \\ \{g' + \gamma\} \cup A \sim \{g' + \gamma, f'\} \cup A, \end{gathered}$$

and as $\gamma \to 0$, $\{f'\} \cup A \sim \{f', g'\} \cup A \sim \{g'\} \cup A$ by Continuity. $\square$

By the lemma, $\kappa_f = \kappa \in (0,1)$ for all potentially tempting acts $f \in \mathcal{H}$. For every $f \in \mathcal{H}$, define $W(f) = U(f) + V(f)$, where

- $V(f) = V_f(f)$ if $f$ is potentially tempting,
- $V(f) = \frac{\kappa}{1-\kappa} U(f)$ if $f$ is never tempting.

The definition of $V$ and Lemma A.3 imply that for all $f \in \mathcal{H}$,

$$V(c_+) \geq V(f) \geq V(e(f)) \geq V(c_-). \tag{A.7}$$

Moreover, for all $\alpha \in [0,1]$ and $c \in \mathcal{C}$,

$$V(\alpha f + (1-\alpha)c) = \alpha V(f) + (1-\alpha)V(c), \tag{A.8}$$
$$W(\alpha f + (1-\alpha)c) = \alpha W(f) + (1-\alpha)W(c). \tag{A.9}$$

If $f$ is never tempting or $\alpha = 0$, then these equalities follow from the linearity of $U$. If $f$ is potentially tempting and $\alpha > 0$, then the act $f_\alpha = \alpha f + (1-\alpha)c$ is also potentially tempting: by Collinear Independence, $\{c'\} \succ \{c', f\} \Longrightarrow \{\alpha c' + (1-\alpha)\, c\} \succ \{\alpha c' + (1-\alpha)\, c, \alpha f + (1-\alpha)\, c'\}$. Therefore, $V_{f_\alpha}(\cdot) = V_f(\cdot)$ on $\mathcal{H}_{f_\alpha}$ by the uniqueness statement in Lemma A.3. Thus,

$$V(f_\alpha) = V_{f_\alpha}(f_\alpha) = V_f(f_\alpha) = \alpha V_f(f) + (1-\alpha)V_f(c) = \alpha V(f) + (1-\alpha)V(c).$$

For every menu $A \in \mathcal{A}$, define

$$\mathcal{U}_{WV}(A) = \max_{f \in A} W(f) - \max_{f \in A} V(f).$$

Later we show that both $W$ and $V$ are continuous and hence, the maxima in the above definition are obtained even if $A$ is not finite.

**Lemma A.5.** *For all binary menus $A \in \mathcal{A}$, $\alpha \in [0,1]$ and $c \in \mathcal{C}$,*

*(i)* $\mathcal{U}(\alpha A + (1-\alpha)\{c\}) = \alpha \mathcal{U}(A) + (1-\alpha)u(c)$;

*(ii)* $\mathcal{U}_{WV}(\alpha A + (1-\alpha)\{c\}) = \alpha \mathcal{U}_{WV}(A) + (1-\alpha)u(c)$;

*(iii)* $\mathcal{U}(A) = \mathcal{U}_{WV}(A)$.

*Proof.* (i) Let $A = \{f, g\}$, $\alpha \in [0,1]$ and $c \in \mathcal{C}$. Because $\{c_+\} \succeq A \succeq \{c_-\}$, there exists $e(A) \in \mathcal{C}$ such that $A \sim \{e(A)\}$. By Collinear Independence, $\alpha A + (1-\alpha)\{c\} \sim \alpha\{e(A)\} + (1-\alpha)\{c\}$. Therefore,

$$\mathcal{U}(\alpha A + (1-\alpha)\{c\}) = \mathcal{U}(\alpha\{e(A)\} + (1-\alpha)\{c\}) = \\ \alpha u(e(A)) + (1-\alpha)u(c) = \alpha \mathcal{U}(A) + (1-\alpha)u(c).$$

(ii) follows from the equalities (A.8) and (A.9).
(iii) Let $A = \{f, g\}$. If both $f$ and $g$ are never tempting, then $\mathcal{U}(A) = \mathcal{U}(\{e(f), e(g)\})$ by Lemma A.2, and

$$\mathcal{U}(\{e(f), e(g)\}) = \max\{u(e(f)), u(e(g)\} = \mathcal{U}_{WV}(A).$$

If $f$ is potentially tempting and $g$ is never tempting, then $\mathcal{U}(A) = \mathcal{U}(\{f, e(g)\})$ by Lemma A.2, and $\mathcal{U}(\{f, e(g)\}) = \mathcal{U}_{WV}(A)$ by Lemma A.3. If both $f$ and $g$ are potentially tempting, then by Lemma A.4, there exist potentially tempting $f' \in \mathcal{H}_f$ and $g' \in \mathcal{H}_g$ such that $\{f'\} \cup A \sim \{g'\} \cup A$ for all $A \in \mathcal{A}$. Take $\gamma \in (0,1]$ and $c' \in \mathcal{C}$ such that $\gamma A + (1-\gamma)\{c'\} = \{\gamma f + (1-\gamma)c', g'\}$. Then

$$\mathcal{U}(\gamma A + (1-\gamma)\{c'\}) = \mathcal{U}(\{\gamma f + (1-\gamma)c', g'\}) = \mathcal{U}(\{\gamma f + (1-\gamma)c', f'\}) =_1 \\ \mathcal{U}_{WV}(\{\gamma f + (1-\gamma)c', f'\}) =_2 \mathcal{U}_{WV}(\{\gamma f + (1-\gamma)c', g'\}) = \mathcal{U}_{WV}(\gamma A + (1-\gamma)\{c'\}).$$

