Anticipation in Observable Behavior

Botond Kőszegi, UC Berkeley*

February 2003

Abstract

This paper asks the natural question: how is utility from anticipation reflected in behavior? I consider a general model of decisionmaking where rationally formed anticipation enters the agent's utility function in addition to physical outcomes, and allow for interactions between these two payoff components. The paper explores three types of behavior made possible by utility from anticipation, and proves that if a decisionmaker who cares about anticipation is distinguishable from one who only cares about physical outcomes, she *has to* exhibit at least one of these phenomena. First, the agent can display informational preferences because she is not indifferent to insecurity, or because she cares about future disappointments. I prove that an agent who is indifferent to insecurity always prefers full to partial information if and only if she is disappointment averse, but a stronger condition is needed for her to prefer more information to less. Second, the agent can be time inconsistent because anticipatory feelings pass by the time she has to "invest" in them, and this time inconsistency can be reflected in intransitivity of choices. Third, the agent can be prone to self-fulfilling expectations. In developing these ideas, I also deal with several modeling difficulties when expectations enter utility.

Keywords: anticipation, time inconsistency, disappointment aversion, non-expected utility.

*Address: UC Berkeley Dept. of Economics, 549 Evans Hall #3880, Berkeley, CA 94720. Email: botond@econ.berkeley.edu.

1 Introduction

After being customary ingredients in economic discussions of intertemporal individual decisionmaking, anticipation and related feelings had gone out of favor with the advent of mathematical models of consumption choices over time (Loewenstein 1992). Though researchers acknowledged the role of psychological factors in affecting *parameter values* in models (Fisher 1930), they believed that including them in the theory itself would make no difference (Samuelson 1939, page 291). By bringing anticipation back into the analysis, several recent papers have challenged this view, recognizing that decisionmakers who care about anticipation can be quite different from those who care exclusively about consumption. Unlike "standard" agents, they can exhibit an intrinsic preference for information, and have a specific tendency toward time inconsistency (Loewenstein 1987, Caplin and Leahy 2001).

The current paper builds on and expands this new work in two ways. By considering a more general framework than previous authors, it identifies a number of additional behavioral consequences of anticipation. More importantly, it also explores the flip side of the question: it identifies conditions under which agents who derive utility from anticipation do look the same as standard ones. I start with a general model of individual decisionmaking where anticipation enters the agent's utility function in addition to physical outcomes. In each of a finite number of periods t, a physical outcome z_t and an anticipation f_t are realized. The agent's choices give her control over physical outcomes, but in general she cannot directly select from possible expectations. Instead, f_t is a rationally formed expectation about the future, and thus is partly in the hands of future selves. Self t maximizes the expectation of a utility function u_t defined over the entire stream of physical outcomes and anticipations, taking into account the physical constraints of the problem, the behavior of future selves, and that anticipation is formed rationally. I call the equilibrium that combines these features personal equilibrium, and show that personal equilibrium exists under general conditions.

Given the goals of the paper, two aspects in which this model extends previous work–most notably Caplin and Leahy (2001)–are crucial. First, the model is explicit about the temporal placement of expectations, formalizing the notion that the pleasure or pain of anticipation is irreversibly realized before the event that precipitates it. Second, the model allows for many kinds of interaction between anticipation and physical outcomes, highlighting that not only the *imme-*

diate effect of anticipation is important, but also its effect on preferences over other outcomes. Disappointment is an obvious example of such a preference.

With the above machinery in hand, I explore three major phenomena through which one can distinguish decisionmakers in my model from those inhabiting two natural comparison worlds: the expected utility (henceforth EU) and Kreps-Porteus (1978, henceforth KP) models. Then I prove the most important result of the paper, that *any* difference of my model from these alternatives has to be related to at least one of the three phenomena.

First, for two distinct reasons, the decisionmaker might in personal equilibrium exhibit a preference for information.[1] One reason for self t to do so is that u_t could be non-linear in the expectations f_t. Intuitively, the agent might not be indifferent to the hedonic experience that results from having to live with insecurity about what will happen to her. This is the natural analogue of informational preferences in previous work (KP and others), where they are determined by the shape of the utility function in an anticipatory component (such as self $t+1$'s expected utility). However, in my model informational preferences can arise for a completely different, as of yet unexplored reason. To illustrate the idea, suppose there are two periods and that either something "good" or something "bad" will happen to the agent in period 2, with equal probability. If the agent's beliefs remain uncertain, she will be disappointed with probability one-half and get a pleasant surprise with probability one-half. Therefore, if she dislikes future disappointments more than she likes future pleasant surprises–she is disappointment averse–she would choose to find out the outcome, thereby averting both possible good news and possible bad news. Section 3.3 formally defines disappointment aversion and shows that it can generate informational preferences that are observationally distinguishable from the KP model (and its extensions). It also fully characterizes the relationship of disappointment aversion and informational preferences: the agent prefers early resolution if and only if she is disappointment averse, but her disappointment aversion has to satisfy an intuitive regularity condition for her to prefer more information to less.

Second, although time inconsistency itself is not a distinguishing feature of utility from anticipation, some of its consequences are. Consider an academic who could do three things next summer: work, relax at home, or take a holiday. Having anticipated work, she might be too stressed and

[1] Informational preferences can also arise in models where the decisionmaker cares only about physical outcomes, as long as the preferences are time-inconsistent. The reason is that information acquisition serves a strategic purpose, manipulating the actions of future selves (Carillo 1997, Carillo and Mariotti 2000, Bénabou and Tirole 2002). With anticipation, informational preferences can arise even in the absence of strategic considerations.

prefer to stay at home, and she might prefer a holiday to relaxing at home. But having anticipated a vacation, a lot of its utility has passed, so a time inconsistent preference to work instead might arise. This circularity results in intransitive choices, and, given rational expectations, generically at least one of her own three choices above is suboptimal from the point of view of the future self. In contrast to standard decisionmakers, the failure to take the utility-maximizing action can happen even if all of the agent's temporal selves have *exactly the same* utility function over the entire stream of expectations and physical outcomes.

Interestingly, identifying the above as well as other effects of (what intuitively seems) time inconsistency raises a new difficulty. As Section 3.2 argues, standard definitions of time consistency fail to extend into this framework. Thus, I offer a generalized definition and use this definition in the formal discussion.

Third, Section 3.1 identifies an additional behavioral manifestation of anticipation, *utility-relevant unpredictability*. Since current expectations depend on future actions while future preferences can also depend on current expectations, the decisionmaker could be trapped in a feedback loop that makes expectations self-fulfilling. For example, if the agent is pessimistic about her future life, she may become disinterested and lethargic, and prefer not to make costly investments like exercising or studying. As a result, she fulfills her pessimistic expectations. But if she is optimistic, she has the energy to do the right things, and once again she fulfills her expectations. Although the agent might not be indifferent between these two personal equilibria, there is no reason to expect that she would be able to coordinate on the one with higher expected utility.

In the key section of the paper (Section 4), I compare my model to the two benchmarks, the EU and KP models. I prove that if a given model incorporating anticipation does not feature at least one of the above phenomena, it does not feature *anything* novel relative to expected utility maximization. That is, if informational preferences, utility-relevant unpredictability, and time inconsistency are ruled out, there is a utility function defined only over physical outcomes such that in every decision problem the agent behaves as if she was maximizing the expectation of that utility function. Similarly, any difference between my model and KP has to do with preferences over disappointment, unpredictability, or time inconsistency. These results provide a partial answer to the question posed at the beginning: how anticipation is reflected in observable behavior.

The paper is organized as follows. Section 2 introduces notation and the concept of personal equilibrium, and deals with existence. Section 3 illustrates the kinds of behavior an "emotional" agent might display. This sets up the observational equivalence results (Section 4). Section 5.1 offers comments on the model, and discusses related literature. Section 5.2 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

2.1 Informal Discussion

This section formulates a mathematical model of decisionmaking when one of the motivating forces behind behavior is anticipation. Although the definition of the agent's preferences and decision problem is notationally heavy, the basic idea is quite intuitive and amenable to a verbal description. To aid in this description as well as that of the formal setup, consider the following four-period model of a simple purchase decision:

> **Example 0** Take the decision problem illustrated in Figure 1. Numbered dots represent decision nodes, while other branches represent nature's moves, each occuring with probability one-half. In period 1, the agent's choice set is degenerate, but she finds out whether her finances are sound (wealth W_H) or poor (wealth W_L). In period 2, she can either visit a ticket agency to consider attending the weekend Broadway show, or she can choose not to do so. In either case, she finds out the price (p_H or p_L). If she is at the agency in period 3, and can afford the ticket (nodes 6, 7, and 10), she has three options: buy it and read reviews about the show, just buy it, or pass on it. Finally, in period 4, the agent again makes no decision, but in case she decided to attend the show without researching it, she now finds out how fun it is (l or h). In period 4, she consumes her leftover money and 0, l or h amount of theater pleasure as shown.

Although simple, and although the only non-trivial physical outcomes occur in period 4, this decision problem can generate a multitude of experiences related to anticipation. Upon learning that her finances are doing well, self 1 might derive pleasure from anticipating the show. Self 2, in case she decides not to visit the agency, cannot look forward to the performance, and having anticipated it in period 1 can make her feel extra bad about this. Self 3, if she has expected to buy at a low price, might find it aversive to buy at the high price at node 7. In addition, if self 3 decides to go to the show, she may be curious to find out about it immediately, or she might want to leave some surprises for period 4. Below, I relate these feelings to formal elements of my model.

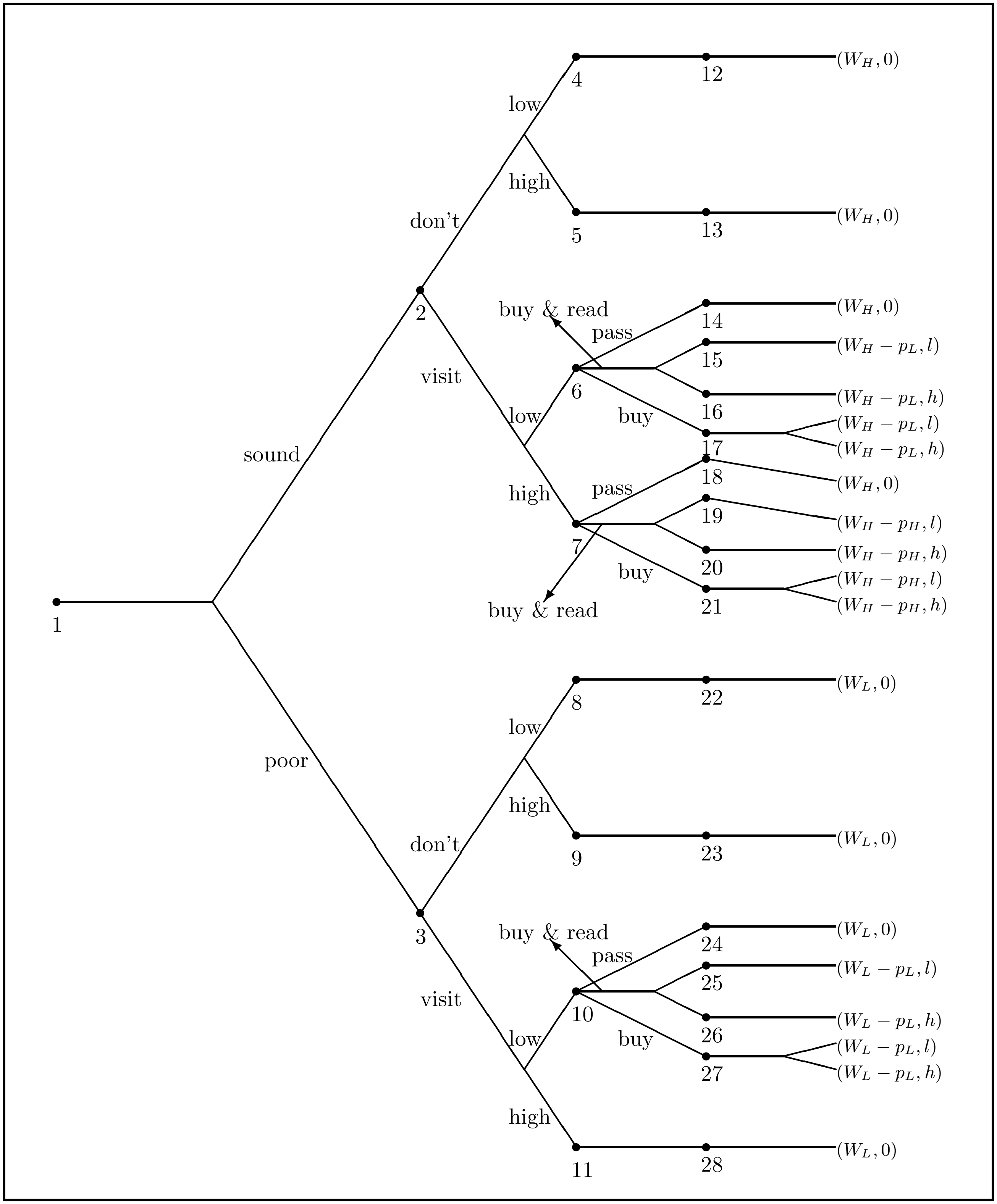

Figure 1: A Simple Purchase Decision

There are three essential aspects of anticipation that the formal model attempts to capture. First, naturally, anticipation is *forward-looking*. Though self 1's pleasure from anticipating the show is immediate, it is about something that will happen later. Realistically, I do not assume that later selves have to comply with self 1's anticipation to attend the show. But even if they do not, at that point it is impossible to change the expectations and the fact that they affected utility. This *irreversibility* is the second important feature of anticipation. Third, the agent only makes choices over physical outcomes (e.g. whether to buy or not), and she has *no direct control* over her expectations. I assume that expectations are formed rationally, so they are determined in equilibrium by the behavior of future selves. Thus, *in equilibrium*, if self 1 expects to go to the show for sure, this expectation will actually be realized. Nevertheless, the fact that expectations need not be correct off the equilibrium path will be crucial for behavior.

