Econ Theory (2010) 44:415–444
DOI 10.1007/s00199-009-0465-x

RESEARCH ARTICLE

Utility from anticipation and personal equilibrium

Botond Kőszegi

Received: 24 August 2007 / Accepted: 15 April 2009 / Published online: 1 May 2009

Abstract I develop a dynamic model of individual decisionmaking in which the agent derives utility from physical outcomes as well as from rational beliefs about physical outcomes ("anticipation"), and these two payoff components can interact. Beliefs and behavior are jointly determined in a *personal equilibrium* by the requirement that behavior given past beliefs must be consistent with those beliefs. I explore three phenomena made possible by utility from anticipation, and prove that if the decisionmaker's behavior is distinguishable from a person's who cares only about physical outcomes, she *must* exhibit at least one of these phenomena. First, the decisionmaker can be prone to self-fulfilling expectations. Second, she might be time-inconsistent even if her preferences in all periods are identical. Third, she might exhibit informational preferences, where these preferences are intimately connected to her attitudes toward disappointments. Applications of the framework to reference-dependent preferences, impulsive behaviors, and emotionally difficult choices are discussed.

Keywords Anticipation · Personal equilibrium · Time inconsistency · Disappointment aversion

JEL Classification B49 · D89

I thank an anonymous referee, George Akerlof, Andrew Caplin, Steven Goldman, Paul Heidhues, Wei Li, Mark Machina, Máté Matolcsi, Markus Möbius, Wolfgang Pesendorfer, Ben Polak, Attila Pór, Matthew Rabin, Luca Rigotti, Tanya Rosenblat, Jacob Sagi, and Chris Shannon for comments.

B. Kőszegi (✉)
Department of Economics, University of California,
508-1 Evans Hall #3880, Berkeley, CA 94720, USA
e-mail: botond@econ.berkeley.edu

1 Introduction

Motivated by the obvious recognition that most people regularly experience emotions related to anticipation, there has been some recent interest in incorporating these emotions into theories of intertemporal individual decisionmaking (Loewenstein 1987; Caplin and Leahy 2001; Kőszegi 2006, for example). In most existing models, the decisionmaker's utility from anticipation is separable from physical sources of utility such as consumption and health. The essential aspect of many applications of economic interest, however, is that anticipation interacts with other carriers of utility. Whether a person is optimistic or pessimistic about the future, and whether she is anxious about something, affects her motivation and energy to carry out costly tasks and to take risks. Conversely, the activity someone is performing at the moment influences—for instance by distracting her from unpleasant thoughts—the anticipatory utility she derives from future events. And in evaluating a given consumption outcome, decisionmakers typically compare it to relevant reference points, which are determined at least in part by the outcomes they had anticipated receiving (Kőszegi and Rabin 2006).

This paper develops a general framework for modeling decisionmaking that can accommodate all the above phenomena. I introduce a model in which preferences depend on anticipation, and propose a solution concept, personal equilibrium, for endogenously determining anticipation and behavior in any environment. I identify several types of behavior that can arise in my model but not in expected-utility and other classical models of individual decisionmaking, and show that these deviations constitute an exhaustive list in a strong sense: if the decisionmaker does not exhibit any of them, she is indistinguishable from a classical decisionmaker. Finally, I discuss the potential importance of the model for a few economic questions.

Section 2 presents the setup. There are two periods, $t = 1, 2$. A physical outcome z_t is realized in each of the two periods, and the decisionmaker's choice problem consists of choosing lotteries over these outcomes. But in addition to physical outcomes, an expectation or anticipation f_1, a probability distribution over future outcomes, is irreversibly realized in period 1. Each self's utility function u_t can depend on the entire vector of outcomes (z_1, f_1, z_2).

To deviate as little as possible from previous models, I assume that the decisionmaker can predict her (stochastic) environment and her own behavior in that environment, so that she has rational expectations in period 1. Given that anticipation can interact with utility from physical outcomes, this assumption raises a conceptual issue: because future preferences (and therefore behavior) can depend on expectations today, and rational expectations depend on future behavior, it is impossible to apply standard dynamic solution concepts based on optimization and backward-induction principles. I define a solution concept, *personal equilibrium*, as a situation where each self maximizes her expected utility, and f_1 coincides with the stochastic outcome implied by behavior in period 2—which typically depends on f_1.[1] Personal equilibrium

[1] This raises a second difficulty relative to previous models. In defining personal equilibrium, I need to be explicit about the temporal placement of expectations, since a self's maximization problem depends crucially on whether she moves before f_1 is realized—so that she can influence f_1—or after it is realized—so that she takes f_1 as given.

endogenizes a decisionmaker's expectations and behavior in any economic situation. Since its introduction in previous versions of the paper, variants of personal equilibrium have been used in a number of theories in the literature (Daido and Itoh 2005; Heidhues and Kőszegi 2005; Stone 2005; Kőszegi and Rabin 2006, 2007, 2009; Masatlioglu and Nakajima 2007; Bénabou 2008; Heidhues and Kőszegi 2008, for example).

In Sect. 3, I explore three types of behavior in this model, all of which have novel features that do not arise in natural comparison worlds: the expected-utility-over-physical-outcomes (henceforth EU), Kreps and Porteus (1978, henceforth KP), and Caplin and Leahy (2001, henceforth CL) models. First, due to the interaction between expectations and behavior, there could be multiple personal equilibria even when only self 2 makes a non-trivial choice. For example, if a person had been pessimistic about what her life would bring, she may have become disinterested and lethargic, and now prefer not to exercise, study, or make other costly investments. As a result, she fulfills her pessimistic expectations. But if she had been optimistic, she is now more energetic, and can more easily make costly investments—again fulfilling her expectations. Although she might not be indifferent between these two personal equilibria, she cannot choose the one with higher expected utility. Hence, within non-trivial bounds, her behavior is not uniquely determined by her preferences.

Second, the fact that self 2 does not internalize her effect on anticipatory feelings—those being irreversibly realized in period 1—can lead to time inconsistency. To demonstrate this, Sect. 3.2 begins by offering a definition of time consistency appropriate for the paper's framework: whenever self 2 prefers to deviate from self 1's expectations, self 1 must have preferred to anticipate and carry out the deviation. As an example of time-inconsistent behavior, suppose a dieter wishes she would exercise perfect self-control in the future, but if that was what she expected, she would splurge a little bit. Similarly to the logic in Loewenstein (1987) and Caplin and Leahy (2001), the dieter's realization that she would splurge feeds back into her current expectations, and may make both selves worse off. But beyond the models in Loewenstein (1987) and Caplin and Leahy (2001), if utility from anticipation and physical outcomes interact, the emotional distress resulting from her realization can exacerbate the dieter's self-control problem. Furthermore, these patterns of time inconsistency can arise—and taking an action that fails to maximize utility can be the unique personal equilibrium—even if both selves have the *same* utility function over the stream of physical outcomes and expectations.

Third, self 1 might exhibit an intrinsic preference for information. In previous models, such preferences (Kreps and Porteus 1978; Caplin and Leahy 2001; Grant et al. 1998, 2000) are determined exclusively by the shape of the utility function in beliefs about the future.[2] The natural analog of this is the shape of u_1 in the expectations f_1,

[2] Informational preferences can also arise in models where the decisionmaker cares only about physical outcomes, but is time-inconsistent. With intertemporal conflicts, information acquisition serves a strategic purpose, manipulating the actions of future selves (Carrillo 1998; Carrillo and Mariotti 2000; Bénabou and Tirole 2002). With anticipation, informational preferences can arise even in the absence of strategic considerations.

determining the decisionmaker's attitude toward the immediate experience of living with uncertainty about what will happen to her. In my model, however, informational preferences can arise even if u_1 is linear in f_1, so that this effect is neutralized. To illustrate the idea, suppose an employee learns that she will receive a bonus that is equally likely to be \$10,000 or \$20,000. If her beliefs remain uncertain about the amount, upon getting the bonus she will be disappointed with probability one-half and pleasantly surprised with probability one-half. If she knows the amount, neither will happen, and because she just learned that she would receive a bonus at all, her current disappointment may also be minimal. Therefore, if she dislikes future disappointments more than she likes future pleasant surprises—she is disappointment averse—she would choose to find out the size of the bonus. Section 3.3 formally defines disappointment aversion and characterizes its relationship with informational preferences: if u_1 is linear in f_1, the decisionmaker prefers early full resolution of uncertainty if and only if she is disappointment averse, and if her disappointment aversion satisfies an intuitive condition, she prefers more information to less even when uncertainty is not fully resolved. Therefore, managing disappointments emerges as an important and complex determinant of informational preferences.

Given the wide variety of emotions related to anticipation (hope, fear, anxiety, disappointment, suspense, savoring, etc.), it may seem that the above are merely a few examples of an unwieldy set of new patterns of behavior implied by utility from anticipation. Instead, Sect. 4 establishes a surprising and important result: if a model incorporating anticipation into decisionmaking is to generate observably different behavior from EU, it must feature at least one of the above three phenomena. That is, if informational preferences, utility-ranked multiple equilibria, and time inconsistency are ruled out, there is a utility function v defined only over physical outcomes such that in every decision problem the agent behaves as if she was maximizing the expectation of v. Analogously, any difference between my model and KP has to do with the interaction between anticipation and physical outcomes or with time inconsistency. These results provide a partial answer to a natural question raised by any model that enriches previous ones: how the addition of new elements—here, utility from anticipation—is reflected in choice behavior.

In Sect. 5, I demonstrate the usability of the model by discussing possible applications. In a reference-dependent model, assuming that the reference point is expectations, and imposing personal equilibrium, leads to the recent model of Kőszegi and Rabin (2006, 2007). In a dynamic choice problem, supposing that a person's impatience depends on anticipatory feelings leads to a theory of the role of emotions in self-regulation. And positing that the act of choice itself affects anticipatory utility by drawing attention to its possible consequences might provide a framework for studying circumstances under which people avoid difficult decisions.

The paper is organized as follows. Section 2 introduces notation and the concept of personal equilibrium, and deals with existence. Section 3 illustrates the novel kinds of behavior in the model. Section 4 gives the observational equivalence results. Section 5 provides possible applications of the theory. Section 6.1 offers comments on the model, and discusses related literature. Section 6.2 deals with issues arising from the fact that preferences in the model do not have a revealed-preference foundation. Proofs of all major results are in the Appendix.

2 Anticipation and decisionmaking

This section formulates a model of decisionmaking when anticipation affects utility. As a motivating example, consider a basketball player going through a period of rehabilitation after knee surgery, during which she must decide how hard to work in her training program. This situation is likely to evoke a multitude of emotions related to anticipation: the athlete might, for instance, derive pleasure from anticipating being back on the court, and she may feel disappointed if her expectations to be in top form again do not materialize. Below, I relate these emotions and the athlete's situation to formal elements of my model.