Here the equality $=_1$ follows from Lemma A.3, and $=_2$ follows from the fact that $W(f') = W(g')$ and $V(f') = V(g')$. By (i) and (ii), $\mathcal{U}(A) = \mathcal{U}_{WV}(A)$. $\square$

**Lemma A.6.** *There exists a convex and closed set $Q$ of probability measures on $S$ such that for all $f \in \mathcal{H}$,*

$$V(f) = \tfrac{\kappa}{1-\kappa} \max_{q \in Q} q \cdot u(f). \tag{A.10}$$

*Moreover, $Q$ is unique and $p \in Q$.*

*Proof.* First show that $V$ is monotonic, continuous, and quasi-convex.
*Monotonicity:* Suppose that $f \in \mathcal{H}$ dominates $f' \in \mathcal{H}$, but $V(f') > V(f)$. Take $c \in \mathcal{C}$ such that $V(f') > V(c) > V(f) \geq V(e(f))$. Then $u(c) > u(e(f)) = U(f) \geq U(f')$. By Lemma A.5 and Monotonicity, $\{c\} \sim \{c, f\} \succ \{f\} \succeq \{f'\}$ and $\{c\} \succ \{c, f'\} \succeq \{f'\}$. For every $\alpha \in [0,1]$, define $f_\alpha = \alpha f + (1-\alpha) f'$ and $\phi(\alpha) = \mathcal{U}(\{c, f_\alpha\})$. Then $\phi(1) = u(c) > \phi(0)$. As $\phi$ is continuous, there exist $\alpha < \beta < 1$ such that

$$\phi(1) = u(c) > \phi(\alpha) > \phi(\beta) > U(f) \geq U(f_\alpha) \geq U(f_\beta).$$

However, both pairs $(c, f_\alpha)$ and $(c, f_\beta)$ are interior, $f_\beta$ dominates $f_\alpha$, and by Monotonicity, $\phi(\alpha) \leq \phi(\beta)$. This is a contradiction.
*Continuity.* Let a sequence of acts $f_n$ converge to $f$ as $n \to \infty$. There exist sequences $\alpha_n$ and $\beta_n$ both converging to zero such that $f + \alpha_n$ dominates $f_n$, and $f_n$ dominates $f - \beta_n$. As $V$ is monotonic,

$$\alpha_n V(c_+) + (1-\alpha_n) V(f) \geq V(f_n) \geq \beta_n V(c_-) + (1-\beta_n) V(f).$$

In the limit, these inequalities imply $V(f) = \lim_{n\to\infty} V(f_n)$.
*Quasi-Convexity.* Suppose that there exist $f, g \in \mathcal{H}$ and $\alpha \in (0,1)$ such that $V(\alpha f + (1-\alpha) g) > V(f) = V(g)$. By continuity of $V$ there exist $\gamma_f, \gamma_g \in (0,1)$ such that $1 > \gamma_f > \alpha > \gamma_g > 0$ and

$$V(\alpha f + (1-\alpha) g) > V(\gamma_f f + (1-\gamma_f) g) = V(\gamma_g f + (1-\gamma_g) g) > V(f) = V(g).$$

Let $f^* = \gamma_f f + (1-\gamma_f) g$, $g^* = \gamma_g f + (1-\gamma_g) g$, and $\alpha^* \in (0,1)$ such that

$$\alpha f + (1-\alpha) g = \alpha^* f^* + (1-\alpha^*) g^*.$$

Take $c \in \mathcal{C}$ such that $V(c) = V(f^*) = V(g^*)$. As $V(c) > V(f) \geq V(e(f))$ and $V(c) > V(g) \geq V(e(g))$, then $u(c) > u(e(f))$ and $u(c) > u(e(g))$, and hence, $W(c) > W(f^*)$ and $W(c) > W(g^*)$. Take $\beta > 0$ sufficiently small such that $W(c-\beta) > W(f^*)$ and $W(c-\beta) > W(g^*)$. By Lemma A.5, both pairs $(c-\beta, f^*)$ and $(c-\beta, g^*)$ are interior, and $\{c-\beta, f^*\} \sim \{c-\beta, g^*\}$. However,

$$\{c-\beta, \alpha^* f^* + (1-\alpha^*) g^*\} \prec \{c-\beta, f^*\},$$

because $V(\alpha^* f^* + (1-\alpha^*) g^*) > V(f^*)$. This is a contradiction with Convex Temptation.