2.2 Rigorous Setup

The decisionmaker is involved in a T-period decision problem. In each period $t \in \{1, \ldots, T\}$, a "physical outcome" $z_t \in Z_t$ and an "anticipation" $f_t \in F_t$ are irreversibly realized. The physical outcomes are standard. In Example 0, the agent's enjoyment of the show and money are physical outcomes in period 4 (Z_4).[2] I assume that each Z_t is a Polish (complete separable metric) space; this ensures that all constructions below lead to Polish spaces in the appropriate topologies. The crucial (and nonstandard) element of the model is the definition of the expectations f_t. I first define the spaces F_t and the agent's preferences. Then, since the agent is not choosing f_t, I need to describe how it is *formed*.

Denote by $\Delta(S)$ the space of Borel probability measures over a Polish space S (endowed with the weak-* topology). Since expectations are about the future, I simply define F_t recursively from the back by

$$F_t = \Delta(Z_{t+1} \times F_{t+1} \times Z_{t+2} \times F_{t+2} \times \cdots \times Z_T).$$

For notational convenience, I will suppress arguments that are deterministic and do not depend on the agent's choice; e.g. z_1, z_2, and z_3 in Example 0. Thus, self 3's expectations are lotteries over money-show consumption pairs; self 2, in turn, anticipates both consumption in period 4 and

[2] In the example, the only relevant physical outcomes occur in period 4. Formally, this is captured by making Z_1, Z_2, and Z_3 all singletons.

expectations in period 3. For any outcome z, let δ_z denote the probability measure that assigns unit mass to z.

The expectations and physical outcomes constitute all the payoffs the decisionmaker cares about. In line with previous research, I assume that for each t, 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 $Z_1 \times F_1 \times Z_2 \times F_2 \times \cdots \times Z_T$ (with generic element $(z_1, f_1, z_2, f_2, \ldots, z_T)$). I assume that u_t is continuous for each t. 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. There may also be normative reasons for expected utility.[3]

The above fully defines the agent's preferences. Let $Y^t = \prod_{s=1}^{t-1}(Z_s \times F_s)$ (with generic element y^t) be the set of all histories up to period t. Conveniently, given y^t, self t's preferences can be thought of as being over F_{t-1}–this space is not only the space of possible expectations in period $t-1$, it is also the set of probability measures over outcomes starting in period t. Throughout the paper, for any measurable function h and probability measure H on h's domain, the shorthand $E_H h(\cdot)$ denotes $\int h(x)dH(x)$. Using this notation, given y^t, self t prefers the measure $l_{t-1} \in F_{t-1}$ over $l'_{t-1} \in F_{t-1}$ if and only if $E_{l_{t-1}} u_t(y^t, \cdot) \geq E_{l'_{t-1}} u_t(y^t, \cdot)$.

With preferences now given, I turn to the formation of expectations. In short, I assume that the agent has correct expectations. But correct expectations are determined only in equilibrium. Therefore, anticipation is part of the specification of equilibrium, for which I first have to define the nature of the agent's decision problem.

In a specific decision problem, the set of feasible physical outcomes in period t is a compact set $Z_t^0 \subset Z_t$. Let D_T^0 be the set of compact subsets of $\Delta(Z_T^0)$. A decision problem in period T is an element $d_T^0 \in D_T^0$. Going backwards, for each $t = 1, \ldots, T-1$, let D_t^0 be the set of compact subsets of $\Delta(Z_t^0 \times D_{t+1}^0)$ (endowed with the topology generated by the Hausdorff metric). A decision

[3] Caplin and Leahy (2001) argue that observed violations of the substitution axiom are not due to the axiom's lack of intuitive appeal, but arise instead from an incomplete specification of preferences. The appeal of the axiom comes from a non-complementarity argument: that once a state is realized, the decisionmaker should not care about states that did not happen. When the agent has unspecified preferences–such as anticipatory feelings–that can depend on unrealized states, the failure to include these preferences in the model shows up as a violation of the substitution axiom for preferences that are included. But once the domain of preferences is fully specified, the intuitive appeal of the axiom comes back in full force. For expectations, given what the agent expects now and what she expected in the past, her preferences should not depend on what else she could have expected today.

problem at time t is some $d_t^0 \in D_t^0$.[4] A decision at time t is an element of d_t^0; thus, a decision induces a lottery over pairs of physical outcomes in that period and future decision problems. When represented as a decision tree, each decision problem corresponds to a decision node, with each outgoing branch being a possible decision. In Example 0, self 4 has no choice, so d_4^0 is always trivial. But it is not always the same singleton set. At node 17, d_4^0 is a singleton with the gamble between $z_4 = (W_H - p_L, h)$ and $z_4' = (W_H - p_L, l)$. But at node 16, d_4^0 has the riskless outcome z_4. At node 6, self 3's problem is to choose between node 14, node 17, and a lottery over nodes 15 and 16. Continuing this way, one can easily construct the other selves' decision problems. This construction allows one to capture decisions regarding information acquisition. Self 3's last two options give the same lottery over Z_4, but with information revealed at different times. For each t, let D_t be the union all possible period t decision problems.

Define expectations F_t^0 and histories Y^{0t} from the feasible Z_t^0's as above. Now, the elements of a personal equilibrium are:

1. A vector $\sigma = (\sigma_1, \ldots, \sigma_T)$ of strategies for each self, where

$$\sigma_t : Y^{0t} \times D_t^0 \to \Delta(Z_t^0 \times D_{t+1}^0)$$

   and $\sigma_t(y^t, d_t^0) \in d_t^0 \ \forall t, \forall y^t \in Y^{0t}, \forall d_t^0$.

2. A vector $\phi = (\phi_1, \ldots, \phi_{T-1})$ of measurable "anticipation functions" in each period, where

$$\phi_t : Y^{0t} \times Z_t^0 \times D_{t+1}^0 \to F_t^0.$$

The first of these elements is the usual vector of strategies of the decisionmakers in the game. The new feature of personal equilibrium is the second element. ϕ_t defines the agent's expectations as a function of what happened up to the end of period t. In Example 0, ϕ_2 specifies, as a function of the history and the decision node self 3 is going to face, what self 2 expects. Just before node 6 (with, say, history y^2 and decision problem d_3^0) it might specify that the agent expects to be at the show without finding out how fun it is until period 4; that is,

$$\phi_2(y^2, d_3^0) = \frac{1}{2}\delta_{(\frac{1}{2}\delta_{z_4}+\frac{1}{2}\delta_{z_4'}, z_4)} + \frac{1}{2}\delta_{(\frac{1}{2}\delta_{z_4}+\frac{1}{2}\delta_{z_4'}, z_4')}, \tag{1}$$

[4] This specification of the agent's decision problem is identical to that of Kreps and Porteus (1978).

since self 2's anticipation includes self 3's expectations as well. Since self t does not choose the anticipation function, she is in general unable to select from all expectations that are possible given her decision problem, even though expectations are realized at time t just as physical outcomes are. Although the agent's problem allows for many expectations in period 1, self 1 has absolutely no control over them. This makes ϕ_t a qualitatively different part of the proposed personal equilibrium.

Note that the specification of ϕ_t embeds an assumption about *timing*: f_t is realized after z_t and d^0_{t+1}, as illustrated in Figure 2. Alternative assumptions about timing would make no difference to the results.

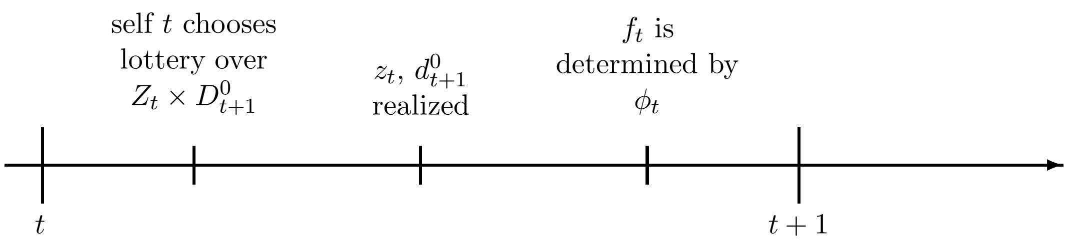

Figure 2: Sequencing of decisions and outcomes within one period

Finally, one more piece of notation is necessary to define the concept of personal equilibrium. Let $l_t(y^t, \gamma_t, \phi_t)$ represent the probability measure that the lottery $\gamma_t \in \Delta(Z^0_t \times D^0_{t+1})$ and the map ϕ_t induce on the vector of outcomes starting in period t, where both f_t and the actual probability measure over future outcomes are given by ϕ_t. Thus, $l_t(y^t, \gamma_t, \phi_t)$ is the probability measure over outcomes self t can expect when she chooses γ_t, her anticipation is given by ϕ_t, and these expectations are realized. In Example 0, if (given history y^3) γ_3 deterministically leads to the decision problem d^0_4, and $\phi_3(y^3, d^0_4) = \frac{1}{2}\delta_{z_4} + \frac{1}{2}\delta_{z'_4}$, then $l_3(y^3, \gamma_3, \phi_3) = \frac{1}{2}\delta_{(\frac{1}{2}\delta_{z_4}+\frac{1}{2}\delta_{z'_4}, z_4)} + \frac{1}{2}\delta_{(\frac{1}{2}\delta_{z_4}+\frac{1}{2}\delta_{z'_4}, z'_4)}$. It is sufficient to focus on such measures over outcomes because in equilibrium the expectations will be required to be correct.

Definition 1 *The profile (σ, ϕ) constitutes a* *personal equilibrium* *if*

1. *(***Optimization***) For all t, all histories y^t, and all decision problems d^0_t,*

$$\sigma_t(y^t, d^0_t) \in argmax_{\gamma_t \in d^0_t} E_{l_t(y^t,\gamma_t,\phi_t)} u_t(y^t, \cdot).$$

2. (**Internal Consistency**) *For all $t < T$, all y^t, z_t, and d^0_{t+1}, we have*

$$\phi_t(y^t, z_t, d^0_{t+1}) = l_{t+1}\left((y^t, z_t, \phi_t(y^t, z_t, d^0_{t+1})), \sigma_{t+1}(y^t, z_t, \phi_t(y^t, z_t, d^0_{t+1}), d^0_{t+1}), \phi_{t+1}\right).$$

Given y^{t_0}, $(\sigma_{t_0}, \ldots, \sigma_T), (\phi_{t_0}, \ldots, \phi_{T-1})$ constitutes a continuation personal equilibrium if conditions 1 and 2 are satisfied for $t \geq t_0$. Given y^{t_0}, z_{t_0}, $(\sigma_{t_0+1}, \ldots, \sigma_T), (\phi_{t_0}, \ldots, \phi_{T-1})$ constitutes a continuation personal equilibrium if condition 1 is satisfied for $t \geq t_0+1$ and condition 2 is satisfied for $t \geq t_0$.

Personal equilibrium can be viewed as an extension of subgame-perfect equilibrium to a preference structure incorporating anticipation. The assumption of optimization at each node is analogous to that in subgame-perfect equilibrium. In personal equilibrium, an additional assumption about the formation of expectations is needed. Namely, given the expectations, it has to be a continuation personal equilibrium to realize the expectations. Returning to the previous example of node 6 in Example 0, suppose that $\phi_2(y^2, d^0_3)$ is given by expression 1. Then, letting d^0_4 be the decision problem at node 17, it has to be the case that $\sigma_3(y^2, \phi_2(y^2, d^0_3), d^0_3) = \delta_{d^0_4}$.[5] In other words, self 3 has to prefer buying the ticket without reading reviews after having expected $\phi_2(y^2, d^0_3)$. A similar internal consistency condition needs to be checked at every decision node.

Section 5.1 offers comments on the model, and discusses related literature. Before I start analyzing the types of behavior made possible by utility from anticipation, I give general conditions under which personal equilibrium exists. As in many other models, it is necessary to allow for mixed strategies to guarantee the existence of personal equilibrium.[6] It turns out that this is not sufficient. Appendix A gives an example in which equilibrium does not exist, and uses the example to motivate an additional assumption: allowing for randomization between equilibria. With that assumption, existence of personal equilibrium is guaranteed:

Theorem 1 *If we allow for randomization between equilibria, a personal equilibrium exists.*

For a formalization of the notion of randomization between equilibria, the proof of Theorem 1, as well as further discussion, see Appendix A.