Formally, the decisionmaker is involved in a two-period decision problem; Kőszegi (2003) presents an extension of the model's setup, definitions, and results to any finite horizon. In periods $t \in \{1, 2\}$, "physical outcomes" $z_t \in Z_t$ are realized, which are standard consequence-based sources of utility usually assumed relevant in economics. In the above example, the state of the athlete's knee is a physical outcome. I assume that each Z_t is a Polish (complete separable metric) space. Most payoff spaces used in applications are Polish; the reason for using this class is that all spaces constructed from Z_1 and Z_2 below are also Polish in the appropriate topologies. Denote by $\Delta(S)$ the space of Borel probability measures over a Polish space S (endowed with the weak-* topology). For any outcome $z \in Z_1, Z_2$, let δ_z be the probability measure that assigns unit mass to z.

In addition to the physical outcomes, the decisionmaker may also care about expectations she forms in period 1 regarding her future physical outcomes. Formally, an anticipatory outcome $f_1 \in F_1 \equiv \Delta(Z_2)$ is realized in period 1, and this happens in an *irreversible* manner: whether or not the athlete ends up as fit as she expected, at that point it is impossible to change the expectations and their effect on utility. The central part of the model is the formation of these expectations, which will be specified in detail below.

The expectations and physical outcomes constitute all the payoffs the decisionmaker cares about. In line with most previous research, I assume that for each $t \in \{1, 2\}$, self t's preferences take the expected-utility form over the enriched outcome space: they are representable by a von Neumann–Morgenstern utility function u_t defined on $Z_1 \times F_1 \times Z_2$, where u_1 and u_2 are continuous. The foremost reason to assume expected utility is methodological—extending past work by both enriching the consequence space and relaxing expected utility would make it difficult to disentangle the effects of these two major changes. Throughout the paper, for any measurable function h and probability measure H, $E_H h(\cdot)$ denotes $\int h(x)dH(x)$.[3] For notational convenience, I will suppress arguments in the utility function that are deterministic and do not depend on the agent's choice.

In this formulation, anticipation can affect a decisionmaker's utility in two important ways. First, f_1 can directly enter self 1's utility, as when the athlete derives plea-

[3] Using this notation, given history z_1, f_1, self 2 prefers the lottery l_2 over l_2' (where both are in $\Delta(Z_2)$) if and only if $E_{l_2} u_2(z_1, f_1, \cdot) \geq E_{l_2'} u_2(z_1, f_1, \cdot)$. And self 1 prefers l_1 over l_1' (both in $\Delta(Z_1 \times F_1 \times Z_2)$) if $E_{l_1} u_1(\cdot) \geq E_{l_1'} u_1(\cdot)$.

sure from anticipating being healthy. Second, expectations can interact with physical outcomes in the determination of self 1's or self 2's utility, as when the athlete is disappointed at not realizing her expectations to be healthy. CL and other previous work have explored the implications of the former possibility, but much less the implications of the latter one. In fact, imposing the condition that u_1 and u_2 are additively separable in $Z_1 \times Z_2$ and F_1, my model is equivalent to CL.

For some statements in the paper, it is necessary that indifference between two options be observable to an outsider. Without additional assumptions, this is not the case because the agent might always choose (say) option 1 over an equally preferable option 2, so that indifference is observationally equivalent to strict preference. To make indifference observable, I assume that for each $t \in \{1, 2\}$, there are outcomes $\underline{z}_t, \overline{z}_t \in Z_2$ such that self t is known to *strictly* prefer $\overline{z}_t$ over $\underline{z}_t$ (holding other outcomes constant). Then, giving the agent an arbitrarily small probability of getting $\overline{z}_t$ rather than $\underline{z}_t$ swings the agent's choice to option 2, revealing that option 1 was not strictly preferred. Along with non-satiation below, this assumption seems very weak, and is likely to hold in most applications.

With preferences given, I turn to modeling the decisionmaking environment. A decision problem in period 2 is a compact set $d_2 \subset \Delta(Z_2)$ satisfying the condition that the set of feasible period-2 outcomes, $\mathrm{Cl}\left[\cup_{l_2 \in d_2} \mathrm{supp} l_2\right]$ (where Cl denotes closure), is compact.[4] Let D_2 be the set of all such period-2 decision problems, endowed with the topology generated by the Hausdorff metric. A decision problem at time 1 is a compact set $d_1 \subset \Delta(Z_1 \times D_2)$ satisfying that the set of feasible period-1 outcomes,

$$\mathrm{Cl}\left\{z_1 | \exists d_2 \in D_2 \text{ s.t. } (z_1, d_2) \in \cup_{l_1 \in d_1} \text{supp } l_1\right\},$$

and the set of feasible period-2 outcomes,

$$\mathrm{Cl}\left\{z_2 | \exists (z_1, d_2) \in Z_1 \times D_2 \text{ s.t. } (z_1, d_2) \in \cup_{l_1 \in d_1} \text{supp } l_1 \text{ and } z_2 \in \cup_{l_2 \in d_2} \text{supp } l_2\right\},$$

are compact. Hence, a decision in period 1 induces a lottery over pairs of physical outcomes in that period and future decision problems. Let D_1 be the set of all possible period-1 decision problems, endowed with the topology generated by the Hausdorff metric. Analogously to the feasibility of period-1 and period-2 outcomes, call a pair $(z_1, d_2) \in Z_1 \times D_2$ feasible if $(z_1, d_2) \in \mathrm{Cl}\left[\cup_{l_1 \in d_1} \text{supp } l_1\right]$, and define the set of feasible period-2 decision problems as $\mathrm{Cl}\left\{d_2 | \exists z_1 \in Z_1 \text{ s.t. } (z_1, d_2) \in \cup_{l_1 \in d_1} \text{supp } l_1\right\}$.

It now remains to specify the decisionmaker's behavior in this environment, especially how expectations are formed in period 1. I propose a solution concept, personal equilibrium, to determine both expectations and behavior in an internally consistent way. The standard ingredient of a personal equilibrium is a pair of history-contingent measurable strategies of the two selves,

---

[4] This condition is a technical one that, along with a similar condition on d_1 below, ensures that the agent's utility over feasible outcomes is bounded in any given decision problem, so that all relevant expectations in the decision problem are well-defined.

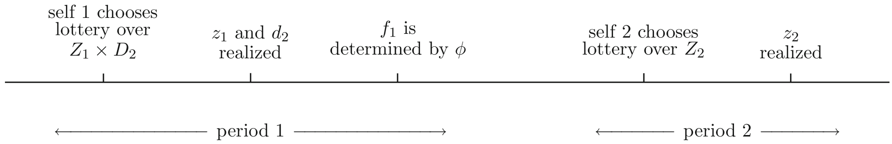

Fig. 1 Timing of decisions and outcomes

$$\sigma_1 \in d_1 \text{ and } \sigma_2 : Z_1 \times F_1 \times D_2 \to \Delta(Z_2),$$

where $\sigma_2(z_1, f_1, d_2) \in d_2$ for all z_1, f_1, d_2. The novel feature of personal equilibrium is a measurable "anticipation function"

$$\phi : Z_1 \times D_2 \to F_1,$$

which determines the decisionmaker's expectations as a function of what happened up to the end of period 1. The specification of ϕ embeds an assumption about *timing*: f_1 is realized after z_1 and d_2, as illustrated in Fig. 1. This is the simplest specification which captures both the agent's incentives to affect her expectations (in period 1), and the possibility that she acts after they are realized (in period 2). The model is sufficiently flexible to accommodate many possibilities about how the decisions and outcomes are spaced. There could be a short or long time between the realization of f_1 and period-2 choice, and self 2's decision may immediately lead to the realization of z_2, or there might be a significant delay. The period-1 or period-2 decision could even capture, in a reduced form, intertemporal choices whose outcomes are spread out over time. Now:

Definition 1 Given a $d_1 \in D_1$, the profile $(\sigma_1, \sigma_2, \phi)$ constitutes a *personal equilibrium* if

1. (**Optimization**)

$$\begin{array}{lll} \sigma_1 \in \text{argmax}_{\gamma \in d_1} & E_\gamma \left[E_{\phi(z_1,d_2)} u_1(z_1, \phi(z_1, d_2), \cdot) \right], & \text{and} \\ \sigma_2(z_1, f_1, d_2) \in \text{argmax}_{\gamma \in d_2} & E_\gamma u_2(z_1, f_1, \cdot) & \text{for all } f_1 \text{ and} \\ & & \text{feasible } (z_1, d_2). \end{array}$$

2. (**Internal Consistency**) For all feasible (z_1, d_2),

$$\phi(z_1, d_2) = \sigma_2(z_1, \phi(z_1, d_2), d_2).$$

Given a feasible (z_1, d_2), ϕ and σ_2 constitute a *continuation personal equilibrium* if for z_1, d_2, condition 1 is satisfied for self 2 and condition 2 is satisfied.

Personal equilibrium can be viewed as an extension of subgame-perfect equilibrium to a model incorporating utility from anticipation, and in fact the two concepts are equivalent whenever utility does not depend on anticipation. The assumption of optimization at each node is analogous to that in subgame-perfect equilibrium; the only complication is that in self 1's optimization problem, $\phi(z_1, d_2)$ determines both expectations in period 1 and the distribution of outcomes in period 2. In addition to maximization at each node, personal equilibrium incorporates an assumption about the formation of expectations. Namely, an internal consistency condition must hold: the distribution of outcomes generated by self 2's optimal behavior given period-1 expectations must coincide with those expectations.

The internal consistency condition in Definition 1 amounts to assuming that the decisionmaker has rational (correct) expectations about her future behavior and outcomes. While this is a strong assumption, it is a structured way to capture the notion that in most situations people would have at least some ability to predict their own behavior. In fact, exactly because personal equilibrium requires someone to predict *only* her own behavior, it is not subject to much of the criticism leveled at classical equilibrium concepts—which require players to predict *others'* beliefs and behavior, and which are being used ubiquitously despite this strong requirement. Rational expectations are also useful to maintain from the methodological vantage point of deviating minimally from previous models.[5]

Implicit in the definition of personal equilibrium is the assumption that the anticipation function is not chosen by either self 1 or self 2. On the one hand, this means that self 1 is in general unable to select from all expectations that are possible given her decision problem, even though expectations are realized at time 1 just as physical outcomes are. On the other hand, by the time self 2 makes her choice, she is also unable to influence anticipation in period 1, even though it depends partly on her expected behavior.