The ranking that the function $V$ represents on $\mathcal{H}$ satisfies all the axioms of the maxmax model — these are the axioms of Gilboa and Schmeidler's multiple-priors model [14], with the exception that their axiom "Uncertainty Aversion", which is convexity of weakly better-than sets, is replaced by convexity of weakly worse-than sets. It follows from [14] that $V$ has the form in (A.10), and that $Q$ is unique. The inclusion $p \in Q$ follows from the fact that for all $f \in \mathcal{H}$, $V(f) \geq V(e(f)) = p \cdot u(f)$. This completes the proof of the lemma. □

**Lemma A.7.** *For all menus $A \in \mathcal{A}$, $\mathcal{U}(A) = \mathcal{U}_{WV}(A)$.*

*Proof.* Fix an arbitrary finite menu $A$. Take $g_A \in \arg\max_{f \in A} W(f)$ and $h_A \in \arg\max_{f \in A} V(f)$. Then for all $f \in A$,

$$\mathcal{U}_{WV}(\{g_A, f\}) \geq \mathcal{U}_{WV}(\{g_A, h_A\}) \geq \mathcal{U}_{WV}(\{f, h_A\}),$$

and by Lemma A.5, $\{g_A, f\} \succeq \{g_A, h_A\} \succeq \{f, h_A\}$. From Set-Betweenness, it follows by induction with respect to the size of the set $A$ that

$$A = \bigcup_{f \in A} \{g_A, f\} \succeq \{g_A, h_A\} \succeq \bigcup_{f \in A} \{f, h_A\} = A,$$

that is, $A \sim \{g_A, h_A\}$. Therefore,

$$\mathcal{U}(A) = \mathcal{U}(\{g_A, h_A\}) = \mathcal{U}_{WV}(\{g_A, h_A\}) = \mathcal{U}_{WV}(A).$$

Finally, as both $\mathcal{U}$ and $\mathcal{U}_{WV}$ are continuous, $\mathcal{U}(\cdot) = \mathcal{U}_{WV}(\cdot)$ on all of $\mathcal{A}$. $\square$

To show the required uniqueness of $(u, p, \kappa, Q)$ in representation (2.1)-(2.3), suppose that this tuple can be replaced by $(u', p', \kappa', Q')$. Then $u'$ is a positive linear transformation of $u$, and hence, $(u, p, \kappa, Q)$ can be replaced by $(u, p', \kappa', Q')$ as well. The uniqueness statements in Lemmas A.1, A.3 and A.6 imply that if $\succeq$ is not strategically rational, then $p = p'$, $\kappa = \kappa'$, and $Q = Q'$.

# B. Appendix: Proofs for Comparative Dissonance

*Proof of Theorem 4.2*: Let $\succeq^*$ and $\succeq$ conform to our model with corresponding tuples $(u^*, p^*, Q^*, \kappa^*)$ and $(u, p, Q, \kappa)$. Suppose that neither preference is strategically rational. Then $k, k^* > 0$ and sufficiency of (4.2) and (4.3) is immediate:

$$\begin{aligned} \{f\} \succ \{f, g\} \quad &\Rightarrow \quad [p \cdot u(f) > p \cdot u(g) \wedge Q \cdot u(g) > Q \cdot u(f)] \quad \Rightarrow \\ &[p^* \cdot u^*(f) > p^* \cdot u^*(g) \wedge Q^* \cdot u^*(g) > Q^* \cdot u^*(f)] \quad \Rightarrow \quad \{f\} \succ^* \{f, g\}. \end{aligned}$$

For necessity, let $\succeq^*$ have greater dissonance than $\succeq$. For all vectors $a \in \mathbb{R}^S$, let

$$Q \cdot a = \max_{q \in Q} q \cdot a \quad \text{and} \quad Q^* \cdot a = \max_{q \in Q^*} q \cdot a. \tag{B.1}$$

**Lemma B.1.** *(i) $u$ and $u^*$ are cardinally equivalent.*

*(ii) For all $a, b \in \mathbb{R}^S$,*

$$p \cdot a > p \cdot b \text{ and } Q \cdot b > Q \cdot a \quad \Rightarrow \quad p^* \cdot a > p^* \cdot b \text{ and } Q^* \cdot b > Q^* \cdot a. \tag{B.2}$$

*(iii) $p = p^*$.*

*Proof.* First, show that for all $c, c' \in \mathcal{C}$,

$$u(c) = u(c') \quad \Rightarrow \quad u^*(c) = u^*(c'). \tag{B.3}$$

Suppose to the contrary that $u(c) = u(c')$ and $u^*(c) > u^*(c')$ for some $c, c' \in \mathcal{C}$. Take $f, g \in \mathcal{H}$ such that $\{f\} \succ \{f, g\}$. The equality $u(c) = u(c')$ implies

$$\{\alpha f + (1-\alpha)c\} \succ \{\alpha f + (1-\alpha)c, \alpha g + (1-\alpha)c'\}.$$

Because $\succeq^*$ has greater dissonance, $\{f\} \succ^* \{f, g\}$. Therefore, the inequality $u^*(c) > u^*(c')$ implies that for sufficiently small $\alpha > 0$,

$$\{\alpha f + (1-\alpha)c\} \sim^* \{\alpha f + (1-\alpha)c, \alpha g + (1-\alpha)c'\}.$$

But this contradicts the hypothesis that $\succeq^*$ has greater dissonance than $\succeq$.