[5] Recall that z_3, being trivial, is suppressed in the notation.

[6] To see this, one can observe that Example 2 in Section 3.2 does not have an equilibrium when self 2's decision problem is $\{\delta_z, \delta_{z'}, \delta_{z''}\}$.

3 Novel Properties of Preferences with Anticipation

This section formally introduces the range of behaviors that are specific to decisionmakers who care about anticipation. In each case, I start by giving an example in which some aspect of the agent's behavior is inconsistent with that of an EU or KP decisionmaker satisfying non-satiation.[7] I then attempt to pinpoint the root of the novel behavior in question, and deal with the complications that arise. In the course of each discussion, I introduce a restriction to be used in the characterization theorems, one that rules out the particular behavior.

3.1 Self-Fulfilling Expectations

Consider the following simple example with two periods.

Example 1 At $t = 1$, choice is degenerate and (z_1, d_2^0) is deterministic. At $t = 2$, there are two actions, z and z', to choose from. Suppressing z_1 in the utility function,[8] make the following assumptions:

$$\begin{aligned} u_2(\delta_z, z) &> u_2(\delta_z, z'), \text{ and} \\ u_2(\delta_{z'}, z') &> u_2(\delta_{z'}, z). \end{aligned}$$

In Example 1, expecting z and carrying it out is a personal equilibrium, and so is expecting z' and carrying it out. Therefore, the agent's expectations are self-fulfilling. This, in itself, is not inconsistent with the behavior of a standard decisionmaker. But the non-indifference between the options in the two equilibria is. If an EU or KP decisionmaker is willing to choose both z and z' in the above decision problem, self 2 must be indifferent between them. Thus, if she is paid an arbitrarily small amount to choose, say, z, she will *always* choose it. Due to the continuity of u_2, this is not true in Example 1. In other words, the unpredictability of choice is more robust.

Crucially, as this discussion shows, neither extending our individual decisionmaker's *physical* consequence space, nor allowing utility to depend on past *physical* outcomes is in itself sufficient to generate unpredictability. In fact, in an EU setting (featuring, in particular, arbitrarily rich physical consequence spaces), any finite game of perfect information generically has only one equilibrium.

[7] For a discussion of revealed preference/testability issues, including why I need non-satiation, see Section 5.2.

[8] Hence, the first component in the utility function is period 1 expectations; $u_2(\delta_z, z')$, for instance, denotes self 2's utility when choosing z' after expecting to choose z for sure.

In a model of loss aversion that builds on the insights of prospect theory (Kahneman and Tversky 1979), Kőszegi and Rabin (2002) provide a natural application for the kind of multiplicity in Example 1. Consider an agent who is making the simple choice of whether to buy wine. Suppose that she compares her outcome to what she expected to get, and in both the money and the wine dimension, she is more sensitive to falling short of her expectations than ending up above them. If this agent goes to the store expecting to buy wine, returning home without wine will feel like a loss, and the money saved in the process will be coded as a gain. Being more sensitive to losses, the agent will tend to buy the wine. But if she expects not to buy, the money will be coded as a loss, and she will indeed tend not to buy. Thus, for a range of price levels, there are multiple equilibria. In that range, the agent's demand is not well-defined.[9]

In the characterization result, I will rule out exactly the kind of behavior in Example 1 by assuming that two equilibrium choices always leave the agent indifferent:

Definition 2 *The decisionmaker is predictable if for any y^t, z_t, and $d_{t+1} \in D_{t+1}$ such that only self $t+1$ makes a choice, if f_t and f'_t are the period t expectations in two continuation equilibria, then*

$$E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot) = E_{f'_t} u_{t+1}(y^t, z_t, f_t, \cdot) \text{ and } E_{f_t} u_{t+1}(y^t, z_t, f'_t, \cdot) = E_{f'_t} u_{t+1}(y^t, z_t, f'_t, \cdot).$$

The self-fulfilling nature of expectations highlights a crucial difference between my model and many prominent non-expected utility theories that take advantage of a recursive structure (Kreps and Porteus 1978, Chew and Epstein 1989, Epstein and Zin 1989, Skiadas 1998, among others). Multiple equilibria of the kind introduced in this section obviously make it impossible to summarize the future by some "certainty equivalent" or other utility measure that replaces the rest of the decision tree. As might be expected and as I discuss below, self-fulfilling expectations are distinct

[9] Although not a perfect fit, Example 1 is also easily adapted to handle the story of optimistic and pessimistic attitudes mentioned in the introduction. The only difference is that the actions now do not have immediate consequences. To adapt Example 1 to handle this, one can simply assume that the effort cost of investing in the future (e.g. exercising) is higher when expectations are more pessimistic, while optimistic and pessimistic agents can care equally about the fruits of the investment.

This formalizes what many self-help books and psychologists have claimed for a long time: that the mindset we have regarding life has a crucial role to play in how our lives turn out (Jones 1977). In particular, negative beliefs are widely viewed as crucial in depression, and the changes in preferences they induce (lethargy, loss of sexual desire and appetite, etc.) are important clinical symptoms of the condition. The therapies many professionals suggest may indeed work, if, in personal equilibrium, undergoing them can change the agent's expectations. Unfortunately, adopting the right mindset is not a simple matter of choice, but an equilibrium phenomenon, which may partially explain why overcoming negative beliefs is so difficult.

from time inconsistency, so a recursive structure does not exist in my model even if the agent is time consistent. Previous models rule out an interaction between preferences and past expectations–which drives unpredictability–by construction, as preferences are taken to depend (at most) on past physical outcomes and the decision problem looking forward.[10]

3.2 Complications Concerning Time Consistency

I motivate the discussion with the following example.

Example 2 $T = 2$, and there is no choice in period 1. In period 2, three possible outcomes are $\{z, z', z''\}$, and $u_2(\delta_{z'}, z') < u_2(\delta_z, z) < u_2(\delta_{z''}, z'')$. Suppose further that for all $\lambda \in [0, 1]$,

$$\begin{aligned} u_2(\lambda\delta_z + (1-\lambda)\delta_{z'}, z') &> u_2(\lambda\delta_z + (1-\lambda)\delta_{z'}, z), \\ u_2(\lambda\delta_{z'} + (1-\lambda)\delta_{z''}, z'') &> u_2(\lambda\delta_{z'} + (1-\lambda)\delta_{z''}, z'), \\ u_2(\lambda\delta_z + (1-\lambda)\delta_{z''}, z) &> u_2(\lambda\delta_z + (1-\lambda)\delta_{z''}, z''). \end{aligned}$$

When self 2's choice set is $\{z, z'\}$, it is a unique equilibrium for the agent to choose z'; when her choice set is $\{z', z''\}$, it is a unique equilibrium for her to choose z''; and it is a unique equilibrium for her to choose z from the set $\{z, z''\}$. Thus, behavior in this example violates transitivity. And since only self 2 makes a decision in this example, the intransitivity cannot be generated with EU or KP preferences.[11]

The reason that transitive preferences do not imply transitivity of choice arises from self 2's ability to deviate from the expectations of self 1. Choice from a given set, of course, depends on such off-equilibrium behavior, and self 2's off-equilibrium preferences are intransitive in the following sense: she prefers option z' after expecting option z, option z'' after expecting option z', and option z after expecting option z''. The story of the academic's summer plans mentioned in the introduction provides such an example.

Intuitively speaking, it is clear that Example 2 is related to time inconsistency in some way. Since self 1 has transitive preferences over outcomes of the form (δ_{z_2}, z_2), self 1 and self 2 have

[10] At the very end of his paper, Skiadas (1998) hints at an extension of his model in which preferences could depend on past expectations. He does not develop this extension formally. However, since Skiadas claims that this would result in a recursive representation, one of his underlying assumptions has to be to rule out preference reversals as a result of different expectations.

[11] It is relatively easy to check that the assumptions in Example 2 are consistent with each other. Since the restrictions only apply to the edges of the simplex, they can be "pasted together" to obtain a continuous u_2.

to "disagree" on at least one of self 2's above choices. Indeed, in the introduction's story, the academic's preference to work after having anticipated a vacation seems time inconsistent: in order for self 1 to enjoy the pleasures of anticipation, self 2 needs to go on vacation.

To formally address these intuitions, I need to define time consistency. I argue below that the two conventional definitions of time consistency–the one based on preferences and the one based on behavior–are unsuited for the current model. I then suggest an appropriate replacement.

According to a version of the preference-based definition, the decisionmaker is time consistent if, for all $1 \leq t < T$, selves t and $t+1$ have the same (possibly history-dependent) preferences over lotteries of outcomes starting in period $t+1$. This definition implies that the agent should be time consistent if she is an expected utility maximizer at all points in time and her von Neumann-Morgenstern utility function is the same for all intertemporal selves.

This definition fails miserably when anticipation is also allowed to enter the utility function. Even if all intertemporal selves have exactly the same utility function over streams of physical outcomes and expectations (that is, u_t equals some given u for all t), the decisionmaker will in general fail to maximize the expectation of u, because later selves do not behave in a way earlier selves would hope. In fact, Example 2 can easily be made into such a counterexample by assuming that $u = u_1 = u_2$. Then, even though both selves prefer the stream $(\delta_{z''}, z'')$ to the stream (δ_z, z), in equilibrium self 2 always chooses z from the two alternatives.

The behaviorally grounded conventional definition of time consistency deems the agent to be time consistent if, without commitment, she always carries out the history-contingent plan of action she would commit to at the beginning of the decision problem. Equivalently, the agent is time consistent if she has no value from commitment.

This definition faces several problems as well. Example 1, which features multiple utility-ranked equilibria, illustrates that the agent might prefer to commit her future action even if she is not time inconsistent in any meaningful sense. Even in the 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. Thus, the *exclusive* role of commitment is to change self 1's expectations, a situation very different from the notion of time inconsistency.[12]

[12] The distinction between time inconsistency and multiple equilibria can be made even sharper if–realistically–anticipatory feelings are allowed to depend randomly on expectations. Then, multiple utility-ranked equilibria can arise even if self 1 does not wish to commit self 2's action. Intuitively, self 1 may not want to commit self 2 to always choose the "good" action, since that could turn out to be a disaster if (by chance) she feels bad. But without such

Therefore, Example 1 puts into doubt whether a decisionmaker who values commitment is time inconsistent. But there is a more serious problem with the commitment-based definition of time consistency. Namely, it is fundamentally problematic to think of commitment decisions in the first place. When self 1 describes the strategy to which she would commit self 2, this should be a function of f_1, period 1 expectations, since these are part of the history for self 2. Thus, self 1 is optimizing for each given f_1. But self 1's expectations are a function of self 2's actions, and to maximize utility at time 1, self 1 has to take this into account. Thus, she is also maximizing as if f_1 is not given. Mathematically, such a problem is nonsensical.[13]

All of the above problems have to do with a pre-post distinction: self t exists before f_t is realized, while self $t+1$ exists afterwards. My proposed definition is an extension of the preference-based definition that deals with this complication.

Definition 3 *The decisionmaker is called time consistent if for all* $t \in \{1, \ldots, T-1\}$, $y^t \in Y^t$, $z_t \in Z_t$, *and* $d_{t+1} \in D_{t+1}$, *if expectations* f_t *and* f'_t *are feasible given* d_{t+1},[14] *we have*

$$E_{f'_t} u_{t+1}(y^t, z_t, f_t, \cdot) > E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot) \Rightarrow E_{f'_t} u_t(y^t, z_t, f'_t, \cdot) > E_{f_t} u_t(y^t, z_t, f_t, \cdot). \tag{2}$$

That is, if self $t+1$ prefers f'_t over f_t ex post, then self t also prefers it ex ante.

According to this definition, preferences over z and z'' in Example 2 are time inconsistent, even if $u_1 = u_2$: if self 1 has expected z'', self 2 prefers to choose z, yet self 1 prefers z'' over z. Intuitively, due to the *timing* of anticipation, self 2 fails to "invest" into period 1 expectations even though she cares about them, as the expectations are already realized by the time she makes her choice. And due to the *forward-looking* nature of anticipation, self 1's realization that self 2 will 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

commitment, she can end up in a low-utility equilibrium. This reinforces the claim that multiple equilibria are about self 1's beliefs being wrong, not self 2's behavior.

[13] The reason we do not encounter a similar difficulty in the standard framework of individual decisionmaking is that in those models the outcomes that are realized between the two periods do not depend on the future.

[14] An expectation f_t is feasible given d_{t+1} if there are internally consistent (Definition 1, part 2) $\phi_t, \ldots, \phi_{T-1}$ and $\sigma_{t+1}, \ldots, \sigma_T$ such that $f_t = \phi_t(y^t, z_t, d_{t+1})$.

different for different selves. The agent's choice over z and z'' in Example 2 formally shows how time inconsistency follows *purely* from the temporal placement of anticipatory feelings and rational expectations, not different preferences.

However, the choice between z and z'' is in itself not specific to agents who care about anticipation. Of course, even a standard decisionmaker's preferences can change, and there is no choice experiment that could reveal that self 1 and self 2 have identical preferences. Time inconsistency creates a novel form of behavior only in conjunction with the other choices in Example 2.[15]

But appearing due to the passage of feelings is not the only way time inconsistency can assert itself. When preferences are time inconsistent, the agent may want to manipulate her own expectations. While this is also true in standard settings (Carillo 1997, Carillo and Mariotti 2000, Bénabou and Tirole 2002), the following example shows that with utility from anticipation expectations can be manipulated even when this would not change the behavior of a standard decisionmaker.