To complete the introduction of personal equilibrium, I show that it exists under general conditions. As in many other models, it is necessary to allow for mixed strategies to guarantee existence.[6]

Theorem 1 *If any feasible d_2 is convex, a personal equilibrium exists.*

In the next sections, I compare the above theory to the EU, KP, and CL models. For this purpose, I formally define what it means for the agent to behave indistinguishably from these alternatives.

Definition 2 The agent behaves according to EU if there is a function $v : Z_1 \times Z_2 \to \mathbb{R}$ such that in any decision problem, $l \in \Delta(Z_1 \times Z_2)$ is a distribution of outcomes induced in personal equilibrium if and only if it maximizes the expectation of v among feasible distributions of outcomes.

The agent behaves according to KP if there are functions $v_1 : Z_1 \times \mathbb{R} \to \mathbb{R}$ and $v_2 : Z_1 \times Z_2 \to \mathbb{R}$ such that i.) σ_1 and σ_2 are personal-equilibrium strategies if and

5 Despite these advantages, there is one way in which rationality may be less reasonable to assume in this context than in more classical settings. Specifically, because rational expectations prevent one from manipulating beliefs, a person who derives utility from anticipation may *prefer* not to be rational to increase her utility (Brunnermeier and Parker 2005; Landier 1999). This is never the case for decisionmakers who only care about physical outcomes. However, incorporating existing reduced-form formalizations of information-processing biases (Rabin and Schrag 1999; Akerlof and Dickens 1982; Gervais and Odean 2001; Manove and Padilla 1999, for example), and especially developing a full-fledged model of how people fool themselves, is beyond the scope of this paper.

6 To see this, observe that Example 3 in Sect. 3.2 does not have a pure-strategy personal equilibrium when self 2's decision problem is $\{\delta_z, \delta_{z'}, \delta_\zeta, \delta_{\zeta'}\}$.

only if σ_1 maximizes the expectation of $v_1\left(z_1, \max_{l_2 \in d_2} E_{l_2} v_2(z_1, \cdot)\right)$ over d_1 and for any f_1 and feasible z_1, d_2, σ_2 maximizes the expectation of $v_2(z_1, \cdot)$ over d_2.

The agent behaves according to CL if u_1 and u_2 are additively separable in f_1 and the physical outcomes.

Section 6.1 further discusses related literature and comments on the model.

3 Key properties of preferences with anticipation

The main contribution of this paper is to put forward a general framework for thinking about individual decisionmaking when utility from anticipation affects choice. At the same time, an important goal is to establish some general properties of the proposed model. In that spirit, this section explores three phenomena made possible by utility from anticipation. I focus on these phenomena because they are likely to be important in many applications, and because (as I establish in Sect. 4) any deviation from the behavioral predictions of more classical models is related to at least one of them.

In each subsection, I develop an example in which some aspect of the agent's behavior is observably inconsistent with EU, KP, or CL models. This means that there is a set of decision problems and a corresponding set of personal-equilibrium choices that are never made by any decisionmaker with EU, KP, or CL preferences. After discussing the example, in each subsection I introduce a restriction that rules out the particular behavior. The characterization theorems in Sect. 4 show that if these restrictions are imposed, no non-standard forms of behavior remain.

3.1 Self-fulfilling expectations

Recall that for an outcome z, δ_z assigns unit mass to z. Consider the following simple example.

Example 1 At $t = 1$, choice is degenerate and (z_1, d_2) is deterministic. At $t = 2$, there are two outcomes, z and z', to choose from. As a function of expectations (δ_z or $\delta_{z'}$) and outcomes (z or z'), u_2 is given by the following table:

	z	z'
δ_z	3	2
$\delta_{z'}$	0	1

In Example 1, expecting z and then choosing it is a personal equilibrium, and so is expecting z' and choosing it. Furthermore, in both of these personal equilibria, self 2 strictly prefers the option she chooses over the alternative. With EU, KP, and CL, in contrast, all options that self 2 is willing to choose from a choice set yield her the same utility, maximizing the expectation of $v(z_1, \cdot)$, $v_2(z_1, \cdot)$, and $u_2(z_1, \cdot)$ in the case of EU, KP, and CL, respectively. In fact, the same is true in non-expected utility models more generally, including Chew and Epstein (1989), Epstein and Zin (1989), and Skiadas (1998).

The above difference can be revealed in behavior by exploiting that self 2 strictly prefers $\overline{z}_2$ to $\underline{z}_2$. For a sufficiently small $\epsilon > 0$, for instance, it is a personal equilibrium for the decisionmaker to choose $(1-\epsilon)\delta_z + \epsilon\delta_{\underline{z}_2}$ from the choice set $\{(1-\epsilon)\delta_z + \epsilon\delta_{\underline{z}_2}, (1-\epsilon)\delta_{z'} + \epsilon\delta_{\overline{z}_2}\}$, and to choose $(1-\epsilon)\delta_{z'} + \epsilon\delta_{\underline{z}_2}$ from the choice set $\{(1-\epsilon)\delta_{z'} + \epsilon\delta_{\underline{z}_2}, (1-\epsilon)\delta_z + \epsilon\delta_{\overline{z}_2}\}$. These choices are inconsistent with EU, KP, and CL, however, because in these models the former choice would mean the agent strictly prefers z to z', and the second choice would mean she strictly prefers z' to z.

In a model of loss aversion that builds on the insights of prospect theory (Kahneman and Tversky 1979), Kőszegi and Rabin (2006) provide a natural example of the kind of multiplicity in Example 1.[7] Suppose that a consumer's utility from shoes and money depends on how her consumption outcome compares to her prior expectations in these two dimensions, and that in both dimensions she is more sensitive to falling short of her expectations than to ending up above them. Then, if she had expected to buy the shoes, she would assess not doing so as a loss in shoes and a gain (of the amount not spent) in money. Being more sensitive to losses, she strictly prefers to conform to her expectations and buy the shoes. In contrast, if she had expected not to buy, she would assess buying as a loss in money and a gain in shoes, and to avoid the loss she strictly prefers not to buy. Hence, for a range of price levels, there are multiple personal equilibria.[8]

In the characterization result, I will rule out the central feature of Example 1—that there are two options the agent may choose, but she is not indifferent between them—by assuming that if there are two different personal-equilibrium choices, self 2 is indifferent between them in each personal equilibrium. Intuitively, this means that the agent's behavior is "stable" in that she can never be observed to make one choice from a choice set, and then to strictly prefer another option from the same set.

Definition 3 The decisionmaker is *stable* if for any $z_1 \in Z_1$ and $d_2 \in D_2$, if f_1 and f_1' are the period-1 expectations in two continuation personal equilibria, then

$$E_{f_1} u_2(z_1, f_1, \cdot) = E_{f_1'} u_2(z_1, f_1, \cdot).$$

Noting the kind of multiplicity issue above, Kőszegi and Rabin (2006, 2007) argue that a person should be able to make any credible plans for her own behavior, and hence propose the refinement that self 1 can select her favorite or "preferred personal equilibrium."[9] Assuming preferred personal equilibrium rather than personal equilibrium, stability is not necessary for the key characterization result (Theorem 3). But while this refinement seems plausible in many settings, there are also some situations—such as the example of optimistic versus pessimistic life outlook in the introduction—where it is not clear that a person can select the best personal equilibrium. Furthermore, Kőszegi and Rabin (2009) show that in a dynamic situation where a decisionmaker does not

[7] Kőszegi and Rabin's model is discussed in more detail in Sect. 5.

[8] Example 1 is also easily adapted to handle the story of optimistic and pessimistic attitudes mentioned in the introduction. In that example, actions are investments into the future rather than decisions with immediate consequences, and the effort cost of investment (e.g. exercising) is higher when expectations are more pessimistic.

[9] The proof of Theorem 1 in the Appendix implies that a preferred personal equilibrium exists.

necessarily start off with preferred-personal-equilibrium beliefs, switching to those beliefs might generate negative utility, so that she does not play the preferred personal equilibrium. For these reasons, the current paper focuses on the less restrictive notion of personal equilibrium.

3.2 Time inconsistency

This section argues that anticipatory utility has important implications related to time inconsistency. To address this issue, however, I need to begin by defining time consistency in this framework. Whenever an existing concept is extended to a new setting, a judgment must be made as to the essence of the notion one is trying to capture. I will think of "intuitive time consistency" (ITC) as a situation where future selves are willing to carry out earlier selves' favorite plans. The following example gives a simple situation in which self 2 is not willing to do so. Based on this example, I argue that two conventional formal definitions of time consistency are unsuited for the current model. I then suggest an appropriate replacement.

Example 2 There is no choice in period 1, and self 2 can choose from the set $\{z, \zeta'\}$. The utilities $u_1 = u_2 \equiv u$ are given by the following table as a function of expectations and outcomes:

	z	ζ'
δ_z	2	3
$\delta_{\zeta'}$	0	1

Since $u_1(\delta_z, z) > u_1(\delta_{\zeta'}, \zeta')$, self 1 would like self 2 to choose z. Yet in the unique personal equilibrium, self 2 chooses ζ'. That is, the two intertemporal selves have the same utility function u over streams of physical outcomes and expectations, yet they fail to maximize the expectation of u. For example, a restaurant-goer's favorite *rational* plan might be to expect to eat a light meal and then eat a light meal, rather than to expect to eat a fat-laden meal and then eat a fat-laden meal. But an even better outcome might be to expect to eat light, and then surprise herself with a fancy fat-laden meal. Not being able to surprise herself in a personal equilibrium, she is stuck with the ex-ante suboptimal option.

Standard definitions of time (in)consistency cannot fully account for the violation of ITC in this example. One definition, which I will call "preference time consistency" (PTC), requires selves 1 and 2 to have the same (possibly history-dependent) preferences over outcomes in period 2. Since the agent in Example 2 (for whom $u_1 = u_2$) is PTC, this definition is clearly inappropriate for the current model.

The other standard definition, which I will call "commitment time consistency" (CTC), requires that the agent never have a value from commitment. This definition faces a serious problem as well. Example 1, which features multiple (possibly utility-ranked) equilibria, violates CTC, yet in that example self 2 is perfectly willing to carry out self 1's favorite course of action—so long as self 1 expects her to do so. In other words, even in a low-utility equilibrium, it is not self 2's behavior that is "wrong," but self 1's expectations, and if self 1 could change her expectations, self 2's preferred action would change as well. Hence, the *exclusive* reason self 1 prefers commitment is to change her own expectations, a situation that does not seem to violate ITC.

The above problems with extending standard notions of time consistency arise from the fact that self 1 exists before f_1 is realized, while self 2 exists afterwards. My proposed definition of time consistency is an extension of PTC that deals with this complication.