Take $c_+, c_- \in \mathcal{C}$ such that $u(c_+) > u(c_-)$ and $u(c_+) \geq u(c) \geq u(c_-)$ for all $c \in \mathcal{C}$. Then for all $c \in \mathcal{C}$,

$$\begin{aligned} c &\sim \frac{u(c) - u(c_-)}{u(c_+) - u(c_-)} c_+ + \frac{u(c_+) - u(c)}{u(c_+) - u(c_-)} c_-, \qquad \text{and by (B.3),} \\ u^*(c) &= \frac{u^*(c_+) - u^*(c_-)}{u(c_+) - u(c_-)} u(c) + \frac{u^*(c_-)u(c_+) - u^*(c_+)u(c_-)}{u(c_+) - u(c_-)}. \end{aligned}$$

Note that $u^*(c_+) \neq u^*(c_-)$ because $\succeq^*$ is not strategically rational and hence non-degenerate. Thus, either $u^*$ is a positive linear transformation of $u$, or $u^*$ is a negative linear transformation of $u$. Next, we show that the former case implies statements (ii) and (iii), and that the latter case is impossible.

*Case 1.* $u^*$ is a positive linear transformation of $u$. Wlog assume that $u = u^*$ and $u(\mathcal{C}) = u^*(\mathcal{C}) = [-1, 1]$. Fix any $a, b \in \mathbb{R}^S$ such that $p \cdot a > p \cdot b$ and $Q \cdot b > Q \cdot a$. Take $\alpha > 0$ such that $\mid \alpha a(s) \mid, \ \mid \alpha b(s) \mid \ \leq 1$ for all $s \in S$. Then $\alpha a = u(f)$ and $\alpha b = u(g)$ for some $f, g \in \mathcal{H}$. (Here $u(f)$ and $u(g)$ are vectors in $\mathbb{R}^S$.) Then

$$\begin{gathered} p \cdot a > p \cdot b \text{ and } Q \cdot b > Q \cdot a \quad \Rightarrow \quad p \cdot u(f) > p \cdot u(g) \text{ and } Q \cdot u(g) > Q \cdot u(f) \quad \Rightarrow \\ \{f\} \succ \{f, g\} \quad \Rightarrow \quad \{f\} \succ^* \{f, g\} \quad \Rightarrow \\ p^* \cdot u(f) > p^* \cdot u(g) \text{ and } Q^* \cdot u(g) > Q^* \cdot u(f) \quad \Rightarrow \quad p^* \cdot a > p^* \cdot b \text{ and } Q^* \cdot b > Q^* \cdot a, \end{gathered}$$

which proves (ii).

To show (iii), suppose that $p \neq p^*$. Let

$$R = \{q \in \mathbb{R}^S : q = p + \alpha(p - p^*) \text{ for } \alpha \geq 0\} = \{q \in \mathbb{R}^S : p \in [q, p^*]\}.$$

Consider two subcases.

(1) $Q \not\subset R$: Let $p' \in Q \backslash R$. Take a hyperplane $b \in \mathbb{R}^S$ that separates the singleton $p$ and the segment $[p', p^*]$:

$$p \cdot b < 0, \quad p' \cdot b > 0, \quad p^* \cdot b > 0.$$

These inequalities violate (B.2) for $a = 0$.

(2) $Q \subset R$: Then $Q$ is a segment with end points $p$ and $p' = p + \alpha(p - p^*)$ for some $\alpha > 0$. Note that $p$ is an interior point of the segment $[p^*, p']$. Take a hyperplane $a \in \mathbb{R}^S$ that separates $p^*$ and $p'$ and passes through $p$:

$$p^* \cdot a > 0, \quad p \cdot a = 0, \quad p' \cdot a < 0.$$

Take a hyperplane $b \in \mathbb{R}^S$ that separates $p'$ and the segment $[p, p^*]$:

$$p' \cdot b > 0, \quad p \cdot b < 0, \quad p^* \cdot b < 0.$$

Wlog $p^* \cdot a > Q^* \cdot b$ (multiply $a$ by a positive scalar if needed). Thus $p \cdot a > p \cdot b$,

$$Q \cdot b \geq p' \cdot b > 0 = \max\{p \cdot a, p' \cdot a\} = Q \cdot a,$$

but $Q^* \cdot a \geq p^* \cdot a > Q^* \cdot b$. This contradicts (B.2).

*Case 2.* $u^*$ is a negative linear transformation of $u$. We show this is impossible.

Wlog assume that $u^* = -u$ and $u(\mathcal{C}) = u^*(\mathcal{C}) = [-1, 1]$. Then, paralleling (B.2) in the previous case,

$$p \cdot a > p \cdot b \text{ and } Q \cdot b > Q \cdot a \quad \Rightarrow \quad p^* \cdot (-a) > p^* \cdot (-b) \text{ and } Q^* \cdot (-b) > Q^* \cdot (-a). \tag{B.4}$$

for all $a, b \in \mathbb{R}^S$. It follows that for some $a, b \in \mathbb{R}^S$, $p \cdot a > p \cdot b$ but $p^* \cdot a < p^* \cdot b$. Thus $p \neq p^*$. Consider two subcases.

(1) $Q \not\subset [p, p^*]$: Let $p' \in Q \backslash [p, p^*]$. Take a hyperplane $b \in \mathbb{R}^S$ that separates $p'$ and $[p, p^*]$:

$$p' \cdot b > 0, \quad p \cdot b < 0, \quad p^* \cdot b < 0.$$

This contradicts (B.4) for $a = 0$.