Example 3 Suppose that $T = 2$, and three possible outcomes in period 2 are $z_2 = 0$, $z_2 = 1$, and $z_2 = 2$. With probability $\frac{2}{3}$, the outcome depends on self 1's choice between 0 and 2, and with probability $\frac{1}{3}$, it depends on self 2's choice between 0 and 1. The uncertainty about how the outcome is determined is revealed after self 2's choice.
Preferences are as follows. $u_2(f_1, z_2) = (E[f_1] - 1) \cdot z_2$. u_1 only depends on z_2, and $u_1(0) = 1$, $u_1(1) = 4$, and $u_1(2) = 0$.

In the unique equilibrium in this example, self 1 chooses 2. Intuitively, self 1 and self 2's interests are aligned if the agent expects $z_2 = 2$ to occur with high probability, but opposite if she expects it to occur with low probability. Thus, self 1 is willing to sacrifice some payoff to make self 2 more optimistic.

In the EU and KP models, the agent's preferences over two outcomes in period t do not depend on what other period t outcomes she finds possible at the time of choice. Thus, the kind of behavior in Example 3 cannot happen in these models. For example, people may buy a lottery ticket partly because the anticipation it buys helps them get through the grind of work every day, overcoming

[15] The time inconsistency in Example 2 highlights one odd aspect of Definition 3. Interestingly, in this example we can conclude that the decisionmaker is time inconsistent without knowing the preferences of self 1. Intuitively, time consistency is a requirement for selves to "agree" with each other. Without knowing the preferences of self 1, how can we say that they do not agree? The reason is that whatever self 1 wants, self 2 is not willing to carry it out (that is, she wants z' instead of z, z'' instead of z', and z instead of z''). Therefore, we do not need to know what self 1 wants to conclude that there is a disagreement.

a self-control problem. But, if anything, working hard only makes a difference if they do not win the lottery. Thus, the motivation must come from anticipation, not the actual outcomes.

Observation 1 below shows that, as intuition suggests, time inconsistency is a necessary ingredient of Examples 2 and 3. Thus, to rule out these behaviors, I use time consistency in the characterization theorems. The observation is also a useful insight for the theorems.

Observation 1 *Suppose that $T = 2$ and the agent is time consistent. Then, for any $z_1 \in Z_1$ and $d_2^0 \in D_2^0$, there is a continuation personal equilibrium with the expectation f_1^* that solves* $\max_{f_1 \in d_2^0} E_{f_1} u_1(z_1, f_1, \cdot)$.

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. □

In both Example 2 and Example 3, utility from expectations and outcomes interact. Thus, to rule out the phenomena in these examples, one could assume instead of time consistency that utility is additively separable in past expectations and other outcomes, *and* that it is additively separable in current expectations and future outcomes.[16] In that case, the only difference between my model and EU is in informational preferences, and the only difference between my model and the KP one is in time inconsistency. These assumptions, however, would (to an extent) trivialize the role of anticipation by isolating it from other forms of utility, as well as automatically eliminate all the other effects in this paper. Therefore, I choose not to use separability just to rule out Examples 2 and 3.[17]

3.3 Informational Preferences

In this section, I need the following notation to handle situations in which outcomes are known with certainty. Let $Z_t^+ = Z_{t+1} \times \cdots \times Z_T$, and for any $z_t^+ = (z_{t+1}, \ldots, z_T) \in Z_t^+$, let $\bar{z}_t^+ = (z_{t+1}, f_{t+1}(z_t^+), \ldots, z_T)$, where $f_{t'}(z_t^+)$ is defined backwards recursively by $f_{t'}(z_t^+) = (z_{t'+1}, f_{t'+1}(z_t^+), \ldots, z_T)$.

[16] While Example 2 relies on the interaction of preferences with past expectations, Example 3 does not: it applies equally well when preferences over *current* expectations and future outcomes interact. For example, we get essentially the same problem if, in Example 3, we have $T = 3$, the decisions and self 1's utility being about z_3, and $u_2(f_2, z_3) = (E[f_2] - 1) \cdot z_3$.

[17] I do in the end use a separability condition in Theorem 4, but this is motivated mostly by Example 1 above and Example 4 below.

That is, $\bar{z}_t^+$ is the vector of all future outcomes when the agent is certain in period $t+1$ that she will get the physical outcomes z_t^+ starting in that period. Let $\bar{Z}_t^+ = \{\bar{z}_t^+ | z_t^+ \in Z_T^+\}$. Clearly, $\Delta(\bar{Z}_t^+) \subset F_t$.

Once again, I motivate the discussion with an example.

Example 4 Suppose that T=2, Z_1 is trivial, and the agent is facing three possible outcomes z, z', z'' in period 2. u_1 is linear in f_1, it satisfies $u_1(\delta_z, z) > u_1(\delta_{z''}, z'') > u_1(\delta_{z'}, z')$, and the following additional conditions:

$$\begin{aligned} u_1(\delta_z, z) + u_1(\delta_{z'}, z') - u_1(\delta_z, z') - u_1(\delta_{z'}, z) &= 1 \\ u_1(\delta_z, z) + u_1(\delta_{z''}, z'') - u_1(\delta_z, z'') - u_1(\delta_{z''}, z) &= 0 \\ u_1(\delta_{z'}, z') + u_1(\delta_{z''}, z'') - u_1(\delta_{z'}, z'') - u_1(\delta_{z''}, z') &= 0. \end{aligned}$$

Example 4 is quite rich, and I will use it to introduce a number of important themes. To start, suppose that the agent is facing a lottery that gives her z or z', with probability one-half each. Would self 1 choose to find out its outcome in advance? If she learns the outcome in period 1, she gets the expected utility

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

while waiting to resolve the uncertainty gives her the expected payoff

$$\frac{1}{2}u_1\left(\frac{1}{2}\delta_z + \frac{1}{2}\delta_{z'}, z\right) + \frac{1}{2}u_1\left(\frac{1}{2}\delta_z + \frac{1}{2}\delta_{z'}, z'\right). \tag{4}$$

Since u_1 is linear in f_1, expression 4 becomes

$$\frac{1}{4}\left[u_1(\delta_z, z) + u_1(\delta_z, z') + u_1(\delta_{z'}, z) + u_1(\delta_{z'}, z')\right]. \tag{5}$$

The difference between self 1's utility level when uncertainty is resolved early (expression 3) and late (expression 5) is then

$$\frac{1}{4}[\underbrace{(u_1(\delta_z, z) - u_1(\delta_z, z'))} - \underbrace{(u_1(\delta_{z'}, z) - u_1(\delta_{z'}, z'))}] = \frac{1}{4}. \tag{6}$$

Thus, our agent prefers to resolve the uncertainty in the first period. In fact, it is easy to prove that when facing *any* lottery between z and z', the agent prefers more information to less.[18]

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

Now suppose that the agent is facing the lottery in which she can get each of z, z', and 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 ignorant is

$$\frac{1}{9}\big[(u_1(\delta_z,z)+u_1(\delta_{z'},z')-u_1(\delta_{z'},z)-u_1(\delta_z,z'))+(u_1(\delta_z,z)+u_1(\delta_{z''},z'')-u_1(\delta_{z''},z)-u_1(\delta_z,z''))+$$
$$+\,(u_1(\delta_{z''},z'')+u_1(\delta_{z'},z')-u_1(\delta_{z'},z'')-u_1(\delta_{z''},z'))\big]=\frac{1}{9}.$$

Thus, the agent would still choose to fully resolve this uncertainty. But suppose that her only learning opportunity is to find out whether the outcome is z'' or one of $\{z,z'\}$. If she chooses to learn this partial information, with probability two-thirds she will find that the outcome is in $\{z,z'\}$. In that case, she faces the same 50-50 lottery as above. Thus, the difference in expected utility between full learning and partial learning is $\frac{2}{3}\cdot\frac{1}{4}=\frac{1}{6}>\frac{1}{9}$. Therefore, the agent would rather remain ignorant than partially learn the truth.

Motivated by these examples, I introduce the following concepts to formally address informational preferences.

Definition 4 *Self t is called <u>one-period information neutral</u> if for all $y^t\in Y^t$, $z_t\in Z_t$, and $f_t\in\Delta(\bar{Z}_t^+)$ with compact support, we have*

$$\int u_t(y^t,z_t,f_t,z)df_t(z)=\int u_t(y^t,z_t,\delta_z,z)df_t(z).$$

She is <u>one-period resolution loving</u> if for all $y^t\in Y^t$, $z_t\in Z_t$, and $f_t\in\Delta(\bar{Z}_t^+)$ with compact support,

$$\int u_t(y^t,z_t,f_t,z)df_t(z)\le\int u_t(y^t,z_t,\delta_z,z)df_t(z).$$

If for all $y^t\in Y^t$ and $z_t\in Z_t$, $\int u_t(y^t,z_t,f_t,z)df_t(z)$ is convex in f_t over the domain $\Delta(\bar{Z}_t^+)$, then self t is <u>one-period information loving</u>. The agent is one-period information neutral, one-period resolution loving, and one-period information loving, respectively, if all of her selves $t\le T-1$ are one-period information neutral, one-period resolution loving, and one-period information loving, respectively.

One-period information neutrality, the condition to be used in Theorem 3, requires the agent to be indifferent to whether decision-irrelevant information about future outcomes is revealed today or tomorrow. One-period resolution lovingness means that the agent would prefer to fully resolve uncertainty today rather than tomorrow. She is one-period information loving if, given that all uncertainty will be resolved tomorrow, she prefers more information to less.

In Example 4, self 1 is one-period resolution loving, but not one-period information loving. Since these preferences violate one-period information neutrality, they are obviously inconsistent with EU. But they are also inconsistent with KP preferences (and their generalization á la Epstein and Zin (1989)), even though those preferences are *designed* to capture choices over information acquisition. In KP, if the agent is one-period information loving for any mixture of her best and worst outcomes (z and z' in this case), she is one-period information loving globally.

In describing informational preferences, previous work (Kreps and Porteus 1978, Epstein and Zin 1989, Grant, Kajii, and Polak 1998, 2000) has focused on the shape of the utility function in what might be called the "anticipatory part," a certainty equivalent associated with the future. If the utility function is linear in this anticipatory part, the agent is one-period information neutral. An analogue of such linearity is u_1's linearity in f_1 in Example 4, yet Example 4 features rather rich informational preferences. Thus, to understand the novelties my model adds relative to previous ones, and to gain insight into how one-period information neutrality can be violated, I now characterize one-period informational preferences when u_t is linear in f_t. Interpreting the shape of u_t in f_t as representing preferences over insecurity, call the agent insecurity neutral if this linearity condition holds.

To understand why the insecurity neutral agent is one-period resolution loving in Example 4, return to expression 6. Since z is a better outcome than z', the first difference inside the square bracket represents the utility loss from the disappointing outcome of z' when expecting the outcome z. Analogously, the second difference is the agent's utility gain from being pleasantly surprised by outcome z after expecting outcome z'. Thus, self 1 prefers early resolution because she is disappointment averse–her utility loss from a future disappointment is greater than her utility gain from a future pleasant surprise. Intuitively, if the agent does not know what will happen to her, she will sometimes be disappointed in the outcome and sometimes be pleasantly surprised. If she is disappointment averse, she prefers to know the outcome in advance in order to avoid disappointments and surprises. To draw a precise connection between one-period informational preferences and disappointment preferences, I formally define disappointment aversion.

Definition 5 *Self t is* <u>*disappointment averse*</u> *if for all y^t, z_t and for all $z_t^+, \zeta_t^+ \in Z_t^+$*

$$u_t(y^t, z_t, \delta_{\bar{z}_t^+}, \bar{z}_t^+) + u_t(y^t, z_t, \delta_{\bar{\zeta}_t^+}, \bar{\zeta}_t^+) \geq u_t(y^t, z_t, \delta_{\bar{z}_t^+}, \bar{\zeta}_t^+) + u_t(y^t, z_t, \delta_{\bar{\zeta}_t^+}, \bar{z}_t^+).$$

If the two sides are equal for all $z_t^+, \zeta_t^+ \in Z_t^+$, self t is <u>*disappointment neutral.*</u>

Definition 5 is a more general version of Gul's (1991) disappointment averse preferences in two senses. First, it applies to more than two periods. Second, even for two periods, Gul's representation satisfies an additional regularity condition to be introduced later (Definition 6).

The intuition for why self 1 is one-period resolution loving applies only to full learning. The only surefire way to avoid further disappointments is to know the outcome with certainty, so partial learning is qualitatively very different. As Example 4 shows, a disappointment averse insecurity neutral agent is not necessarily one-period information loving. Roughly, the agent is most sensitive to the disappointment of getting z' after expecting z (and does not mind getting z'' after expecting z or getting z' after expecting z'' so much). Thus, learning that the outcome is z or z' sets her up for a major disappointment. For example, the disappointment from a demotion might be much more devastating when the alternative was a nice promotion than when it was, say, no change. If so, ruling out the possibility of no change could decrease expected utility.