Definition 4 The decisionmaker is *time consistent* if for all $z_1 \in Z_1, d_2 \in D_2$, and $f_1, f_1' \in d_2$,

$$E_{f_1'} u_2(z_1, f_1, \cdot) > E_{f_1} u_2(z_1, f_1, \cdot) \implies E_{f_1'} u_1(z_1, f_1', \cdot) > E_{f_1} u_1(z_1, f_1, \cdot). \quad (1)$$

Definition 1 says that if self 2 prefers f_1' over following through on an expectation to choose f_1, then self 1 also prefers f_1' to f_1. According to this definition, preferences in Example 2 are time inconsistent ($u_2(\delta_z, \zeta') > u_2(\delta_z, z)$, yet $u_1(\delta_z, z) > u_1(\delta_{\zeta'}, \zeta')$). Intuitively, while self 2 cares as much about period-1 expectations as self 1 does, these are already realized by the time she makes her choice, so she does not internalize her effect on them. And self 1's realization that self 2 would not do so decreases both selves' utility. This argument for why the passage of feelings can lead to time inconsistency was first given by Loewenstein (1987) and Caplin and Leahy (2001). But both papers then assumed time-inconsistent preferences in the standard way, by making the utility functions different for different selves. Example 2 formally shows that time inconsistency can follow *purely* from the feedback of self 2's actions into self 1's expectations, not from different preferences.

When the utility from later physical outcomes interacts with earlier expectations, an important manifestation of time inconsistency is the context-dependence of preferences. That is, a person's behavior can depend on unchosen alternatives even when only a single decision is made:

Example 3 There is no choice in period 1. In period 2, four possible outcomes are $\{z, z', \zeta, \zeta'\}$. Utilities $u_1 = u_2 \equiv u$ are given by the following table as a function of expectations and outcomes:

	ζ	z	z'	ζ'
δ_ζ	12	13	10	11
δ_z	14	16	15	17
$\delta_{z'}$	5	3	4	2
$\delta_{\zeta'}$	10	7	9	8

If the decisionmaker's choice set is $\{z, z', \zeta'\}$, the unique pure-strategy personal equilibrium is to choose z'. For example, the athlete in rehab used to motivate the model in Sect. 2 may prefer to look forward to training hard and being at full strength again (z), but be frustrated in this if self 2 has a choice to do a mediocre job (ζ') instead. This undermines the athlete's optimism, and she ends up not training at all after her injury (z'). But if the decisionmaker's choice set is $\{z, z', \zeta\}$, the unique pure-strategy personal equilibrium is to choose z. Intuitively, if she has the option of using a motivating trainer (choosing ζ), the athlete cannot get stuck not training. In fact, realizing this changes her attitudes toward rehab, and she chooses to train even by herself. The option of working with a trainer is necessary not because she ends up using that option, but because it ensures to her that she will not get stuck in a bad personal equilibrium. Therefore, the nature of a athlete's unchosen alternatives—whether they bolster or undermine motivating expectations—are crucial in determining her behavior.[10]

Example 3 is inconsistent with EU, KP, or CL models. For an EU, KP, or CL decisionmaker to choose z' from $\{z, z', \zeta'\}$, self 2 must prefer it to z. Then, z cannot be the unique preferred choice from $\{z, z', \zeta\}$.[11] Similarly to Sect. 3.1, one can elicit these preferences by exploiting that self 2 prefers $\overline{z}_2$ to $\underline{z}_2$. For a sufficiently small ϵ, the unique personal equilibrium from the choice set $\{(1-\epsilon)\delta_z + \epsilon\delta_{\overline{z}_2}, (1-\epsilon)\delta_{z'} + \epsilon\delta_{\underline{z}_2}, \zeta'\}$ is $(1-\epsilon)\delta_{z'} + \epsilon\delta_{\underline{z}_2}$, and the unique personal equilibrium from the choice set $\{(1-\epsilon)\delta_z + \epsilon\delta_{\underline{z}_2}, (1-\epsilon)\delta_{z'} + \epsilon\delta_{\overline{z}_2}, \zeta\}$ is $(1-\epsilon)\delta_z + \epsilon\delta_{\underline{z}_2}$. This is impossible with EU, KP, or CL preferences, because in these models the former choice would mean the decisionmaker strictly prefers z' to z, and the latter choice would mean she strictly prefers z to z'.

Observation 1 below shows that if the decisionmaker is time consistent, choosing self 1's favorite option from a choice set is a personal equilibrium. This directly rules out Example 2, and also rules out Example 3 because self 1 cannot strictly prefer both z and z' over the other.

Observation 1 *Suppose the agent is time consistent. Then, for any $z_1 \in Z_1$ and $d_2 \in D_2$, if f_1^* solves $\max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot)$, there is a continuation personal equilibrium with expectation f_1^*.*

Proof Suppose by contradiction that f_1^* is not a continuation equilibrium. Then, there is an f_1' such that $E_{f_1'} u_1(z_1, f_1^*, \cdot) > E_{f_1^*} u_1(z_1, f_1^*, \cdot)$. Then, by time consistency, f_1^* does not solve the above maximization problem. □

While Observation 1 shows that time consistency is sufficient to rule out the kind of behavior in Example 3, it is clearly not necessary for that purpose. Ideally, for characterizing the differences between decisionmakers with and without emotions related to anticipation, one would like to rule out only "anticipation-related" time inconsistency. Unfortunately, it is not clear what such a concept would mean, so I use the general notion of time consistency in Definition 1.

3.3 Informational preferences

This section shows that the decisionmaker has informational preferences that are closely related to her disappointment aversion, a relationship that is both largely unex-

10 Relatedly, when preferences are time inconsistent, the agent may want to manipulate her own expectations to influence her future behavior. This is also true in standard settings (Carrillo 1998; Carrillo and Mariotti 2000; Bénabou and Tirole 2002).

11 Behavior can also depend on unchosen alternatives in time-inconsistent models with standard preferences. In that case, however, unchosen alternatives can only influence behavior through the strategic interaction between different selves, and hence not in situations where only a single decision is made.

plored in previous research and that can lead to observably different behavior from EU and KP. For related results on the connection between informational preferences and disappointment aversion, see Dillenberger (2008) and Kőszegi and Rabin (2009).

Throughout the section, I will use the following measure of disappointment aversion:

$$T_{z_1}(z_2, z_2') \equiv (u_1(z_1, \delta_{z_2}, z_2) - u_1(z_1, \delta_{z_2}, z_2')) - (u_1(z_1, \delta_{z_2'}, z_2) - u_1(z_1, \delta_{z_2'}, z_2')).$$

To interpret this expression, suppose that the agent prefers z_2 to z_2'. Then, the first difference represents the utility loss from the disappointing outcome of z_2' when expecting the outcome z_2. The second difference is the person's utility gain from the pleasant surprise of getting z_2 after expecting z_2'. Hence, $T_{z_1}(z_2, z_2')$ measures how much more self 1 dislikes future disappointments than she likes future pleasant surprises. Based on this measure, I define:

Definition 5 The decisionmaker is *disappointment averse* if $T_{z_1}(z_2, z_2') \geq 0$ for all $z_1 \in Z_1$ and $z_2, z_2' \in Z_2$. The decisionmaker is *disappointment neutral* if $T_{z_1}(z_2, z_2') = 0$ for all $z_1 \in Z_1$ and $z_2, z_2' \in Z_2$.

Throughout this section, I will assume that u_1 is linear in f_1. In most previous work on informational preferences (KP; CL; Epstein and Zin 1989; Grant et al. 1998, 2000), if utility is linear in beliefs about the future the agent is indifferent to decision-irrelevant information. Hence, this assumption allows me to isolate informational preferences that are unique to the current theory. Psychologically, the assumption means that the decisionmaker is indifferent to the current experience of insecurity about what will happen to her.

The following example will motivate the main results.

Example 4 Z_1 is trivial, and there are three possible outcomes z, z', z'' in period 2. u_1 is linear in f_1, it satisfies $u_1(z_1, f_1, z) > u_1(z_1, f_1, z'') > u_1(z_1, f_1, z')$ for all f_1, and $T_{z_1}(z, z') = 24$, $T_{z_1}(z, z'') = 2$, $T_{z_1}(z', z'') = 1$.

Suppose that the decisionmaker in Example 4 is facing a lottery that gives her z or z' with probability one-half each. Would self 1 choose to find out the outcome in advance?[12] If she learns it, she gets the expected utility

$$\frac{1}{2}u_1(\delta_z, z) + \frac{1}{2}u_1(\delta_{z'}, z'), \tag{2}$$

while waiting to resolve the uncertainty gives her the expected payoff

$$\frac{1}{2}u_1\left(\frac{1}{2}\delta_z + \frac{1}{2}\delta_{z'}, z\right) + \frac{1}{2}u_1\left(\frac{1}{2}\delta_z + \frac{1}{2}\delta_{z'}, z'\right). \tag{3}$$

[12] Following KP, the way to represent this decision in the framework of Sect. 2 is the following. Early resolution of uncertainty corresponds to a choice that leads with probability one-half to the singleton period-2 choice set containing z, and with probability one-half to the singleton period-2 choice set containing z'. Late resolution of uncertainty is the choice that leads, with probability one, to the singleton choice set with the fifty-fifty lottery between z and z'.

Using that u_1 is linear in f_1, the difference between Expressions (2) and (3) is $\frac{1}{4}T_{z_1}(z, z') = 6$. Hence, the decisionmaker prefers to resolve the uncertainty in the first period, and does so exactly because she is disappointment averse. Intuitively, leaving her beliefs uncertain exposes the agent to either a disappointment (if she gets the worse outcome z') or to a pleasant surprise (if she gets the better outcome z). If she dislikes future disappointments more than she likes future pleasant surprises, she would rather avoid both by learning the truth now.[13]

Now suppose that the decisionmaker is facing the lottery in which she can get z, z', or z'' with probability one-third each. Using a similar calculation to that above, the difference in expected utility between fully resolving this uncertainty and remaining completely ignorant is 3. But consider the opportunity to learn only whether the outcome is z'' or one of $\{z, z'\}$. If the person chooses to receive this information, with probability $2/3$ she will find that the outcome is in $\{z, z'\}$. In that case, she faces the same 50–50 lottery as above. Hence, the difference in expected utility between full learning and partial learning is $(2/3) \cdot 6 = 4 > 3$. Despite being disappointment averse, therefore, the person would rather remain ignorant than partially learn the truth. This example illustrates that—since the only surefire way to avoid future disappointments is to know the outcome with certainty—the implications of disappointment aversion for partial learning are qualitatively different from those for full learning above. Intuitively, of all possible disappointments, the decisionmaker is most sensitive to the frustration of getting z' after expecting z. Thus, learning that the outcome is z or z' sets her up for a major disappointment. For example, a job seeker's disappointment from getting her least favorite job might be much more devastating when the alternative was a great job than when it was, say, a mediocre job. If so, first finding out whether she has received the mediocre job decreases her expected utility.