(2) $Q \subset [p, p^*]$: Then $Q$ is a segment with end points $p$ and $p' = \alpha p^* + (1 - \alpha)p$ for some $\alpha > 0$. Take a hyperplane $a \in \mathbb{R}^S$ that separates $p$ and $[p', p^*]$:

$$p \cdot a = 0, \quad p' \cdot a < 0, \quad p^* \cdot a < 0.$$

Take another hyperplane $b \in \mathbb{R}^S$ that separates $p$ and $[p', p^*]$:

$$p \cdot b < 0, \quad p' \cdot b > 0, \quad p^* \cdot b > 0.$$

Wlog $p^* \cdot (-a) > Q^* \cdot (-b)$ (multiply $a$ by a positive scalar if needed). Then $p \cdot a = 0 > p \cdot b$,

$$Q \cdot b \geq p' \cdot b > 0 = \max\{p' \cdot a, p \cdot a\} = Q \cdot a,$$

but $Q^* \cdot (-a) \geq p^* \cdot (-a) > Q^* \cdot (-b)$. This contradicts (B.4). $\square$

The following method of proof is analogous to the one used by Kopylov [20]. Let $\mathbb{D}$ be the set of all points $a \in \mathbb{R}^S$ at which the convex functions $Q \cdot a$ and $Q^* \cdot a$ are both differentiable. By [28, Theorem 25.5], the complement of the set $\mathbb{D}$ has measure zero. Thus $\mathbb{D}$ is dense. For every $a \in \mathbb{D}$, let

$$q(a) = \nabla(Q \cdot a) \quad \text{and} \quad q^*(a) = \nabla(Q^* \cdot a)$$

be the derivatives of $Q \cdot a$ and $Q^* \cdot a$ respectively. Let $\vec{1} = (1, \ldots, 1) \in \mathbb{R}^S$.

**Lemma B.2.** *The functions $q(\cdot), q^*(\cdot) : \mathbb{D} \to \mathbb{R}^S$ have the following properties:*

*(i) For all $a \in \mathbb{D}$ and $q \in Q$, $q = q(a)$ iff $Q \cdot a = q \cdot a$.*

*(ii) For all $a \in \mathbb{D}$ and $q \in Q^*$, $q = q^*(a)$ iff $Q^* \cdot a = q \cdot a$.*

*(iii) If $a \in \mathbb{D}$, $\alpha > 0$ and $\gamma \in \mathbb{R}$, then*

$$\alpha a + \gamma \vec{1} \in \mathbb{D}, \quad q(\alpha a + \gamma \vec{1}) = q(a), \quad q^*(\alpha a + \gamma \vec{1}) = q^*(a).$$

*(iv) For any $a \in \mathbb{D}$, there exists $\epsilon_a \in [0,1]$ such that $q(a) = \epsilon_a q^*(a) + (1 - \epsilon_a)p$.*

*(v) There exists $\epsilon \in [0,1]$ such that $q(a) = \epsilon q^*(a) + (1 - \epsilon)p$ for all $a \in \mathbb{D}$.*

*Proof.*

(i) Fix $a \in \mathbb{D}$ and $q \in Q$ such that $Q \cdot a = q \cdot a$. For all $b \in \mathbb{R}^S$ and $\delta \in \mathbb{R}$,

$$Q \cdot a + \delta(q \cdot b) = q \cdot (a + \delta b) \leq Q \cdot (a + \delta b) = Q \cdot a + \delta(q(a) \cdot b) + o(\delta).$$

Then $q \cdot b = q(a) \cdot b$ for all $b \in \mathbb{R}^S$, that is, $q = q(a)$. Similarly for (ii).

(iii) Fix $a \in \mathbb{D}$, $\alpha > 0$ and $\gamma \in \mathbb{R}$. Then $\alpha a + \gamma \vec{1} \in \mathbb{D}$ because the superposition $Q \cdot b = \alpha Q \cdot \left(\frac{b - \gamma \vec{1}}{\alpha}\right) + \gamma$ is differentiable at $\alpha a + \gamma \vec{1}$. By (i), $q(\alpha a + \gamma \vec{1}) = q(a)$ because $Q \cdot (\alpha a + \gamma \vec{1}) = \alpha(Q \cdot a) + \gamma = q(a) \cdot (\alpha a + \gamma \vec{1})$. Similarly for $Q^*$ and $q^*(\cdot)$.

(iv) Suppose that for some $a$ no such $\epsilon_a$ exists. Let $b$ separate $q(a)$ from the segment $[q^*(a), p]$, so that $q^*(a) \cdot b < 0$, $p \cdot b < 0$, but $q(a) \cdot b > 0$. Then for sufficiently small $\delta > 0$, $Q^* \cdot (a + \delta b) = Q^* \cdot a + \delta(q^*(a) \cdot b) + o(\delta) < Q^* \cdot a$, but also

$$p \cdot a > p \cdot (a + \delta b) \quad \text{and} \quad Q \cdot (a + \delta b) \geq q(a) \cdot (a + \delta b) > q(a) \cdot a = Q \cdot a.$$

By (B.2), $Q^* \cdot (a + \delta b) > Q^* \cdot a$, a contradiction.