Loosely speaking, in order for the agent to be one-period information loving, her sensitivity to disappointment has to be more regular than in Example 4. A sufficient regularity condition can be motivated by the following consideration. Psychologically, being disappointed means that one gets a worse outcome than one hoped for or expected, causing disutility above and beyond the bad outcome itself. The converse holds for pleasant surprises. It is reasonable to expect the disappointment and pleasant surprise, and therefore also disappointment aversion, to be related to the relative desirability of the two outcomes. Disappointment aversion is regular if this relationship is linear: $u(\delta_{z_2}, z_2) + u(\delta_{\zeta_2}, \zeta_2) - u(\delta_{\zeta_2}, z_2) - u(\delta_{z_2}, \zeta_2) = \alpha |u(\delta_{z_2}, z_2) - u(\delta_{\zeta_2}, \zeta_2)|$ for some $\alpha \geq 0$. The three conditions in Example 4 clearly cannot be consistent with a single such α. The following definition generalizes this regularity condition to multiple dimensions, which allows disappointment aversion to depend on more than one attribute of the outcomes in question. The generalization applies, for example, to a consumer who is consuming multiple goods and is disappointment averse in each of the consumption dimensions separately.

Definition 6 *Self t is* <u>*regular disappointment averse*</u> *if there are finitely many bounded measurable* *functions* $c_1, \ldots, c_K : \bar{Z}_t^+ \to \mathbb{R}$ *and a symmetric function* $g : \mathbb{R}_+^K \to \mathbb{R}_+$ *such that* g *increases in each component and satisfies homogeneity of degree one, and for all* $\bar{z}_t^+, \bar{\zeta}_t^+ \in \bar{Z}_t^+, y^t, z_t$,

$$\begin{aligned} & u_t(y^t, z_t, \delta_{\bar{z}_t^+}, \bar{z}_t^+) + u_t(y^t, z_t, \delta_{\bar{\zeta}_t^+}, \bar{\zeta}_t^+) - u_t(y^t, z_t, \delta_{\bar{z}_t^+}, \bar{\zeta}_t^+) - u_t(y^t, z_t, \delta_{\bar{\zeta}_t^+}, \bar{z}_t^+) = \\ = \; & g(|c_1(\bar{z}_t^+) - c_1(\bar{\zeta}_t^+)|, \ldots, |c_K(\bar{z}_t^+) - c_K(\bar{\zeta}_t^+)|). \end{aligned}$$

The key aspect of Definition 6 is the symmetry of g. If disappointment aversion does not depend on the different attributes symmetrically, partial learning can lead the agent to expect further disappointments in attributes in which she is most sensitive to them, and thus to prefer less information to more.

Now, the following theorem formalizes and proves the intuitions developed so far.

Theorem 2 *Suppose that self* t *is insecurity neutral. Then*

1. *self* t *is one-period resolution loving if and only if she is disappointment averse;*

2. *if self* t *is regular disappointment averse, she is one-period information loving.*

Theorem 2 deals in essence with the relationship between preference for *early* information and aversion to *future* disappointments. Intuitively, it seems like a corresponding relationship should exist between preference for *late* information and aversion to *current* disappointments. That is, the agent's aversion to being disappointed today should make her reluctant to receive information. The reason this effect does not enter the discussion is that it is formally (and also psychologically) difficult to distinguish from preference for insecurity: u_t could be nonlinear in f_t for both disappointment and insecurity reasons. By imposing more structure, one could possibly distinguish disappointment aversion from insecurity lovingness by making the former the part that is dependent on past expectations. This interesting exercise is beyond the scope of this paper.

To show that a violation of one-period information neutrality always reduces to a combination of non-neutrality to insecurity and non-neutrality to disappointment, and thus I have fully characterized differences from previous models, observe the following:

Observation 2 *If self* t *is insecurity neutral and disappointment neutral, she is one-period information neutral.*

Proof. Obvious from the proof of Theorem 2. □

4 How is Utility from Anticipation Reflected in Behavior?

4.1 The Observational Equivalence Theorems

Section 3 has identified three major observable differences between agents who derive utility from anticipation and those who do not. These differences revolve around informational preferences, unpredictability, and time inconsistency issues. The natural question arises: when these phenomena are not present, are there other distinguishing forms of behavior? When preferences satisfy a small extra assumption, a version of non-satiation, the answer is no. The agent is defined to satisfy non-satiation if for all t and y^t, the image of $\bar{Z}_{t-1}^{+}$ under u_t is unbounded from above. This last assumption holds in virtually all applications of interest.[19]

Theorem 3 *If the decisionmaker is time consistent, one-period information neutral, predictable, and satisfies non-satiation, then she is observationally equivalent to a time consistent decisionmaker whose utility function depends only on physical outcomes. That is, there is a utility function v defined on $Z_1 \times \cdots \times Z_T$ such that she acts as if she maximized the expectation of v in every decision problem.*

The proof contained in the appendix proceeds by backward induction, proving at each stage that the statement is true starting in period t. The basic idea of the inductive step is the following. By a logic analogous to that of Observation 1, maximizing self t's expected utility (given the constraint that she has to be rational) is a personal equilibrium starting in period t. Next, predictability means (loosely) that self $t+1$ is indifferent between all continuation equilibria after z_t is realized. By time consistency, self t is also indifferent. Combined with the previous claim, this means that any personal equilibrium starting in period t maximizes self t's expected utility (again, constrained by rationality). Now, since the future selves behave like EU maximizers, and the agent is time consistent, self t is indifferent as to when in the future information is revealed. Since she is also one-period information neutral, she is completely indifferent as to the timing of resolution of uncertainty. In conclusion, selves t through T also behave as EU maximizers.

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

[19] Although randomization between equilibria is necessary to guarantee the existence of personal equilibrium in general, this assumption is not necessary under the conditions of Theorems 3 and 4. But both results would still be true (with essentially the same proof) if we did allow for randomization between equilibria.

with observed violations of one-period information neutrality. However, Example 4 shows that some non-monotonic informational preferences due to the interaction of expectations and future outcomes *cannot* be captured in their framework. To rule out such "intricacies" in behavior, I use a separability condition between expectations and the corresponding outcomes. In addition, Kreps and Porteus rule out time inconsistency by assumption, so choices in Examples 2 and 3 violate their model. The following theorem proves that when time consistency and an additive separability condition are imposed, there do not remain any behavioral differences between a model incorporating anticipation and KP.[20] Perhaps surprisingly, the same separability condition is sufficient to rule out both "intricate" informational preferences and unpredictability, even though the former is about inseparabilities between current expectations and future outcomes, while the latter is about inseparabilities between past expectations and current outcomes. The reason is that time consistency ties the two together.

Theorem 4 *Suppose that the decisionmaker is time consistent and satisfies non-satiation. Furthermore, suppose that for each* $1 < t \leq T$*,* u_t *is additively separable in past expectations and other outcomes. Then there are functions* $v_t : Z_1 \times \cdots \times Z_t \times D_{t+1} \to \mathbb{R}$ *such that i.) given* $z_1, \ldots, z_{t-1}$*, self* t *bahaves as is maximizing the expectation of* v_t*, and ii.) for any* $d_{t+1}, d'_{t+1} \in D_{t+1}$ *and* $z^{t+1} \in Z_1 \times \cdots \times Z_t$*,*

$$v_t(z^{t+1}, d_{t+1}) \geq v_t(z^{t+1}, d'_{t+1}) \Leftrightarrow \max_{l_{t+1} \in d_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot) \geq \max_{l_{t+1} \in d'_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot).$$

Theorems 3 and 4 are tight in the sense that they become false if any of the assumptions is removed; nor does any one of the assumptions follow from the others.

4.2 Discussion of the Main Result

Theorems 3 and 4 might be quite surprising. A reasonable first reaction to the model of this paper is that it is not at all restrictive, that preferences over anticipation, by virtue of their inordinate

[20] Allowing for preferences over the timing of resolution of uncertainty as in the KP model, Chew and Epstein (1989) study the novelties this adds to an expected utility model. Although they do not explicitly model anticipation, some of their results have a similar flavor to the characterization theorems. In their 2-period KP type model, time consistency and neutrality to the timing of the resolution of uncertainty imply an expected utility representation. Time consistency does not seem to play a central role in their result; it only ensures that self 2's preferences over lotteries in that period are linear, which, in an expected utility model as mine, would be the case anyway (since it is the last period).

complexity, can rationalize nearly any kind of behavior. Theorems 3 and 4 show, in constrast, that the kinds of novel behavior fall in one of only three categories.[21] Thus, utility from anticipation creates an important but limited set of new phenomena.

By limiting the set of new effects to be analyzed, Theorem 3 (and Theorem 4) offers an aid for thinking about the additional types of behavior that are made possible by introducing anticipation into motivation. Once we have a detailed understanding of the few phenomena ruled out in the theorem, and how they connect to key economic variables, we get an elegant and parsimonious theory that can be applied in a variety of contexts. In any model that incorporates anticipation and purports to expand, say, EU, one can first check whether it violates time consistency, one-period information neutrality, or predictability. If it does, one can identify which of these conditions fails, and then apply what we know about the concept to gain insight into the model. If, on the other hand, the model lacks all three phenomena, it cannot be justified from an empirical point of view. Of course, much more work is needed to arrive at a detailed understanding sufficient to do this. But, for instance, this paper's analysis of disappointment aversion and informational preferences is a step in that direction.

However, it is also important to emphasize some points of caution regarding Theorems 3 and 4. I have assumed rationality in the definition of personal equilibrium, thereby eliminating a possible behavioral difference between emotional and non-emotional decisionmakers right in the setup of the problem. The reason for maintaining rationality is the same as remaining in the expected utility framework: I want to deviate only one step at a time from the standard model. But rationality not only seems empirically violated in some situations, there is a theoretical reason against it as well. Specifically, a person who cares about anticipation might want to fool herself–*choose to* abandon rationality–because rational expectations bar her from forming beliefs that make her feel better.[22] This is never the case for decisionmakers who only care about physical outcomes. Although I am not aware of a convincing formal model of how people fool themselves, it

[21] The characterization results are interesting even though the assumptions required to get them are admittedly strong. Time consistency and information neutrality are known to put strong restrictions on standard preferences, but it was not known what kind of restrictions they put on utility from anticipation *relative to* standard preferences.

[22] Eliaz and Spiegler (2002) arrive at a similar conclusion–that rationality may not be the best assumption when the decisionmaker derives utility from anticipation–in a different way. They argue convincingly that plausible patterns of informational preferences that seem to come from anticipatory feelings in fact cannot be reconciled with such preferences if the decisionmaker is Bayesian rational.

is reasonable to expect that such a model would generate additional differences in behavior.[23] But even if a model in which the decisionmaker can to an extent reward herself through self-deception is psychologically very plausible, this does not render the exercise in the current paper meaningless. To categorize differences in behavior that a richer model of that sort allows, we can first look for the behavioral manifestations of self-deception, and if those are not present, only the behavioral differences identified in this paper remain.

In addition, Theorem 3 should probably be viewed as one possible–though, given the history of work on this subject, natural–*categorization* of the differences between hot and cold decisionmakers. In other words, there may be other ways to "cover" the same set of behavioral phenomena. For example, the discussion on time consistency, as well as the proof of Theorem 3, make it clear that this concept is instrumental in ruling out a variety of different behaviors.

5 Further Comments

5.1 Comments on the Model and Related Literature

First, as mentioned in the introduction, the setup of the model is an extension of the psychological expected utility model of Caplin and Leahy (2001) in two respects: in capturing the temporal placement of expectations, and in allowing for interactions between expectations and later outcomes. These features are responsible for a majority of the novel phenomena I introduce; in fact, Examples 1 through 4 all generate behavior that would not happen with Caplin and Leahy's preferences.

The assumption that expectations affect the preference ordering over other outcomes is justified by extensive evidence.[24] But besides the obvious aim of exploring the resulting important conse-

[23] There do exist several models, mostly in the quasi-Bayesian paradigm, in which agents fool themselves in some way, possibly because they want to make themselves feel better. See for example Rabin and Schrag (1999), Akerlof and Dickens (1982), Gervais and Odean (2001), or Manove and Padilla (1999). These papers, however, are mostly interested in the consequences of a given bias, and do not model how the bias arises from preferences over anticipation (or something else).

[24] There are three distinct situations that the interaction of utility with past expectations might capture. When expectations are formed almost simultaneously with actions, or when feelings are drawn out over time, the model's formalism represents the influence of lingering moods on subsequent actions. Anticipation has a well-documented effect on preferences along these lines, as in the case of depression mentioned in Section 3.1 (see Footnote 9).

The dependence of u_t on past expectations can also depict the possibility that the utility self t derives from anticipation or physical outcomes depends on what she expected in the past. There is strong evidence that our preferences have this facet as well (Mellers 2000, Mellers, Schwartz, and Ritov 1999, for example). (Kőszegi and Rabin (2002) argue that the wide array of experimental results on loss aversion and the endowment effect are best understood in terms of expectations-dependent preferences.) In Example 0, self 2's dejection from not being able to quences of anticipation, another major reason for considering a more general model than Caplin and Leahy (2001) has to do with the different goals of this paper. Though formulating a unified framework, Caplin and Leahy focus on the novel insights their model can deliver in specific applications (e.g. asset pricing). Importantly, this paper also aims to theoretically explore the boundaries of behavior generated by anticipation, an exercise that calls for the most general framework possible.