This intuition illustrates that the decisionmaker does not like information if her disappointment aversion for some pairs of outcomes is disproportionately larger than for other pairs. The following is a sufficient condition that rules out such possibilities.

Definition 6 The agent is *regular disappointment averse* if for some $v_1, \ldots, v_K : Z_1 \times Z_2 \to \mathbb{R}$,

$$T_{z_1}(z_2, z_2') = \sum_{k=1}^{K} |v_k(z_1, z_2) - v_k(z_1, z_2')|.$$

Regular disappointment aversion can be motivated by the following consideration. Being disappointed means that one gets a worse outcome than one hoped for or expected, causing disutility beyond the bad outcome itself. The converse holds for pleasant surprises. Presumably, the disappointment or pleasant surprise, and therefore

[13] The example and discussion above ignore the possibility that a person may be averse to *current* disappointments. Intuitively, a person's aversion to being disappointed today should make her reluctant to receive information. To analyze this motive, however, a longer-horizon model is needed: with two periods, only self 1 can make decisions regarding timing of the resolution of uncertainty, and there are no past beliefs in period 1 relative to which self 1 can be disappointed. Hence, the examples and results in this section only identify the relationship between informational preferences and aversion to future disappointments, and are not intended as a complete analysis of disappointment-management motives.

also disappointment aversion, is related to aspects of the desirability of the outcomes in question. In this view, disappointment aversion is regular if it is linearly related to some of these aspects, captured in the functions v_k.[14] Note that by the triangle inequality, the three conditions in Example 4 clearly violate regular disappointment aversion.

I introduce the following formal concepts for the kinds of informational preferences in the above example.

Definition 7 The decisionmaker is *information neutral* if for all $z_1 \in Z_1$ and $f_1 \in \Delta(Z_2)$ with compact support,

$$\int u_1(z_1, f_1, z) df_1(z) = \int u_1(z_1, \delta_z, z) df_1(z).$$

She is *resolution loving* if for all $z_1 \in Z_1$ and $f_1 \in \Delta(Z_2)$ with compact support,

$$\int u_1(z_1, f_1, z) df_1(z) \leq \int u_1(z_1, \delta_z, z) df_1(z).$$

If for all $z_1 \in Z_1$, $\int u_1(z_1, f_1, z) df_1(z)$ is convex in f_1 over probability measures in $\Delta(Z_2)$ with compact support, then the decisionmaker is *information loving*.

Information neutrality means that the decisionmaker is indifferent to whether decision-irrelevant information about future outcomes is revealed in period 1 or 2. Resolution lovingness says that she prefers to fully resolve uncertainty in period 1. She is information loving if she prefers more information to less, even when she has no option to fully resolve uncertainty. The decisionmaker in Example 4 is resolution loving, but not information loving. She is also not information neutral.

Since the decisionmaker is not information neutral, she is inconsistent with EU. Moreover, the combination of resolution lovingness and lack of information lovingness in the example is also inconsistent with KP. In Example 4, the decisionmaker is information loving when facing any lottery between z and z', her best and worst outcomes.[15] For a KP decisionmaker who exhibits the same behavior, $v_1(z_1, \cdot)$ would have to be convex in its second argument between the values of $v_2(z_1, z')$ and $v_2(z_1, z)$. But this would mean that, unlike the decisionmaker in Example 4, she is information loving globally, so she would not strictly prefer not to learn whether the outcome is z''.

[14] It could be that only the overall desirability matters, in which case the natural specification is $K = 1$ and $v_1(z_1, z_2) = u_1(z_1, \delta_{z_2}, z_2)$. But in other situations—for instance, when a consumer consumes multiple goods and cares about disappointments separately in the different goods—disappointment preferences can take a more complex form.

[15] To see this, note that for any $p \in [0, 1]$, $pu_1(z_1, p\delta_z + (1-p)\delta_{z'}, z) + (1-p)u_1(z_1, p\delta_z + (1-p)\delta_{z'}, z') = pu_1(z_1, \delta_z, z) + (1-p)u_1(\delta_{z'}, z') - p(1-p)(u_1(z_1, \delta_z, z) + u_1(z_1, \delta_{z'}, z') - u_1(z_1, \delta_{z'}, z) - u_1(z_1, \delta_z, z'))$, which is convex in p.

Similarly to the previous subsections, these differences can be elicited from behavior by taking advantage of the fact that self 1 is known to strictly prefer $\overline{z}_1$ to $\underline{z}_1$. All of self 1's choices above remain the same if her preferred alternative comes with a small probability of getting $\underline{z}_1$ rather than $\overline{z}_1$.

The following theorem formalizes the relationship between disappointment aversion and informational preferences. Part 1 shows that when u_1 is linear in f_1—so that the source of informational preferences present in previous models is ruled out—informational preferences are about disappointment preferences. Parts 2 and 3 formalize the intuitions developed from Example 4.

Theorem 2 *Suppose u_1 is linear in f_1. Then*

1. *the agent is information neutral if and only if she is disappointment neutral.*
2. *the decisionmaker is resolution loving if and only if she is disappointment averse;*
3. *if the decisionmaker is regular disappointment averse, she is information loving.*

4 How is utility from anticipation reflected in behavior?

Section 3 has identified three observable differences between decisionmakers who derive utility from anticipation and those who do not. One reason to study these phenomena is that they are important and likely common deviations from standard models. But when preferences satisfy a small extra assumption, a version of non-satiation, these phenomena are special in another way: *all* novelties vis-á-vis previous models are related to them. More precisely, if the decisionmaker does not exhibit any of the phenomena, she is observationally equivalent to a time-consistent agent whose utility function depends only on physical outcomes. Self 2 is defined to satisfy non-satiation if for all $z_1 \in Z_1$ and $f_1 \in F_1$, the image of Z_2 under $u_2(z_1, f_1, \cdot)$ is unbounded from above. This last assumption holds in virtually all applications of interest.

Theorem 3 *Suppose self 2 satisfies non-satiation. If the decisionmaker is time consistent, information neutral, and stable, she behaves according to EU.*

The logic of the proof of Theorem 3 is the following. By Observation 1, maximizing self 1's expected utility (given the constraint that expectations have to be correct) is a continuation personal equilibrium once z_1 and d_2 are realized. Stability means (loosely) that self 2 is indifferent between all continuation equilibria. By time consistency, self 1 is also indifferent. Combined with the previous claim, this means that any continuation personal equilibrium maximizes the expectation of u_1. Finally, since self 1 is information neutral, her utility function can be written in the expected-utility form over physical outcomes.

It bears emphasizing that although stability, time consistency, and information neutrality are defined in terms of the decisionmaker's utility function, they correspond directly to observable behavior in that they rule out the phenomena in Sect. 3. That is, if any of these conditions is violated, there are circumstances under which this could be observed in behavior; and conversely, if one of the three phenomena is never displayed in behavior, the corresponding condition is satisfied. In this light, Theorem 3 provides the converse of Sect. 3, which has demonstrated three types of behavior that cannot be rationalized in an EU framework. Namely, the theorem shows that if the decisionmaker's behavior cannot be rationalized using EU, at least one of these behaviors will be observed in some decision problems.

This might be quite surprising. A reasonable first reaction to the model of this paper is that preferences over anticipation, by virtue of their inordinate complexity, lead to infinitely richer and less restrictive behavior than EU. Theorem 3 shows, in contrast, that any novel phenomenon is some combination of only three basic types of behavior. By limiting the set of new effects to be analyzed, Theorem 3 could prove useful for thinking about and organizing the additional types of behavior made possible by introducing anticipation into motivation. In any model that incorporates anticipation and purports to expand EU, one can first check whether it violates time consistency, information neutrality, or stability. If it does, one can identify which of these conditions fails, and apply what we know about the concept to facilitate the understanding of the sources of the model's new predictions.

At the same time, Theorem 3 also implies circumstances under which a model incorporating anticipation is unnecessary to make novel behavioral predictions. If a theory lacks all three phenomena in Sect. 3, any predictions it makes could have been made by a standard model as well.[16]

Of course, the three phenomena are quite broad. As a result, anticipatory utility will lead to a rich set of behavioral consequences, and much more work is needed to arrive at a detailed characterization of the three phenomena sufficient to understand these consequences. For instance, the reference-dependent models of Kőszegi and Rabin (2006) and Heidhues and Kőszegi (2005) feature a wide set of phenomena related to multiple equilibria and time inconsistency.

The next result compares my model to its other natural point of reference in the literature, the KP model. Since their model allows for informational preferences, it can of course be consistent with observed violations of information neutrality. However, Example 4's non-monotonic informational preferences due to disappointment aversion cannot be captured in their framework. To rule out such preferences, I use a separability condition between expectations and the corresponding outcomes. In addition, KP assume time consistency, so Examples 2 and 3 violate their model. The following theorem proves that if time consistency and an additive separability condition are imposed, there do not remain any behavioral differences between a model incorporating anticipation and KP.[17]

Theorem 4 *Suppose that the decisionmaker is time consistent, self 2 satisfies nonsatiation, and u_2 is additively separable in f_1 and other outcomes. Then the decisionmaker behaves according to KP.*

[16] This does not, however, necessarily mean that the model incorporating anticipation is not useful. For instance, two models may have similar or identical behavioral predictions, but radically different welfare implications. The focus of the current paper is exclusively on observable behavior.

[17] The same separability condition is sufficient to rule out both non-monotonic informational preferences and instability, even though the former is about inseperabilities between current expectations and future outcomes, while the latter is about inseperabilities between past expectations and current outcomes. The reason is that time consistency ties the two together.

Theorems 3 and 4 are tight in that they become false if any of the assumptions is removed.

5 Applications of personal equilibrium

The previous sections introduced a new model of individual decisionmaking when expectations influence preferences, and the analysis so far has concerned general properties of the model. But the theory is intended for use in economic applications, and in any specific application, more structure must be imposed on the utility functions u_1 and u_2. I now outline three possibilities for how this framework can inform central economic problems.

5.1 Reference-dependent preferences

Countless experimental studies demonstrate that preferences are reference-dependent: people do not evaluate an economic outcome based solely on how it ranks on an absolute scale, but also care about how it compares to relevant reference points. Perhaps the most important feature of such preferences is loss aversion: decisionmakers are more averse to falling short of the reference point than they are keen about achieving an equal-sized gain over it (Kahneman et al. 1990, for example). Based on a variety of evidence and a reinterpretation of classical experiments, Kőszegi and Rabin (2006) propose a model in which a person's reference point is given by her recent expectations about what outcomes she would get. A child who had been expecting a nice Christmas gift feels a loss if she gets only a mediocre one; but a seller who never expected to keep an item in her possession may not feel much of a loss from giving it up. With expectations being the reference point, Kőszegi and Rabin (2006) close the model by assuming that behavior corresponds to a personal equilibrium.