(v) Let $a, b \in \mathbb{D}$ be such that $q^*(a) \neq p$ and $q^*(b) \neq p$, and prove $\epsilon_a = \epsilon_b$. (Note that if $q^*(a) \neq p$, then $\epsilon_a$ is unique, and if $q^*(a) = p$, then $\epsilon_a \in [0,1]$ is arbitrary. ) As $q^*(a) \neq p$ and $p = p^* \in Q^*$, then by (iii), $Q^* \cdot a > p \cdot a$. Similarly, $Q^* \cdot b > p \cdot b$. Let

$$a' = \frac{a - (p \cdot a)\vec{1}}{Q^* \cdot a - p \cdot a} \quad \text{and} \quad b' = \frac{b - (p \cdot b)\vec{1}}{Q^* \cdot b - p \cdot b}.$$

By (iii) and (iv), $a', b' \in \mathbb{D}$, $q^*(a') = q^*(a)$, $q^*(b') = q^*(b)$, and

$$q(a') = q(a) = \epsilon_a q^*(a) + (1 - \epsilon_a)p \quad \text{and} \quad q(b') = q(b) = \epsilon_b q^*(b) + (1 - \epsilon_b)p.$$

By construction, $p \cdot a' = p \cdot b' = 0$, $Q^* \cdot a' = Q^* \cdot b' = 1$, $Q \cdot a' = \epsilon_a$, and $Q \cdot b' = \epsilon_b$. Suppose that $\epsilon_a \neq \epsilon_b$; wlog let $\epsilon_a < \epsilon_b$. Then for sufficiently small $\gamma > 0$,

$$p \cdot (a' + \gamma \vec{1}) = \gamma > p \cdot b', \quad Q \cdot (a' + \gamma \vec{1}) = \epsilon_a + \gamma < \epsilon_b = Q \cdot b',$$

but $Q^* \cdot (a' + \gamma \vec{1}) = 1 + \gamma > Q^* \cdot b'$. This contradicts (B.2). Thus $\epsilon_a = \epsilon_b$. □

Conclude that $Q \cdot a = \epsilon(Q^* \cdot a) + (1-\epsilon)(p \cdot a)$ for all $a \in \mathbb{D}$ and hence, by continuity, for all $a \in \mathbb{R}^S$. It follows that $Q = \epsilon Q^* + (1-\epsilon)p$; $\epsilon > 0$ because $\succeq$ is not strategically rational. This completes the proof of Theorem 4.2.

*Proof of Theorem 4.3*: Let $\succeq^*$ and $\succeq$ conform to our model with corresponding tuples $(u^*, p^*, Q^*, \kappa^*)$ and $(u, p, Q, \kappa)$. Suppose that neither preference is strategically rational.

Let $P = (1-\kappa)\{p\} + \kappa Q$ and $P^* = (1-\kappa^*)\{p^*\} + \kappa^* Q^*$. The conditions (4.2), (4.3), and $\kappa^* \geq \epsilon\kappa$ imply

$$P = \left(1 - \frac{\epsilon\kappa}{\kappa^*}\right)\{p\} + \frac{\epsilon\kappa}{\kappa^*}P^*.$$

Sufficiency of these conditions now follows from:

$$\begin{aligned} \{f\} \succ \{f,g\} \sim \{g\} \quad &\Rightarrow \quad [p \cdot u(f) > p \cdot u(g) \wedge P \cdot u(g) > P \cdot u(f)] \quad \Rightarrow \\ &[p^* \cdot u^*(f) > p^* \cdot u^*(g) \wedge P^* \cdot u^*(g) > P^* \cdot u^*(f)] \quad \Rightarrow \quad \{f\} \succ^* \{f,g\} \sim^* \{g\}. \end{aligned}$$

For necessity, let $\succeq^*$ be more self-justifying than $\succeq$. Then $\succeq^*$ has more dissonance than $\succeq$, and Theorem 4.3 implies (4.2) and (4.3). Moreover, for all $a, b \in \mathbb{R}^S$,

$$p \cdot a > p \cdot b \text{ and } P \cdot b > P \cdot a \quad \Rightarrow \quad p^* \cdot a > p^* \cdot b \text{ and } P^* \cdot b > P^* \cdot a. \tag{B.5}$$

To prove this claim, fix any $a, b \in \mathbb{R}^S$. Take $\alpha > 0$ and $f, g \in \mathcal{H}$ such that $\alpha a = u(f)$ and $\alpha b = u(g)$. Then

$$\begin{gathered} p \cdot a > p \cdot b \text{ and } P \cdot b > P \cdot a \quad \Rightarrow \quad p \cdot u(f) > p \cdot u(g) \text{ and } P \cdot u(g) > P \cdot u(f) \quad \Rightarrow \\ \{f\} \succ \{f,g\} \sim \{g\} \quad \Rightarrow \quad \{f\} \succ^* \{f,g\} \sim^* \{g\} \quad \Rightarrow \\ p^* \cdot u(f) > p^* \cdot u(g) \text{ and } P^* \cdot u(g) > P^* \cdot u(f) \quad \Rightarrow \quad p^* \cdot a > p^* \cdot b \text{ and } P^* \cdot b > P^* \cdot a. \end{gathered}$$

Use the condition (B.5) to replace $Q$ and $Q^*$ by $P$ and $P^*$ in Lemma B.2 and obtain $0 < \theta \leq 1$ such that $P = (1-\theta)\{p\} + \theta P^*$. In particular, $P \subset P^*$ and therefore also

$$(1-\kappa\epsilon)\{p\} + \kappa\epsilon Q^* \subset (1-\kappa^*)\{p\} + \kappa^* Q^*.$$

As $Q^*$ is a nonsingleton, $\kappa\epsilon \leq \kappa^*$. □

# C. Appendix: Examples

We provide some details for Example 1.