Second, many non-expected utility models that allow for preferences over the resolution of uncertainty (KP, Skiadas 1998, Grant, Kajii, and Polak 1998, 2000, and others) are special cases of my model. These papers (as well as the infinite-horizon models of Epstein and Zin 1989 and Weil 1990) approach questions similar to mine from a completely different perspective. Their starting point is preferences over the 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 shapes preferences and influences behavior, and is often singled out as a factor in informational preferences. Therefore, it is a reasonable alternative to start from these primitives. In addition, starting from the origins of informational preferences allows one to identify the full range of behaviors these underlying motivations can induce, including a complete understanding of informational preferences themselves.

One important feature of my model that derives from its different perspective is the complete separation between preferences and how equilibrium behavior is determined. In decision theory, the agent's expectations about future behavior is often a component of preferences.

Third, my model also incorporates as a special case the assumption that the agent cares about the *utility* of past or future selves. This assumption is often made in one-sided (Barro 1974, Bernheim and Ray 1987) and two-sided (Kimball 1987, Hori and Kanaya 1989, Bergstrom 1999) intergenerational altruism models. Agents in these models are still standard in the sense that they do not care about anticipation: if altruism is sufficiently weak (so that the interdependent utility functions do not "blow up"), these utility functions are equivalent to utility functions over consumption alone.

Other special cases of my model, and ones that are closest to it in spirit, are two-sided inter-

derive pleasure from anticipation after having done so earlier, and self 3's aversion to a high price are both examples of such a preference.

Similarly, the utility from future outcomes can depend on current expectations, just as the utility from current outcomes can depend on past expectations. An example is self 3's preference for a surprise to come in period 4.

generational altruism models in which each generation takes the utility of the past generation as given (Hori and Kanaya 1989, Bergstrom 1999), just as my decisionmakers take past expectations as given. In contrast to mine, the main concern of these papers is not behavior (no equilibrium concept is offered), but how the set of interdependent utility functions can be represented by utility functions over allocations and the past generation's utility. Also, although these authors note that preferences over consumption sequences are typically time inconsistent, they do not mention the possibility of time inconsistency due to sunk past expectations, which is central to Example 2.

Thus, from a formal point of view, this paper can be considered an addition to the intergenerational altruism literature: it compares behavior in intergenerational models in which generations do versus do not derive utility from anticipation. However, I find this application less interesting, because generations are less likely to care about past generations' expectations than an individual cares about her own past expectations.

Fourth, importantly, there are aspects of decisionmaking under uncertainty that are impossible to capture in this framework. For example, the agent cannot be affected by any information that would only be relevant had she made a different decision in the past. If, in Example 0, the agent is not at the agency in period 2, her expectations cannot be affected by the price she learns. As a result, her behavior does not depend on whether or not she learns the price if she does not visit the agency. In particular, she cannot be motivated by *regret* about not going to the agency when she finds that the price is low.[25] Similarly, while Section 3.3 shows that utility from anticipation can account for rich informational preferences, Eliaz and Spiegler (2002) demonstrate that the framework does put important restrictions on the information acquisition choices agents can make.

Fifth, in an influential paper, Geanakoplos, Pearce, and Stacchetti (1989) introduce a static multiple-player equilibrium concept (psychological equilibrium) for preferences that can depend on beliefs about others' strategies. Due to belief-dependent preferences, some phenomena are similar to those in this paper (e.g. there might not be a pure-strategy equilibrium in a game where there would be one with standard preferences). But many of the phenomena in the current paper are dynamic by nature.[26]

[25] Also, utility does not depend on how the agent arrives at a given distribution of outcomes; e.g. if she ends up without the ticket with probability 1, it does not matter whether she achieves this through staying away from the agency or by not buying once there. In particular, the agent is not *tempted* by other options.

[26] Also, anticipation in my setting is very different from beliefs about strategies in at least one way: later selves know what earlier selves anticipated, while players generally cannot fully learn other players' strategies. Partly for

5.2 Revealed Preference Issues

The main subject of this paper is how extending the agent'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. Thus, I blur the line between testable and untestable aspects of the model, creating a slight tension in the paper. This section attempts to clarify these issues. For the discussion below, I maintain the non-satiation assumption as defined in Section 4. Non-satiation is likely to hold in almost any setting of interest. It is also 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.

Examples 1 through 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 agent's behavior in a few decisionmaking problems could notice the behaviors these examples demonstrate.[27] Similarly, any failure of the key restrictions used in Theorem 3 (predictability, time consistency, and one-period information neutrality) can be revealed in behavior.

But this does not mean that *all* aspects of the preferences I used in the paper can be revealed in choice. Examples 1, 2, and 3, and the separability assumption in Theorem 4, make restrictions on the interaction of later preferences with earlier expectations. In Example 2, for instance, I make the assumption that self 2 prefers the stream $(\delta_{z''}, z)$ over the stream $(\delta_{z''}, z'')$. Clearly, it is impossible to set up a choice experiment in which self 2 would get the opportunity to reveal this preference: if self 1 expected to be able to choose between z'' and z (with some probability), *and* self 2 actually had the preferences in Example 2, self 1 could not have expected z'' for sure.

This is a minor problem. Assuming continuity, it is easy to extract self 2's preferences over Z_2 for any given f_1. To decide whether she likes z_2 or z'_2, we give the agent the distribution f_1 with probability $1-\epsilon$, and let self 2 choose between z_2 and z'_2 with probability ϵ. Letting ϵ go to zero,

this reason, it is difficult to convincingly extend the psychological equilibrium concept to dynamic settings. See Dufwenberg and Kirchsteiger (1998) for a discussion.

[27] A stronger revealed preference concept would require the phenomenon to be displayed *in equilibrium.* At the cost of some complication, one can extend these examples so that all novel phenomena are exhibited in equilibrium. Loosely, one can give the agent a lottery over a well-chosen set of decision problems, and ask her to predict her behavior in each of them. Using non-satiation, we can ensure that she gives correct predictions. In essence, this turns our observation of a single decision problem into an observation of multiple ones.

we obtain self 2's ordering between the two options. With this method, one can in general elicit how the agent's preferences over physical outcomes interact with past expectations.

However, the same method does not work to elicit the interaction of future physical outcomes and current expectations, on which Example 4 depends. The key distinction is that a current choice changes current expectations, but it does not change past expectations. Looking *backward*, the agent's past expectations and current choices can diverge, since information has been revealed since the expectations were formed. But if expectations are rational, the agent's expectations and the distribution of future outcomes are the same. Thus, looking *forward*, we cannot gauge the agent's preferences over streams in which outcomes diverge from expectations. In fact, it is easy to give a pair of utility functions that lead to the same behavior in all decisionmaking problems: for $T = 2$ and $Z_2 = \mathbb{R}$, 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, Section 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.[28] More fundamentally, the above untestability might highlight a problem with revealed preference, not a model incorporating anticipation. Kőszegi (2001) shows that a doctor might act differently if her patient has anxiety about her medical condition than if she does not, even if the patient behaves exactly the same way in all individual decisionmaking problems. If the patient has anxiety, the doctor uses her superior information to try to make her feel better. Crucially, despite rational expectations by the patient, 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 doctor-patient problem.

Depending on the interpretation, these arguments seem to indicate that one has to either reject the idea that models should have a revealed preference foundation, or reject a model incorporating anticipation. If one is willing to accept a more general version of revealed preference, I believe neither need to be the case. 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 once again create a divergence between expectations and future

[28] In addition, when we impose more structure on how expectations can enter utility (restricting what insecurity preferences can capture), only disappointment preferences may be able to generate the complex behavior in Example 4. A psychologically plausible structural assumption of this sort is that only a scalar function of f_t enters u_t; roughly, f_t affects utility only in as much as it means "high" or "low" expectations.

outcomes, and thus elicit all preferences used in this paper. Carrying out this exercise is left for future work.

6 Conclusion

The goal of this paper is to provide a perspective on the decisionmaking consequences of utility from anticipation. I approach this question by assuming the most general expected utility function over anticipation and physical outcomes, and then catalogue the novel forms of behavior it generates. There are differences between standard decisionmakers and those in my model relating to time inconsistency, informational preferences, and unpredictability. After discussing these phenomena in detail, the paper proves that any novelty in behavior has to be related to at least one of them.

It remains to apply this framework to specific decisions of economic importance, fulfilling the promise that it aids in analyzing important questions. An immediate application is a more refined understanding of risk aversion. Specifically, consider what we can learn about the famous equity premium puzzle. As was first pointed out by Mehra and Prescott (1984), the coefficient of risk aversion that is backed out of historically observed excess returns on equities seems puzzlingly high compared to risk aversion exhibited in other settings. If a model based on utility from anticipation can help explain this puzzle, it has to violate time consistency, one-period information neutrality, or predictability. Self-fulfilling expectations do not seem important for this application. Informational preferences do seem promising, because they substantially enrich risk preferences. In particular, decisionmakers can be more risk averse in a dynamic, than in a static, setting for two reasons: disappointment aversion and insecurity aversion. At first sight, both of these can account for the equity premium puzzle. But consider insecurity aversion in itself first (by assuming that utility from expectations is separable from other outcomes). As Theorem 4 indicates, unless the agent is time inconsistent, she will behave in the same way as a decisionmaker with KP preferences. Weil (1989) argues, however, that reasonable parameterizations of the KP model cannot account for the equity premium puzzle. And it seems unlikely that risk preferences would be driven by time inconsistency: people do not generally express a concern that they will take too little risk in the future. Therefore, if anticipation is to help explain the equity premium puzzle, disappointment aversion has to be involved to a certain extent. All this is a heuristic argument. But if made precise, Theorem 3 says that we have one complete description of how utility from anticipation helps explain the equity premium puzzle.

A bolder and more unusual direction for future research is to develop the implications of the *vulnerability* that anticipation introduces into human motivation. Since the agent has no real control over her expectations–them being in a large part in the hands of future selves–if they get worse, there is little she can do in the short run to make herself feel better. How people deal with these kinds of threats is an exciting question that has not been addressed in economic theory.

A Existence of Personal Equilibrium

This section proves that personal equilibrium exists under fairly general conditions. Example 2 (which does not have an equilibrium when self 2's decision problem is $\{\delta_z, \delta_{z'}, \delta_{z''}\}$) demonstrates that we have to allow for mixed strategies in order to guarantee existence of personal equilibrium. To motivate the need for allowing randomization between equilibria, I give an example below in which equilibrium does not exist. Then I move on to the existence result.

Example 5 The agent is involved in a four-period decision problem. Relevant physical outcomes occur in period 4, and there are three of them: z, z', and z''. The decision tree is illustrated in Figure 3. In addition to the physical outcomes, self 2's utility depends on expectations in period 1 and self 4's utility depends on expectations in periods 1 and 3. The utility functions satisfy the following restrictions:

$$\begin{aligned}
u_2(f_1, z) &> u_2(f_1, z')\,\forall f_1 && \text{A1}\\
\forall f_1 = \lambda\delta_{z''} + \frac{1-\lambda}{2}(\delta_z + \delta_{z'}),\ u_2(f_1, z) &> u_2(f_1, z''). && \text{A2}\\
\forall \lambda \in [0,1], u_2(\lambda\delta_z + (1-\lambda)\delta_{z''}, z) - u_2(\lambda\delta_z + (1-\lambda)\delta_{z''}, z'') &= 1 - 2\lambda && \text{A3}\\
u_3(z) - u_3(z'') &= u_3(z'') - u_3(z') > 0 && \text{A4}\\
u_4(\delta_z, f_3, z') &> u_4(\delta_z, f_3, z)\,\forall f_3 && \text{A5}\\
f_1 = \frac{1}{2}\delta_z + \frac{1}{2}\delta_{z''} \Rightarrow u_4(f_1, f_3, z) &> u_4(f_1, f_3, z')\forall f_3 && \text{A6}\\
f_1 = \frac{3}{4}\delta_z + \frac{1}{4}\delta_{z''} \Rightarrow u_4(f_1, \delta_z, z) > u_4(f_1, \delta_z, z') \text{ and } u_4(f_1, \delta_{z'}, z') &> u_4(f_1, \delta_{z'}, z) && \text{A7}
\end{aligned}$$

With these assumptions, there is no personal equilibrium. A certain outcome of z' and z'' are ruled out by A1 and A2. If the agent expects z in period 1, self 2 prefers the outcome z'' (A3), which is, by A4 and A5, the unique continuation equilibrium if she stays in. A1 also rules out any mixture between z and z'. Clearly, a mixture between z' and z'' cannot be an equilibrium, since (by A4) if self 3 expected self 4 to choose z' for sure, she would rather choose z''. Also by A4, if self 3 expected self 4 to choose z for sure, she would stay in with probability 1. Thus, a