The framework of Kőszegi and Rabin (2006) is readily applicable to many economic situations. For instance, a prediction of the theory in the realm of consumer behavior is that a person's willingness to pay for a good is not fixed, but depends on the market environment and how she expects to respond to it. Specifically, her valuation is higher if she had been expecting to buy the good with greater probability, and lower if she had been expecting to acquire it at a cheaper price. Hence, the theory is a platform to study psychological phenomena behind pricing strategies and marketing and sales techniques.[18]

5.2 Emotions and self-regulation

One of the most successful and widely applied ideas in behavioral economics is the formalization of intrapersonal conflicts in intertemporal choice (Strotz 1956; Phelps and Pollak 1968; Laibson 1997), capturing the idea that people tend to overweight

[18] Heidhues and Kőszegi (2005, 2008) study the implications of the model for the pricing behavior of profit-maximizing firms selling to loss-averse consumers.

immediate pleasures relative to what is in their long-run self-interest. Yet economists have devoted little attention to the psychological foundations of such impulsive behaviors.

Psychological evidence indicates that emotions play a crucial role in impulse control linked to a broad spectrum of personal and social problems. Emotional distress tends to increase overeating in overweight people (and food consumption in normal adults), undermines attempts at smoking cessation, leads to increases in drinking alcohol, and to gambling and compulsive shopping. Emotions also regulate social behaviors such as aggression and helping others. Moreover, Tice et al. (2001) demonstrated that—at least in some settings—emotional distress changes behavior by directly affecting instantaneous preferences: if subjects believe that their mood is unchangeable, they do not show an increase in food consumption or procrastination.

This paper provides a framework to model these self-control phenomena. While emotions affect impulsive behaviors and self-control, whether people think they can focus on long-term goals also likely affects their anticipatory emotions. To illustrate some possible consequences of this interaction, consider the following simple example. Suppose $z_1 \in \mathbb{R}$ is exogenous and random, and self 2 chooses $z_2 \in [0, M]$. A higher z_2 corresponds to more forward-looking behavior by self 2.[19] Posit that $u_1(z_1, \delta_{z_2}, z_2) = z_2 + z_1$ and $u_2(z_1, \delta_{z_2}, z_2') = z_2 + z_1 + [z_2' - (z_2 + z_1)]^2$. Intuitively, if the person expects to choose z_2 in period 2, her emotions in both periods 1 and 2 are given by $z_2 + z_1$. More positive emotions lead to higher utility, and also to more forward-looking choices (as self 2 aims for an action as close as possible to $z_2 + z_1$).

Now notice that if $z_1 > 0$, the unique pure-strategy personal equilibrium is to choose $z_2 = M$, giving relatively high utility to both selves. To see this, note that if self 1 expected self 2 to choose $z_2 < M$, self 2 would want to choose something higher. But if $z_1 < 0$, the unique pure-strategy personal equilibrium is to choose $z_2 = 0$, giving possibly drastically lower utility to both selves. Hence, self-control can be fragile to even small changes in circumstances. Intuitively, once the person starts feeling bad for some external reason, she may compensate by focusing too much on short-term pleasures over long-term goals. But as she realizes that this is what she is going to do, she starts feeling even worse, exacerbating the problem. Hence, her lack of self-control can spiral out of control.

5.3 Intimidating decisions

Neoclassical utility theory (and most of economics) conceptualizes consumer decisions as a simple matter of choosing, in a detached and composed way, the best of the available options in question. In selecting a family car, for example, a mother is supposed to coldly weigh the important attributes (safety, comfort, price, style, etc.) of the options in consideration, and decide which combination is best for her. Holding a person's information and decision constant, the act of choice itself is not generally thought to affect her utility.

[19] Although the formal model has no periods beyond period 2, as mentioned above the period-2 utility function can be interpreted as the reduced form for a longer-horizon problem.

Intuition and evidence suggests, however, that many consumer choices are more emotionally involving. In the process of making a decision, a consumer has to consider the tradeoffs involved. By doing so, she is reminded of what is at stake and what could go wrong, often negatively affecting her utility. It is difficult for a mother to calmly assess the tradeoff between money and her family's safety, because this inevitably leads her to think about the possibility that her children might be in a crash. It is tough for an employee to consider the advantages of her investment options in a cool way, because this reminds her that she may end up poorer than she would hope. Healthcare, education, and career choices all involve similar unpleasant acknowledgments of conflict as well. Indeed, psychological evidence indicates that focusing on threats increases stress (Miller 1987; Neufeld 1976); and inducing such negative emotions increases the frequency with which the default option is chosen in hypothetical decisions (Luce 1998). In real choices in supermarkets (Iyengar and Lepper 2000) and retirement savings (Iyengar and Jiang 2004), increasing conflict by increasing the number of options induces people to avoid deciding altogether.[20]

Choice conflict can also be naturally modeled in the framework of this paper. While the act of making a choice may not directly affect a person's beliefs about the future, it affects the *utility* she derives from a given set of beliefs because it forces her to focus on those beliefs. Hence, the process of making a careful decision influences a person's utility from anticipation; at the same time, the quality of her decisions also affects her anticipatory utility. Thus, preferences over anticipation and physical outcomes interact, and the decisionmaker's behavior should be determined in personal equilibrium. A model of this kind might shed light on the findings mentioned above, and help in designing policies for giving consumers the right options in emotionally loaded decisions.

6 Discussion

6.1 Comments on the model and related literature

First, the model is an extension of the psychological expected utility model of CL in allowing for an interaction between expectations and later outcomes. Two further theoretical novelties follow from this. In CL, equilibrium can be obtained using standard backward induction: self 2's utility function does not depend on self 1's expectations, so her maximization problem immediately determines the set of period-1 expectations consistent with future decisionmaking. And since self 2's behavior is independent of expectations, it is not crucial to define when those expectations are realized. In the current model, behavior has to be determined jointly with expectations, giving rise to the personal-equilibrium concept. And to define personal equilibrium, I need to be explicit about the temporal placement of expectations. These features are responsible

20 In addition, choice conflict might provide a reasonable and parsimonious explanation for default effects and other striking regularities in retirement savings decisions (see for example Madrian and Shea 2001; Choi et al. 2004). If people find investment decisions intimidating and therefore try to avoid them, their outcomes will be greatly influenced by what choices they are forced to make and the rules that govern their financial assets before they make them.

for a majority of the phenomena in Sect. 3; in fact, Examples 1, 3, and 4 all generate behavior that could not happen with CL preferences.

Second, most previous models that allow for informational preferences (Epstein and Zin 1989; Grant et al. 1998, 2000; KP; Skiadas 1998, and others) approach the question from a completely different perspective than that taken in this paper. Their starting point is preferences over the timing of resolution of uncertainty, without explicitly formalizing utility from anticipation. My goal is to model the primitives that give rise, among others, to informational preferences. As mentioned above, there is plenty of evidence that anticipation influences behavior, and is often singled out as a factor in informational preferences. It therefore seems reasonable to start from these primitives.

Third, there are aspects of decisionmaking under uncertainty that are impossible to capture in this framework. For example, the decisionmaker cannot be affected by any information that would only be relevant had she made a different decision in the past. The athlete recovering from a knee injury (whose example was used to motivate the model) cannot be influenced by *regret* about not choosing an alternative treatment that she later finds could have been more effective.[21] Similarly, while Sect. 3.3 shows that utility from anticipation can lead to rich informational preferences, Eliaz and Spiegler (2006) demonstrate that even with such preferences, there are important restrictions on the information-acquisition choices agents can make.

Fourth, in an influential paper, Geanakoplos et al. (1989) introduce a multiple-player equilibrium concept (psychological equilibrium) for preferences that can depend on beliefs about others' strategies. Psychological equilibrium also imposes the internal-consistency condition that beliefs are correct, and the optimization condition that each player maximizes her belief-dependent utility. In this sense, personal equilibrium is a single-player—but multiple-self—modification of psychological equilibrium in which both selves' preferences depend on self 1's beliefs about self 2's strategy. In the dynamic version of their model, however, Geanakoplos et al. (1989) assume that a player's utility depends only on beliefs at the *beginning* of the game. Unlike personal equilibrium, therefore, their model does not allow actions to directly influence beliefs.[22]

[21] Loomes and Sugden (1982) introduce a model of choice under uncertainty in which the anticipation of regret affects decisions.

[22] Fifth, my model also incorporates as a special case the assumption that the decisionmaker cares about the *utility* of past or future selves, a feature reminiscent of intergenerational-altruism models. Closest in spirit to my model is two-sided intergenerational altruism, where each generation takes the utility of the past generation as given, and this utility in turn depends on current outcomes (Hori and Kanaya 1989; Bergstrom 1999). In contrast to mine, the main concern of these papers is not behavior—no solution concept is offered—but how the set of interdependent utility functions can be simplified. Thus, from a formal point of view, this paper can be considered an addition to the intergenerational-altruism literature: it provides a model of behavior in intergenerational models in which generations derive utility from previous generations' expectations. However, this application seems less interesting, because people are likely to care far less about past generations' expectations than about their own past expectations.

6.2 Revealed preference issues

One main subject of this paper is how extending the decisionmaker's consequence space to include anticipation is reflected in her behavior. At the same time, I do not offer a revealed-preference foundation for the enriched preferences—it is not clear to what extent the decisionmaker's utility function can be extracted from her behavior. This creates a slight tension in the paper, which the current section attempts to address.

Examples 1, 3, and 4 are meant to illustrate the behavioral manifestations of anticipation, and are therefore by definition "revealed" consequences of it. To be precise, this means that an observer who could witness the decisionmaker's behavior in a few well-chosen decisionmaking problems could notice the key aspects of behavior these examples demonstrate, and conclude that they are inconsistent with EU, KP, and CL preferences.[23]

Similarly, any failure of the key restrictions used in Theorem 3 (stability, time consistency, and information neutrality) can be revealed in behavior.

Examples 1, 2, and 3, and the separability assumption in Theorem 4, make restrictions on the interaction of self 2's preferences with self 1's expectations. With the auxiliary assumption of continuity, it is easy to extract these preferences.[24] To elicit a preference ordering between (f_1, z_2) and (f_1, z_2'), we give the agent the distribution f_1 (and leave her no choice) with probability $1 - \epsilon$, and let self 2 choose between z_2 and z_2' with probability ϵ. Letting ϵ go to zero, we obtain self 2's ordering between the two options.