Set-Betweenness: suppose that $\mathcal{U}(A) \leq \mathcal{U}(B)$. Then

$$\mathcal{U}(A \cup B) = \frac{\max_{h \in A \cup B}[U(h)V(h)]}{\max_{h' \in A \cup B} V(h')}$$

$$= \frac{\max\{\max_{h \in A}[U(h)V(h)],\ \max_{h \in B}[U(h)V(h)]\}}{\max\{\max_{h \in A} V(h),\ \max_{h \in B} V(h)\}}$$

$$\leq \frac{\max\left\{\frac{\max_{h\in A} V(h)}{\max_{h\in B} V(h)}\max_{h\in B}\left[U\left(h\right)V\left(h\right)\right],\ \ \max_{h\in B}\left[U\left(h\right)V\left(h\right)\right]\right\}}{\max\left\{\max_{h\in A} V\left(h\right),\ \ \max_{h\in B} V\left(h\right)\right\}}.$$

If $\frac{\max_{h\in A} V(h)}{\max_{h\in B} V(h)} \leq 1$, then above equals $\frac{\max_{h\in B}[U(h)V(h)]}{\max_{h\in B} V(h)} = \mathcal{U}\left(B\right)$. If $\frac{\max_{h\in A} V(h)}{\max_{h\in B} V(h)} \geq 1$, then above equals

$$\frac{\frac{\max_{h\in A} V(h)}{\max_{h\in B} V(h)}\max_{h\in B}\left[U\left(h\right)V\left(h\right)\right]}{\max_{h\in A} V\left(h\right)} = \mathcal{U}\left(B\right).$$

Thus $\mathcal{U}\left(A\cup B\right) \leq \mathcal{U}\left(B\right)$.

The verification of $\mathcal{U}\left(A\right) \leq \mathcal{U}\left(A\cup B\right)$ is symmetric:

$$\mathcal{U}\left(A\cup B\right) = \frac{\max\left\{\max_{h\in A}\left[U\left(h\right)V\left(h\right)\right],\ \ \max_{h\in B}\left[U\left(h\right)V\left(h\right)\right]\right\}}{\max\left\{\max_{h\in A} V\left(h\right),\ \ \max_{h\in B} V\left(h\right)\right\}}$$

$$\geq \frac{\max\left\{\max_{h\in A}\left[U\left(h\right)V\left(h\right)\right],\ \frac{\max_{h\in B} V(h)}{\max_{h\in A} V(h)}\max_{h\in A}\left[U\left(h\right)V\left(h\right)\right]\right\}}{\max\left\{\max_{h\in A} V\left(h\right),\ \ \max_{h\in B} V\left(h\right)\right\}}$$

If $\frac{\max_{h\in B} V(h)}{\max_{h\in A} V(h)} \leq 1$, then above equals $\frac{\max_{h\in A}[U(h)V(h)]}{\max_{h\in A} V(h)} = \mathcal{U}\left(A\right)$. If $\frac{\max_{h\in B} V(h)}{\max_{h\in A} V(h)} \geq 1$, then above equals

$$\frac{\frac{\max_{h\in B} V(h)}{\max_{h\in A} V(h)}\max_{h\in A}\left[U\left(h\right)V\left(h\right)\right]}{\max_{h\in B} V\left(h\right)} = \mathcal{U}\left(A\right).$$

Thus $\mathcal{U}\left(A\cup B\right) \geq \mathcal{U}\left(A\right)$.

Calculate that

$$\{g\} \succ \{g,h\} \Longleftrightarrow [U\left(g\right) > U\left(h\right) \wedge V\left(g\right) < V\left(h\right)], \text{ and}$$

$$\{g\} \succ \{g,h\} \sim \{h\} \Longleftrightarrow [U\left(g\right) > U\left(h\right) \wedge UV\left(g\right) < UV\left(h\right)].$$

Thus

$$\{g\} \succ \{g,h\} \Longrightarrow \mathcal{U}\left(\{g,h\}\right) = \max\left\{\frac{U\left(g\right)V\left(g\right)}{V\left(h\right)}, U\left(h\right)\right\} \text{ and}$$

$$\{g\} \succ \{g,h\} \succ \{h\} \Longrightarrow \mathcal{U}\left(\{g,h\}\right) = \frac{U\left(g\right)V\left(g\right)}{V\left(h\right)}.$$

Constants-Do-Not-Tempt: $U\left(f\right) > U\left(c\right) \Longrightarrow V\left(f\right) \geq U\left(f\right) > U\left(c\right) = V\left(c\right)$.

Monotonicity: (i) obvious. (ii) If $(c,f)$ and $(c,g)$ are interior, then

$$\mathcal{U}\left(\{c,f\}\right) = \frac{U\left(c\right)V\left(c\right)}{V\left(f\right)} \leq \frac{U\left(c\right)V\left(c\right)}{V\left(g\right)} = \mathcal{U}\left(\{c,g\}\right)$$

if $f$ dominates $g$.