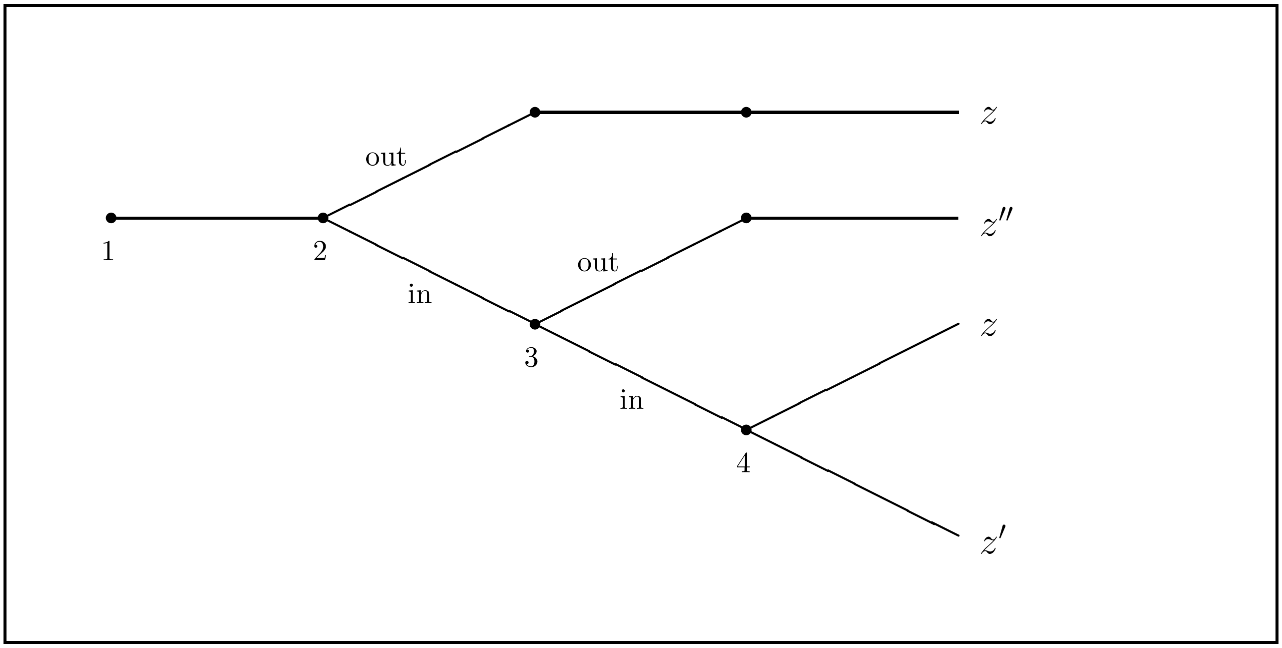

Figure 3: An Example Where Personal Equilibrium Does Not Exist

mixture between z and z'' is only possible if self 2 uses a mixed strategy. But when self 2 is indifferent between z and z'', by A4 and A6 the unique continuation equilibrium results in z. Finally, by A2, a mixture between all three outcomes cannot be an equilibrium, either: when self 3 is willing to mix, self 2 will exit.

Allowing for "public" randomization between equilibria eliminates this non-existence problem. For $f_1 = \frac{3}{4}\delta_z + \frac{1}{4}\delta_{z''}$, there are multiple continuation equilibria after self 3's choice to stay in: by A7, there is one in which self 4 chooses z with probability one, and one in which she chooses z' with probability one. If self 4 would choose z', self 3 would rather exit and choose z'', while she would stay in if self 4 would choose z. Thus, if the agent could randomize between the two equilibria, she could mix between z and z''. This would be sufficient to guarantee equilibrium in the current example.

The need to allow for randomization between equilibria is similar to that in Harris, Reny, and Robson's (1995) proof of existence of subgame-perfect equilibrium in a large class of games. In their case, public randomization is necessary to correlate the expectations of players who move simultaneously, thereby convexifying the set of equilibrium outcomes. The current paper deals with individual decisionmaking, so the players (the intertemporal selves) never move simultaneously. In fact, the decisionmaker's problem in Example 5 satisfies the conditions of Harris (1985), where randomization is not necessary to guarantee the existence of subgame-perfect equilibrium. However, the self-fulfilling nature of expectations creates a similar non-convexity problem as self-fulfilling expectations due to multiple equilibria.

At first sight, randomization between equilibria might seem to call for a new concept of personal equilibrium, in which the ability to randomize is explicitly incorporated into the definition. But randomization is easily captured in the framework of Definition 1 by sufficiently enriching the physical consequence spaces with outcomes that the agent does not care about, but which can be used to randomize between equilibria. More precisely, between any two periods, we insert a "randomization period" with a lottery over physical outcomes that do not affect the agent's utility either directly or through anticipation. The claim, then, is that if the consequence spaces in these randomization periods are sufficiently rich, a personal equilibrium exists. However, instead of this more complicated approach, in the proof below I simply assume at every step that a lottery over personal equilibrium outcomes is also a personal equilibrium outcome. The two approaches are obviously equivalent.

Now we are ready to prove the existence result.

Theorem 1 *Under the conditions of Section 2, if we allow for randomization between equilibria, a personal equilibrium exists.*

Proof. We prove by backward induction that the set of personal equilibrium outcomes starting in period t is a non-empty, convex-valued, closed correspondence of the histories and decision problems $Y^{0t} \times D_t^0$. The difficult part is in establishing the non-emptiness of the equilibrium set; convexity follows from the assumption that randomization between equilibria is possible, and the correspondence has a closed graph by the continuity of u_t and the compactness of decision problems.

The statement is obvious for $t = T$. Now suppose that the statement is true for $t + 1$. By the Kakutani-Fan-Glicksberg theorem (Aliprantis and Border 1994, page 484), for each y^t, z_t, d_{t+1}^0, the set of expectations $\phi_t^c(y^t, z_t, d_{t+1}^0)$ in period t compatible with future decisionmaking is compact and nonempty. Furthermore, since the correspondence of future equilibria is closed, ϕ_t^c is a closed correspondence. For each y^t, z_t, d_{t+1}^0, consider the function

$$\eta(y^t, z_t, d_{t+1}^0) = \max_{f_t \in \phi_t^c(y^t, z_t, d_{t+1}^0)} E_{f_t} u_t(y^t, z_t, f_t, \cdot)$$

and the correspondence

$$\mu(y^t, z_t, d_{t+1}^0) = \{f_t \in \phi_t^c(y^t, z_t, d_{t+1}^0) | E_{f_t} u_t(y^t, z_t, f_t, \cdot) = \eta(y^t, z_t, d_{t+1}^0)\}.$$

By construction, D_{t+1}^0 is compact Polish. Since a closed correspondence between compact Hausdorff spaces is measurable (Aliprantis and Border 1994, Theorem 14.68, page 494), and the functions are bounded by continuity and the compactness of their domains, the measurable maximum theorem (Aliprantis and Border 1994, page 508) applies. Thus, η is measurable and μ admits a measurable selection. Call one such selection ϕ_t.

Notice that since ϕ_t^c is closed, the function η is upper semicontinuous. By theorem 12.4 of Aliprantis and Border (1994), the function $\gamma_t \mapsto \int \eta(y^t, z_t, d_{t+1}^0) d\gamma_t(z_t, d_{t+1}^0)$ from $\Delta(Z_t^0 \times D_{t+1}^0)$ to $\mathbb{R}$ is upper semicontinuous. This means that if the continuation distribution of outcomes is determined by ϕ_t, self t's maximization problem has a solution. This establishes the existence of equilibrium starting in period t. $\square$

B Proofs

Theorem 2 *Suppose that self t is insecurity neutral. Then*

1. *self t is one-period resolution loving if and only if she is disappointment averse;*
2. *if self t is regular disappointment averse, she is one-period information loving.*

Proof. 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.

I prove the converse for each y^t, z_t. For notational simplicity, let $u(\cdot) = u_t(y^t, z_t, \cdot)$. Suppose the agent is facing a probability measure over future outcomes $F \in \Delta(\bar{Z}_t^+)$. If all uncertainty is resolved early, her expected utility is

$$\int_{\bar{Z}_t^+} u(\delta_z, z) dF(z),$$

whereas remaining completely ignorant gives an expected utility of

$$\int_{\bar{Z}_t^+} u(F, z) dF(z).$$

Since the agent is insecurity neutral, the above equals

$$\int_{\bar{Z}_t^+} \int_{\bar{Z}_t^+} u(\delta_\zeta, z) dF(\zeta) dF(z).$$

Noticing that

$$\int_{\bar{Z}_t^+} u(\delta_z, z) dF(z) = \int_{\bar{Z}_t^+} \int_{\bar{Z}_t^+} u(\delta_z, z) dF(\zeta) dF(z),$$

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

$$\frac{1}{2} \int_{\bar{Z}_t^+} \int_{\bar{Z}_t^+} [u(\delta_z, z) - u(\delta_z, \zeta) + u(\delta_\zeta, \zeta) - u(\delta_\zeta, z)] dF(\zeta) dF(z). \tag{7}$$

Since the agent is disappointment averse, we know that 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 7 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.) To do so, we give a geometric interpretation to the integral 7. Since disappointment aversion is regular, we can rewrite 7 as

$$\frac{1}{2} \int_{\bar{Z}_t^+} \int_{\bar{Z}_t^+} g(|c_1(z) - c_1(\zeta)|, \ldots, |c_K(z) - c_K(\zeta)|) dF(z) dF(\zeta). \tag{8}$$

As far as preference for information is concerned, we can think of elements of $\bar{Z}_t^+$ as points in a K-dimensional Euclidean space, where $z \in \bar{Z}_t^+$ corresponds to $(c_1(z), \ldots, c_K(z)) \in \mathbb{R}^K$. On these points, g defines a semimetric. Then, expression 8 is the expected distance of two points independently chosen according to the distribution G, the probability measure F induces on $\mathbb{R}^K$ through the maps c_k. We prove that this distance is concave in G for a symmetric semimetric d which is homogeneous of degree one.

We first simplify the problem in several steps. First, notice that it is sufficient to prove our claim for probability measures over finitely many points–by "lumping" together points that are close to each other into a single point, we can approximate the expected distance arbitrarily closely. More precisely, for any $\epsilon > 0$, we can cover $\mathbb{R}^K$ with disjoint K-dimensional cubes of diameter ϵ. Consider the distribution $G'(\epsilon)$ that, for any cube that intersects the support of F, places the same weight on the center of the cube as G does on the entire cube. By the triangle inequality, as ϵ approaches zero, the expected distance between two points chosen independently according to G' approaches the expected distance between two points chosen independently according to G.

Second, we can further restrict our attention to uniform distributions. For any distribution over finitely many points, we can arbitrarily closely approximate the probabilities placed on each by point with rational numbers of the same denominator. Then, we "subdivide" each point into a number equal to the numerator in its probability. A uniform distribution over these points arbitrarily closely approximates the expected distance.

Third, in considering mixtures between two distributions, we can assume that they are over the same number of points–otherwise, once again we just "subdivide" the points in the support of each distribution.

Fourth, it is sufficient to consider mixtures of distributions with an equal weight on each. It is easy to show that if the expected distance is concave for even mixtures, then it is concave for any mixture–we can approximate any mixing weight arbitrarily closely by repeatedly taking even mixtures.

With these simplifications, the problem becomes the following:

Lemma 1 *For two sets of points $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ in a K-dimensional Euclidean space with a symmetric semimetric d that is homogeneous of degree one (in the vector of coordinate distances), $\sum_{i<j} d(a_i, a_j) + \sum_{i<j} d(b_i, b_j) \leq \sum_{i,j} d(a_i, b_j)$.*

To complete the proof, we prove this Lemma by induction on K. In one dimension, take any point p on the real line and consider how many times p is between pairs of points in set A, how many times it is between pairs of points in set B, and how many times between a pair of points from different sets. Suppose there are l_1 points in set A and l_2 points in set B to the left of p. Thus, p is between $l_1(n - l_1)$ pairs of points from set A, $l_2(n - l_2)$ pairs of points from set B, and $l_1(n - l_2) + l_2(n - l_1)$ pairs of points from different sets. It is sufficient to prove

$$l_1(n - l_1) + l_2(n - l_2) \leq l_1(n - l_2) + l_2(n - l_1),$$

which is true since $l_1^2 + l_2^2 - 2l_1 l_2 = (l_1 - l_2)^2 \geq 0$.

The following lemma is sufficient to establish the inductive step by contradiction:

Lemma 2 *Suppose there are two sets of line segments L_1 and L_2 in a K-dimensional Euclidean space with symmetric semimetric d that is homogeneous of degree one. If the total length of segments in L_1 is greater than the total length of segments in L_2, then there is a projection of these segments into a $K-1$-dimensional space where the same is true.*

For each unit vector, project L_1 and L_2 onto the $K - 1$-dimensional subspace to which the vector is normal. Each of these projections gives a total length for the two sets of segments. Integrate these lengths over the set of unit vectors according to the uniform distribution. Since d is symmetric and homogeneous of degree one, for each individual line segment, the integral is proportional to the length of the line segment. Thus, the integral is greater for L_1 than for L_2. This implies that there must be a projection for which the claim holds. $\square$

Theorem 3 *If the decisionmaker is time consistent, one-period information neutral, predictable, and satisfies non-satiation, then she is observationally equivalent to a time consistent decisionmaker whose utility function depends only on physical outcomes. That is, there is a utility function v defined on $Z_1 \times \cdots \times Z_T$ such that she acts as if she maximized the expectation of v in every decision problem.*

Proof. 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

$$E_{f_t'} u_{t+1}(y^t, z_t, f_t, \cdot) \geq E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot)$$

but

$$E_{f_t'} u_t(y^t, z_t, f_t', \cdot) < E_{f_t} u_t(y^t, z_t, f_t, \cdot).$$

Then, the first inequality must in fact be an equality, otherwise time consistency would be violated. Since the decisionmaker satisfies non-satiation, we can perturb f_t' a little bit to make it strictly preferred to f_t. But by continuity, if the perturbation is small enough, the second inequality will not be reversed, in contradiction to time consistency.