However, the same method does not work to fully elicit self 1's preferences over future physical outcomes and current expectations, on which Example 4 depends. For a rational decisionmaker, self 1's expectations and the distribution of future outcomes have to be the same. Thus, we cannot gauge her preferences over streams in which outcomes diverge from expectations.[25] In fact, it is easy to give a pair of utility functions that lead to the same behavior in all decisionmaking problems: for $Z_2 = \mathbb{R}$, take $u_1(z_1, f_1, z_2) = z_2$ and $u_1(z_1, f_1, z_2) = E[f_1]$.

Although Example 4 has features that are outside the grasp of revealed preference, Sect. 3.3 shows that these features can nevertheless be used in a very intuitive characterization of *observable* novelties in informational preferences, and are thus not superfluous. Furthermore, when preferences depend on expectations, the classical revealed-preference requirement of founding all features of the theory on observed individual behavior is overly narrow and excludes discussion of econom-

[23] This holds for EU, KP, CL preferences satisfying non-satiation. Non-satiation is a reasonable assumption to maintain because allowing a utility function to be constant makes it consistent with any behavior, and renders the exercise in this paper trivial and uninteresting.

[24] This is not possible without continuity. In Example 3, for instance, self 2 prefers the stream (δ_z, ζ') over the stream (δ_z, z). Clearly, it is impossible to set up a choice experiment in which self 2 would get the opportunity to directly reveal this preference: if self 1 knew self 2 would choose between z and ζ' (with some probability), *and* self 2 actually had the preferences in Example 3, self 1 could not have expected to receive z with probability one.

[25] The reason the same problem does not arise in extracting self 2's preferences is that in period 2, the decisionmaker's past expectations and current choices can diverge, since information has been revealed since the expectations were formed.

ically interesting issues. Kőszegi (2006) shows that an informed agent might communicate differently with an uninformed principal if the principal has anticipatory emotions than if she does not, even if the principal behaves exactly the same way in all individual decisionmaking problems. If the principal has emotions, the agent attempts to convey her superior information in a way to make the principal feel better. Despite rational expectations by the principal, asymmetric information creates a divergence between her expectations and future outcomes, and the failure of revealed preference to pick up preferences for this case means that it is insufficient for the "emotional agency" problem.

The above argument does not, however, imply that one needs to fully reject the notion of revealed preference when it comes to anticipation—only that an extended version of it is necessary. Specifically, suppose that the agent has a "representative" who knows her preferences and makes decisions on her behalf. By giving choice problems to the representative instead of the agent, we can create a divergence between expectations and future outcomes, and elicit all preferences used in this paper.

7 Conclusion

The goal of this paper is to provide a framework for modeling and studying the consequences of utility from anticipation when it interacts with utility from physical outcomes. I assume a general expected-utility function over anticipation and physical outcomes, and propose a solution concept, personal equilibrium, for the determination of the decisionmaker's behavior. There are differences between standard decisionmakers and those in my model relating to time inconsistency, informational preferences, and instability. After discussing these phenomena, the paper proves that any novelty in behavior has to be related to at least one of them.

Appendix: Proofs

Proof of Theorem 1 For a given z_1, d_2, consider the correspondence that maps any $f_1 \in F_1$ to the set of optimal choices for self 2, when the history is z_1, f_1 and the decision problem is d_2. Since d_2 is compact, this correspondence is nonempty valued. Since self 2 can use mixed strategies, it is convex valued. And since u_2 is continuous, it has a closed graph. Therefore, by the Kakutani–Fan–Glicksberg theorem (Aliprantis and Border 1999, p. 550), its set of fixed points is nonempty and compact. Therefore, let $\phi^c(z_1, d_2)$ denote the set of its fixed points; $\phi^c(z_1, d_2)$ is the set of possible continuation personal equilibrium expectations given z_1 and d_2, and ϕ^c is a nonempty and compact valued correspondence.

We now prove that ϕ^c is a closed correspondence. Consider a feasible sequence $(z_1^n, d_2^n) \to (z_1, d_2)$, and suppose the sequence f_1^n satisfying $f_1^n \in \phi^c(z_1^n, d_2^n)$ has $f_1^n \to f_1$. Since d_2^n approaches d_2 in the Hausdorff metric, we must have $f_1 \in d_2$. Consider any $f_1' \in d_2$. Again since d_2^n approaches d_2 in the Hausdorff metric, there is a sequence $f_1^{n\prime}$ such that $f_1^{n\prime} \in d_2^n$ and $f_1^{n\prime} \to f_1'$. Because f_1^n is a continuation personal equilibrium given (z_1^n, d_2^n), we know that

$$E_{f_1^n} u_2(z_1, f_1^n, \cdot) \geq E_{f_1^{n\prime}} u_2(z_1, f_1^n, \cdot).$$

By continuity of u_2, this implies that

$$E_{f_1} u_2(z_1, f_1, \cdot) \geq E_{f_1'} u_2(z_1, f_1, \cdot).$$

Hence, $f_1 \in \phi^c(z_1, d_2)$, completing the proof that ϕ^c is closed.

Consider the function

$$\eta(z_1, d_2) = \max_{f_1 \in \phi^c(z_1, d_2)} E_{f_1} u_1(z_1, f_1, \cdot)$$

and the correspondence

$$\mu(z_1, d_2) = \{f_1 \in \phi^c(z_1, d_2) | E_{f_1} u_1(z_1, f_1, \cdot) = \eta(z_1, d_2)\}.$$

Intuitively, given z_1 and d_2, $\mu(z_1, d_2)$ is the set of continuation personal equilibria that are best from the point of view of self 1, and $\eta(z_1, d_2)$ is the highest expected utility self 1 can achieve in a continuation personal equilibrium.

Now since ϕ^c is a closed correspondence between compact Hausdorff spaces, it is measurable (Aliprantis and Border 1999, Theorem 17.19, p. 571). Furthermore, $E_{f_1} u_1(z_1, f_1, \cdot)$ is continuous (Aliprantis and Border 1999, Theorem 14.3, p. 476). Therefore, the measurable maximum theorem (Aliprantis and Border 1999, p. 570) applies. As a result, η is measurable and μ admits a measurable selection. Call one such selection ϕ.

Notice that since ϕ^c is closed, the function η is upper semicontinuous. By Theorem 14.5 of Aliprantis and Border (1999, p. 479), the function $\gamma \mapsto \int \eta(z_1, d_2) d\gamma(z_1, d_2)$ from $\Delta(Z_1 \times D_2)$ to $\mathbb{R}$ is upper semicontinuous. This means that if the continuation distribution of outcomes is determined by ϕ, self 1's maximization problem has a solution (Aliprantis and Border 1999, Theorem 2.40, p. 43). This establishes the existence of personal equilibrium. □

Proof of Theorem 2

Part 1. We have already established in the text that if the agent wants to receive all information about a fifty-fifty gamble over two outcomes, she is disappointment averse. This establishes the "only if" part of the theorem.

We prove the converse for each given z_1. For notational simplicity, let $u(\cdot) = u_1(z_1, \cdot)$. Suppose the agent is facing a probability measure over future outcomes $F \in \Delta(Z_2)$. If all uncertainty is resolved early, her expected utility is

$$\int u(\delta_z, z) dF(z),$$

whereas remaining completely ignorant gives an expected utility of

$$\int u(F, z) dF(z).$$

Since the agent is insecurity neutral, the above equals

$$\int u(\delta_\zeta, z)dF(\zeta)dF(z).$$

Noticing that

$$\int u(\delta_z, z)dF(z) = \int u(\delta_z, z)dF(\zeta)dF(z),$$

the difference between the full-information and full-ignorance levels of expected utility is

$$\frac{1}{2}\int [u(\delta_z, z) - u(\delta_z, \zeta) + u(\delta_\zeta, \zeta) - u(\delta_\zeta, z)]dF(\zeta)dF(z). \tag{4}$$

Since the agent is disappointment averse, the integrand is positive everywhere; this proves that the agent always wants full information over no information.

Part 2. To prove that the agent always wants more information, we prove that the difference (4) is concave in F, the probability measure over future outcomes. This is sufficient because expected utility from full information is linear in the probability measure of future outcomes the agent is facing, so it implies that expected utility is convex in F.

Since disappointment aversion is regular, we can rewrite (4) as

$$\frac{1}{2}\int \sum_{k=1}^{K} |v_k(z) - v_k(\zeta)|dF(z)dF(\zeta).$$

We will prove that

$$\int |v_k(z) - v_k(\zeta)|dF(z)dF(\zeta) \tag{5}$$

is concave in F for each k. To do so, we give a geometric interpretation to the integral (5). Note that through the map v_k, points in Z_2 map to points on the real line. Then, expression (5) is the expected distance of two points independently chosen according to the distribution G, the probability measure F induces on $\mathbb{R}$ through the map v_k. We prove that this distance is concave in G.

Consider two cumulative distribution functions G_1 and G_2 on the reals, as well as a convex combination $\lambda G_1 + (1-\lambda)G_2$ of them. Take any point p on the real line. The probability that p is on a line segment between two points randomly chosen according to G_1 is $2G_1(p)(1 - G_1(p))$. The same probability for G_2 is $2G_2(p)(1 - G_2(p))$, and for the convex combination it is $2(\lambda G_1(p) + (1-\lambda)G_2(p))(1-\lambda G_1(p) - (1-\lambda) G_2(p))$. To complete the proof, it is sufficient to prove that

$$\begin{aligned}&\lambda G_1(p)(1 - G_1(p)) + (1-\lambda)G_2(p)(1 - G_2(p))\\ &\quad \le (\lambda G_1(p) + (1-\lambda)G_2(p))(1 - \lambda G_1(p) - (1-\lambda)G_2(p)).\end{aligned}$$

This is equivalent to

$$0 \leq \lambda(1-\lambda)(G_1(p) - G_2(p))^2,$$

which holds for any p. □

Proof of Theorem 3 Let

$$v(z_1, z_2) = u_1(z_1, \delta_{z_2}, z_2).$$

We first use non-satiation and time consistency to show that a modified definition of time consistency holds as well, in which the strict inequalities are replaced by weak ones. Suppose by contradiction that

$$E_{f_1'} u_2(z_1, f_1, \cdot) \geq E_{f_1} u_2(z_1, f_1, \cdot)$$

but

$$E_{f_1'} u_1(z_1, f_1', \cdot) < E_{f_1} u_1(z_1, f_1, \cdot). \tag{6}$$

Then, the first inequality must in fact be an equality, otherwise time consistency would be violated. Therefore, by non-satiation and continuity, there is an f_1'' such that $E_{f_1''} u_2(z_1, f_1, \cdot) > E_{f_1} u_2(z_1, f_1, \cdot)$ but $E_{f_1''} u_1(z_1, f_1'', \cdot) < E_{f_1} u_1(z_1, f_1, \cdot)$. This contradicts time consistency.