Convex Temptation: Given interior pairs,

$$\mathcal{U}(\{c,f\}) = \mathcal{U}(\{c,g\}) \Longrightarrow \frac{U(c)\,V(c)}{V(f)} = \frac{U(c)\,V(c)}{V(g)} \Longrightarrow V(f) = V(g)\,.$$

If also $(c, \alpha f + (1-\alpha)\,g)$ is interior, then

$$\mathcal{U}(\{c, \alpha f + (1-\alpha)\,g\}) = \frac{U(c)\,V(c)}{V(\alpha f + (1-\alpha)\,g)}.$$

Therefore, $\{c, \alpha f + (1-\alpha)\,g\} \succeq \{c,f\}$ iff

$$V(\alpha f + (1-\alpha)\,g) \leq V(f) = V(g)\,,$$

that is, if $V$ is quasiconvex.

NC: $\{f'\} \succ \{f,f'\} \sim \{f\} \Longrightarrow [U(f') > U(f) \wedge UV(f') < UV(f)]$.
But choice out of $\{f,f',g\}$ solves $\max_{h=f,f',g} U(h)\,V(h)$ - follows that $f'$ could not be chosen. Thus $\{f,f'\} \sim \{f,f',g\}$ as argued in the text.

Similarly if $\{f\} \sim \{f,f'\} \succ \{f'\}$. The bottom line is that NC is satisfied because the temptation ranking is represented by $V$ and choice out of menus is rationalizable by $UV$.

# References

[1] G. Akerlof and W.T. Dickens, The economic consequences of cognitive dissonance, *Amer. Econ. Rev.* 72 (1982), 307-319.

[2] C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis*, Springer, 1994.

[3] E. Aronson, *The Social Animal*, 7th edition, W.H. Freeman, New York, 1996.

[4] R. Baumeister, The optimal margin of illusion, *J. Soc. and Clinical Psych.* 8 (1989), 176-189.

[5] S.H. Chew, A generalization of the quasilinear mean with application to the measurement of income inequality and decision theory resolving the Allais Paradox, *Econometrica* 51 (1983), 1065-1092.

[6] E. Dekel, B. Lipman and A. Rustichini, A unique subjective state space for unforeseen contingencies, *Econometrica* 69 (2001), 891-934.

[7] E. Dekel, B. Lipman and A. Rustichini, Temptation driven preferences, 2005.

[8] J.H. Dreze, Decision theory with moral hazard and state-dependent preferences, Ch. 2 in *Essays on Economic Decisions under Uncertainty*, Cambridge, U.K. 1987.

[9] L.G. Epstein, An axiomatic model of non-Bayesian updating, *Rev. Econ. Stud.*, forthcoming.

[10] L.G. Epstein and M. Marinacci, Coarse contingencies, 2005.

[11] L.G. Epstein, J. Noor and A. Sandroni, Non-Bayesian updating: a theoretical framework, 2005.

[12] L. Festinger, *A Theory of Cognitive Dissonance*, Stanford U. Press, 1957.

[13] D. Fudenberg and D.K. Levine, A dual model of impulse control, 2005.

[14] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique priors, *J. Math. Econ.* 18 (1989), 141-153.

[15] W.N. Goetzmann and N. Peles, Cognitive dissonance and mutual fund investors, *J. Financial Research* 20 (1997), 145-158.

[16] F. Gul and W. Pesendorfer, Temptation and self-control, *Econometrica* 69 (2001), 1403-35.

[17] R. Haagsma and N. Koning, Endogenous mobility-reducing norms, *J. Econ. Behavior and Organization* 49 (2002), 523-547.

[18] R.E. Knox and J.A. Inkster, Postdecision dissonance and post time, *J. Personality and Social Pysch.* 8 (1968), 319-323.

[19] I. Kopylov, Temptations in general settings, 2005.

[20] I. Kopylov, A parametric model of ambiguity aversion, 2005.

[21] D. Kreps, A representation theorem for 'preference for flexibility', *Econometrica* 47 (1979), 565-577.

[22] D. Kreps, *Notes on the Theory of Choice*, Westview, 1988.

[23] D. Kreps, Static choice in the presence of unforeseen contingencies, in *Essays in Honour of F. Hahn*, P. Dasgupta et al eds., MIT Press, 1992.

[24] M. Machina, Temporal risk and the nature of induced preferences, *J. Econ. Theory* 33 (1984), 199-231.

[25] J. Noor, Temptation, welfare and revealed preference, Boston U., 2005.

[26] M. Rabin, Cognitive dissonance and social change, *J. Econ. Behav. and Organization* 23 (1994), 177-194.

[27] M. Rabin and J.L. Schragg, First impressions matter: a model of confirmatory bias, *Quart. J. Econ.* 114 (1999), 37-82.

[28] R.T. Rockafeller, *Convex Analysis*, Princeton U. Press, 1972.

[29] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57, 571-587, 1989.

[30] J. Smith, Cognitive dissonance and the overtaking anomaly: psychology in the principal-agent relationship, Princeton, 2004.

[31] R.H. Strotz, Myopia and inconsistency in dynamic utility maximization, *Rev. Ec. Studies* 23 (1955), 165-180.

[32] S.E. Taylor, *Positive Illusions: Creative Self-Deception and the Healthy Mind*, Basic Books, New York: 1989.

[33] S.E. Taylor and J.B. Brown, Illusion and well-being: a social psychological perspective on mental health, *Psych. Bulletin* 103 (1988), 193-210.