Now we proceed to proving the statement of the theorem. First, by Observation 1, for any z_1 and d_2 there is a continuation personal equilibrium in which self 1 expects and self 2 chooses the f_1^* that solves $\max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot)$. That is, there is a personal equilibrium that maximizes self 1's utility subject to the constraint that she has to be rational.

Next, we prove that for any z_1 and d_2, all continuation personal equilibria give self 2 the same expected utility. Suppose that f_1 and f_1' are the period 1 expectations in two continuation equilibria. Then by stability,

$$E_{f_1} u_2(z_1, f_1, \cdot) = E_{f_1'} u_2(z_1, f_1, \cdot) \text{ and } E_{f_1'} u_2(z_1, f_1', \cdot) = E_{f_1} u_2(z_1, f_1', \cdot). \tag{7}$$

Now by the above definition of time consistency, the first part of Eq. (7) implies

$$E_{f_1} u_1(z_1, f_1, \cdot) \leq E_{f_1'} u_1(z_1, f_1', \cdot),$$

while the second part implies the same inequality the other way. Therefore, all personal equilibria give self 1 the same expected utility.

From the above two claims, we can conclude that *all* continuation personal equilibria have the property that they maximize self 1's expected utility; that is, they solve $\max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot)$. Conversely, by Observation 1, any f_1^* that solves this problem is a continuation personal equilibrium of outcomes. This implies that the set of personal equilibrium outcomes and the outcomes generated by the solution to

$$\max_{\gamma \in d_1} E_\gamma \left[\max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot) \right] \tag{8}$$

are identical. By information neutrality,

$$E_{f_1} u_1(z_1, f_1, \cdot) = \int u_1(z_1, \delta_{z_2}, z_2) df_1(z_2) = E_{f_1} v(z_1, \cdot).$$

completing the proof. □

Proof of Theorem 4 Let

$$v_1(z_1, d_2) = \max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot).$$

Since time consistency and non-satiation are satisfied, we know from the proof of Theorem 3 that the alternative definition of time consistency (in which the strict inequalities are replaced with weak ones) holds. Also, as in the proof of that theorem, for any z_1 and d_2, there is a continuation personal equilibrium that maximizes self 1's expected utility constrained by rationality. Now, since self 2's utility is separable from past expectations, if f_1 and f_1' are period 1 expectations in two continuation equilibria, the equalities in (7) are automatically satisfied. By time consistency, all continuation equilibria therefore give self 1 the same expected utility.

Therefore, just as in Theorem 3, we conclude that the set of personal equilibrium strategies and the outcomes generated by the solution to (8) are identical. Also, the set of self 1's personal equilibrium choices coincides with the solutions to $\max_{\gamma \in d_1} v_1(\cdot)$. Now since self 2's utility is separable from f_1, there is a function $v_2 : Z_1 \times Z_2 \to \mathbb{R}$ such that for any $z_1 \in Z_1$ and $d_2 \in D_2$, self 2's personal equilibrium choices coincide with the maximizers of the expectation of v_2.

By time consistency

$$E_{f_1} u_1(z_1, f_1, \cdot) \geq E_{f_1'} u_1(z_1, f_1', \cdot) \iff E_{f_1} v_2(z_1, \cdot) \geq E_{f_1'} v_2(z_1, \cdot)$$

for any f_1, f_1'. This implies that

$$\begin{aligned} &\max_{f_1 \in d_2} E_{f_1} u_1(z_1, f_1, \cdot) \geq \max_{f_1 \in d_2'} E_{f_1} u_1(z_1, f_1, \cdot) \\ &\iff \max_{f_1 \in d_2} E_{f_1} v_2(z_1, \cdot) \geq \max_{f_1 \in d_2'} E_{f_1} v_2(z_1, \cdot) \end{aligned}$$

or equivalently

$$v_1(z_1, d_2) \geq v_1(z_1, d_2') \iff \max_{f_1 \in d_2} E_{f_1} v_2(z_1, \cdot) \geq \max_{f_1 \in d_2'} E_{f_1} v_2(z_1, \cdot).$$

This completes the proof. □

References

Akerlof, G., Dickens, W.T.: The economic consequences of cognitive dissonance. Am Econ Rev **72**(3), 307–319 (1982)

Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis: A Hitchhiker's Guide, 2nd edn. Berlin: Springer (1999)

Bénabou, R.: Groupthink: collective delusions in organizations and markets. Working Paper, Princeton University (2008)

Bénabou, R., Tirole, J.: Self-confidence and personal motivation. Q J Econ **117**(3), 871–915 (2002)

Bergstrom, T.C.: Systems of benevolent utility functions. J Public Econ Theory **1**(1), 71–100 (1999)

Brunnermeier, M., Parker, J.: Optimal expectations. Am Econ Rev **95**(4), 1092–1118 (2005)

Caplin, A., Leahy, J.: Psychological expected utility and anticipatory feelings. Q J Econ **116**(1), 55–80 (2001)

Carrillo, J.: Self-control, moderate consumption, and craving. CEPR Discussion Paper # 2017 (1998)

Carrillo, J., Mariotti, T.: Strategic ignorance as a self-disciplining device. Rev Econ Stud **67**, 529–544 (2000)

Chew, S.H., Epstein, L.G.: The structure of preferences and attitudes towards the timing of the resolution of uncertainty. Int Econ Rev **30**(1), 103–117 (1989)

Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A.: For better or for worse: default effects and 401(k) savings behavior. In: Wise, D. (ed.) Perspectives in the Economics of Aging, pp. 81–121. Chicago: University of Chicago Press (2004)

Daido, K., Itoh, H.: The pygmalion effect: an agency model with reference-dependent preferences. CESifo Working Paper Series No. 1444 (2005)

Dillenberger, D.: Preferences for one-shot resolution of uncertainty and allais-type behavior. Working Paper, University of Pennsylvania (2008)

Eliaz, K., Spiegler, R.: Can anticipatory feelings explain anomalous attitudes to information?. Games Econ Behav **56**, 87–104 (2006)

Epstein, L.G., Zin, S.E.: Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica **57**(4), 937–969 (1989)

Geanakoplos, J., Pearce, D., Stacchetti, E.: Psychological games and sequential rationality. Games Econ Behav **1**(1), 60–79 (1989)

Gervais, S., Odean, T.: Learning to be overconfident. Rev Financial Stud **14**(1), 1–27 (2001)

Grant, S., Kajii, A., Polak, B.: Intrinsic preference for information. J Econ Theory **83**, 233–259 (1998)

Grant, S., Kajii, A., Polak, B.: Preference for information and dynamic consistency. Theory Decis **48**, 263–286 (2000)

Heidhues, P., Kőszegi, B.: The impact of consumer loss aversion on pricing. CEPR Discussion Paper 4849 (2005)

Heidhues, P., Kőszegi, B.: Competition and price variation when consumers are loss averse. Am Econ Rev **98**(4), 1245–1268 (2008)

Hori, H., Kanaya, S.: Utility functionals with nonpaternalistic intergenerational altruism. J Econ Theory **49**(2), 241–265 (1989)

Iyengar, S.S., Jiang, W.: Choosing not to choose: the effect of more choices in retirement savings decisions. Working Paper (2004)

Iyengar, S.S., Lepper, M.R.: When choice is demotivating: can one desire too much of a good thing? J Personal Soc Psychol **79**(6), 995–1006 (2000)

Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979)

Kahneman, D., Knetsch, J.L., Thaler, R.H.: Experimental tests of the endowment effect and the coase theorem. J Political Econ **98**(6), 1325–1348 (1990)

Kőszegi, B.: Anticipation in observable behavior. http://emlab.berkeley.edu/users/botond/feelings.pdf (2003)

Kőszegi, B.: Emotional agency. Q J Econ **121**(1), 121–155 (2006)

Kőszegi, B., Rabin, M.: A model of reference-dependent preferences. Q J Econ **121**(4), 1133–1166 (2006)

Kőszegi, B., Rabin, M.: Reference-dependent risk attitudes. Am Econ Rev **97**(4), 1047–1073 (2007)

Kőszegi, B., Rabin, M.: Reference-dependent consumption plans. Am Econ Rev (2009, in press)

Kreps, D.M., Porteus, E.L.: Temporal resolution of uncertainty and dynamic choice theory. Econometrica **46**(1), 185–200 (1978)

Laibson, D.: Golden eggs and hyperbolic discounting. Q J Econ **112**(2), 443–477 (1997)

Landier, A.: Wishful thinking and beliefs dynamics. Working Paper, New York University (1999)
Loewenstein, G.: Anticipation and the valuation of delayed consumption. Econ J **97**, 666–684 (1987)
Loomes, G., Sugden, R.: Regret theory: an alternative theory of rational choice under uncertainty. Econ J **92**(368), 805–824 (1982)
Luce, M.F.: Choosing to avoid: coping with negatively emotion-laden consumer decisions. J Consumer Res **24**(4), 409–433 (1998)
Madrian, B.C., Shea, D.F.: The power of suggestion: inertia in 401(k) participation and savings behavior. Q J Econ **116**(4), 1149–1187 (2001)
Manove, M., Padilla, A.J.: Banking (conservatively) with optimists. Rand J Econ **30**(2), 324–350 (1999)
Masatlioglu, Y., Nakajima, D.: Choice by constraint elimination. Working Paper, University of Michigan (2007)
Miller, S.M.: Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Personal Soc Psychol **52**(2), 345–353 (1987)
Neufeld, R.W.: Evidence of stress as a function of experimentally altered appraisal of stimulus aversiveness and coping adequacy. J Personal Soc Psychol **33**(5), 632–646 (1976)
Phelps, E.S., Pollak, R.A.: On second-best national saving and game-equilibrium growth. Rev Econ Stud **35**, 185–199 (1968)
Rabin, M., Schrag, J.: First impressions matter: a model of confirmatory bias. Q J Econ **114**(1), 37–82 (1999)
Skiadas, C.: Recursive utility and preferences for information. Econ Theory **12**, 293–312 (1998)
Stone, R.: Loss aversion and self-control. Working Paper (2005)
Strotz, R.H.: Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud **23**, 165–180 (1956)
Tice, D.M., Bratslavsky, E., Baumeister, R.F.: Emotional distress regulation takes precedence over impulse control: if you feel bad, do it! J Personal Soc Psychol **80**(1), 53–67 (2001)