We now prove by backward induction that for each t, there is a function $v_t : Y^t \times \prod_{s=t}^T Z_s \to \mathbb{R}$ such that selves t through T behave as if maximizing the expectation of v_t. This is sufficient since $v = v_1$ establishes the claim of the theorem.

In the proof below, we will take advantage of the following notation:

- For any $f_t \in F_t$, let $p_t(f_t)$ denote the marginal of Z_t^+.
- For any probability measure g over Z_t^+, let $\bar{g}$ be the probability measure in $\Delta(\bar{Z}_t^+)$ such that $\bar{g}(S) = g\left(\{z_t^+ \in Z_t^+ | \bar{z}_t^+ \in S\}\right)$ for any measurable set $S \subset \bar{Z}_t^+$.
- Let $\bar{d}_t^0$ denote the set of marginal probability measures over $\prod_{s=t}^{T-1}(Z_s \times F_s) \times Z_T$ that are feasible given the decision problem d_t^0.

Besides proving the above statement, we also prove that selves t through T behave as if solving $\max_{f_{t-1} \in \bar{d}_t^0} E_{f_{t-1}} u_t(y^t, \cdot)$; to be precise, "behave as if solving" a given problem and "behave as if maximizing" a utility function will mean that the set of continuation personal equilibrium measures is identical to the set of solutions to the problem. For $t = T$, the statement is true by definition of self T's utility function. Now assume that the statement is true for self $t+1$. We prove that it is true for t in several steps.

1. First, we prove by contradiction that for any z_t and d_{t+1}^0, it is a continuation personal equilibrium for selves $t+1$ through T to act as if solving $\max_{f_t \in \bar{d}_{t+1}^0} E_{f_t} u_t(y^t, z_t, f_t, \cdot)$. Suppose f_t solves this maximization problem, and there is an $f_t' \in \bar{d}_{t+1}^0$ such that $E_{p_t(f_t')} v_{t+1}(y^t, z_t, f_t, \cdot) > E_{p_t(f_t)} v_{t+1}(y^t, z_t, f_t, \cdot)$. Then, since future selves

behave as if maximizing the expectation of u_{t+1}, self $t+1$ prefers f_t' over f_t. Then, by time consistency, f_t is not a solution to the above maximization problem.

2. Next, we prove that for any y^t, z_t, and d_{t+1}^0, all continuation personal equilibria give self t the same expected utility. Suppose that f_t and f_t' are the period t expectations in two continuation equilibria. Then, since selves $t+1$ through T behave as if maximizing the expected utility of self $t+1$, we have

$$E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot) \geq E_{f_t'} u_{t+1}(y^t, z_t, f_t, \cdot) \text{ and } E_{f_t'} u_{t+1}(y^t, z_t, f_t', \cdot) \geq E_{f_t} u_{t+1}(y^t, z_t, f_t', \cdot). \tag{9}$$

Now note that there is a decision problem $d_{t+1}^0(f_t) \in D_{t+1}$ such that all selves' choices are trivial in all states of the world, and the unique continuation equilibrium with this decision problem generates expectations f_t in period t. Namely, for each self and any history, we limit her choice to the one she would choose in the equilibrium that generates expectations f_t. Define $d_{t+1}^0(f_t')$ similarly. Then, the decision problem $d_{t+1}^0(f_t) \cup d_{t+1}^0(f_t')$ satisfies the condition that only self $t+1$ has a choice, and one of her choices leads to the probability measure f_t over outcomes starting in period $t+1$, and the other leads to f_t'. Using inequalities 9, this implies that for y^t, z_t, and the decision problem $d_{t+1}^0(f_t) \cup d_{t+1}^0(f_t')$, f_t and f_t' are both possible continuation equilibrium expectations. Then by predictability, inequalities 9 should in fact hold with equality; that is,

$$E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot) = E_{f_t'} u_{t+1}(y^t, z_t, f_t, \cdot) \text{ and } E_{f_t'} u_{t+1}(y^t, z_t, f_t', \cdot) = E_{f_t} u_{t+1}(y^t, z_t, f_t', \cdot). \tag{10}$$

Now by the above definition of time consistency, the first part of equation 10 implies

$$E_{f_t} u_{t+1}(y^t, z_t, f_t, \cdot) \leq E_{f_t'} u_{t+1}(y^t, z_t, f_t', \cdot),$$

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

3. From the above two claims, we can conclude that *all* continuation personal equilibria have the property that they maximize self t's expected utility (that is, they solve $\max_{f_t \in \bar{d}_{t+1}^0} E_{f_t} u_t(y^t, z_t, f_t, \cdot)$).

4. Therefore, selves t through T behave as if solving $\max_{f_{t-1} \in \bar{d}_t^0} E_{f_{t-1}} u_t(y^t, \cdot)$.

5. For any $f_t, f_t' \in F_t$ such that $p_t(f_t) = p_t(f_t')$, we clearly have

$$E_{p_t(f_t')} v_{t+1}(y^t, z_t, f_t, \cdot) = E_{p_t(f_t)} v_{t+1}(y^t, z_t, f_t, \cdot) \text{ and } E_{p_t(f_t')} v_{t+1}(y^t, z_t, f_t', \cdot) = E_{p_t(f_t)} v_{t+1}(y^t, z_t, f_t', \cdot).$$

Therefore, since future selves behave as if maximizing self $t+1$'s utility, self $t+1$ is indifferent between these two distributions as well. By the above version of time consistency, this means that self t is indifferent between f_t and f_t' as well. In particular, self t is indifferent between any f_t and an alternative one that has the same marginal over Z_t^+ and all uncertainty is resolved in period $t+1$. Formally, for any $f_t \in F_t$,

$$E_{f_t} u_t(y^t, z_t, f_t, \cdot) = E_{\overline{p_t(f_t)}} u_t(y^t, z_t, \overline{p_t(f_t)}, \cdot).$$

6. Let $v_t(y^t, z_t, z_t^+) = u_t(y^t, z_t, \delta_{\bar{z}_t^+}, \bar{z}_t^+)$. By one-period information neutrality,

$$E_{\overline{p_t(f_t)}} u_t(y^t, z_t, \overline{p_t(f_t)}, \cdot) = E_{p_t(f_t)} v_t(y^t, z_t, \cdot).$$

Thus, we have proven that selves t through T behave as if maximizing v_t, satisfying the requirements of the inductive step. □

Theorem 4 *Suppose that the decisionmaker is time consistent and satisfies non-satiation. Furthermore, suppose that for each* $1 < t \leq T$, u_t *is additively separable in past expectations and other outcomes. Then there are functions* $v_t : Z_1 \times \cdots \times Z_t \times D_{t+1} \to \mathbb{R}$ *such that i.) given* $z_1, \ldots, z_{t-1}$, *self* t *bahaves as is maximizing the expectation of* v_t, *and ii.) for any* $d_{t+1}, d'_{t+1} \in D_{t+1}$ *and* $z^{t+1} \in Z_1 \times \cdots \times Z_t$,

$$v_t(z^{t+1}, d_{t+1}) \geq v_t(z^{t+1}, d'_{t+1}) \Leftrightarrow \max_{l_{t+1} \in d_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot) \geq \max_{l_{t+1} \in d'_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot).$$

Proof. We use a similar induction argument to that in the proof of Theorem 3. That is, we prove both that the statement is true starting in period t and that selves t through T behave as if maximizing the expected utility of self t constrained by rationality. The statement is obviously true for period T. Suppose now that it is true for period $t+1$.

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, step 1 of the proof of that theorem carries over as well, so for any y^t, z_t, and d^0_{t+1}, there is a continuation personal equilibrium that maximizes self t's expected utility constrained by rationality. Now, since future selves behave as if maximizing the expected utility of self $t+1$, and that utility is separable from past expectations, if f_t and f'_t are period t expectations in two continuation equilibria, the equalities in 10 are automatically satisfied. By time consistency, all continuation equilibria give self t the same expected utility.

Therefore, just as in Theorem 3, we conclude that selves t through T behave as if maximizing the expected utility of self t. By the separability condition, this utility function leads to identical behavior as one of the form

$$\hat{v}_t(z^{t+1}, f_t, z_{t+1}, f_{t+1}, \ldots, z_T, f_T).$$

Let

$$v_t(z^{t+1}, d^0_{t+1}) = \max_{f_t \in \bar{d}^0_{t+1}} E_{f_t} \hat{v}_t(z^{t+1}, f_t, \cdot).$$

By time consistency

$$\max_{f_t \in \bar{d}^0_{t+1}} E_{f_t} \hat{v}_t(z^{t+1}, f_t, \cdot) \geq \max_{f_t \in \bar{d}^{0\prime}_{t+1}} E_{f_t} \hat{v}_t(z^{t+1}, f_t, \cdot) \Leftrightarrow \max_{f_t \in \bar{d}^0_{t+1}} E_{f_t} \hat{v}_{t+1}(z^{t+1}, \cdot) \geq \max_{f_t \in \bar{d}^{0\prime}_{t+1}} E_{f_t} \hat{v}_{t+1}(z^{t+1}, \cdot). \tag{11}$$

Now by self $t+1$'s decision problem, and since selves $t+1$ through T act as if maximizing the expected utility of self $t+1$,

$$\max_{l_{t+1} \in d^0_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot) = \max_{f_t \in \bar{d}^0_{t+1}} E_{f_t} \hat{v}_{t+1}(z^{t+1}, \cdot), \tag{12}$$

and similarly for $d^{0\prime}_{t+1}$. Combining 11 and 12, we get exactly

$$v_t(z^{t+1}, d^0_{t+1}) \geq v_t(z^{t+1}, d^{0\prime}_{t+1}) \Leftrightarrow \max_{l_{t+1} \in d^0_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot) \geq \max_{l_{t+1} \in d^{0\prime}_{t+1}} E_{l_{t+1}} v_{t+1}(z^{t+1}, \cdot). \ \square$$

References

Akerlof, G., and W. T. Dickens (1982): "The Economic Consequences of Cognitive Dissonance," *The American Economic Review*, 72(3), 307–319.

Aliprantis, C. D., and K. C. Border (1994): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Berlin: Springer-Verlag.

Barro, R. J. (1974): "Are Government Bonds Net Wealth?," *Journal of Political Economy*, 82(6), 1095–1117.

Bénabou, R., and J. Tirole (2002): "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, forthcoming.

Bergstrom, T. C. (1999): "Systems of Benevolent Utility Functions," *Journal of Public Economic Theory*, 1(1), 71–100.

Bernheim, B. D., and D. Ray (1987): "Economic Growth with Intergenerational Altruism," *Review of Economic Studies*, 54(2), 227–241.

Caplin, A., and J. Leahy (2001): "Psychological Expected Utility and Anticipatory Feelings," *Quarterly Journal of Economics*, 116(1), 55–80.

Carillo, J. (1997): "Self-control, Moderate Consumption, and Craving," Mimeo.

Carillo, J., and T. Mariotti (2000): "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies*, 67, 529–544.

Chew, S. H., and L. G. Epstein (1989): "The Structure of Preferences and Attitudes Towards the Timing of the Resolution of Uncertainty," *International Economic Review*, 30(1), 103–117.

Dufwenberg, M., and G. Kirchsteiger (1998): "A Theory of Sequential Reciprocity," Tilburg CentER for Economic Research Discussion Paper: 9837.

Eliaz, K., and R. Spiegler (2002): "Anticipatory Feelings and Attitudes to Information," Mimeo, New York University; URL: http://homepages.nyu.edu/ ke7/jeea.pdf.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Fisher, I. (1930): *The Theory of Interest*. New York: Macmillan.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1(1), 60–79.

Gervais, S., and T. Odean (2001): "Learning To Be Overconfident," *Review of Financial Studies*, 14(1), 1–27.

Grant, S., A. Kajii, and B. Polak (1998): "Intrinsic Preference for Information," *Journal of Economic Theory*, 83, 233–259.

——— (2000): "Preference for Information and Dynamic Consistency," *Theory and Decision*, 48, 263–286.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59(3), 667–686.

Harris, C. (1985): "Existence and Characterization of Perfect Equilibrium in Games of Perfect Information," *Econometrica*, 53(3), 613–628.

Harris, C., P. Reny, and A. Robson (1995): "The Existence of Subgame-Perfect Equilibrium In Continuous Games With Almost Perfect Information: A Case For Public Randomization," *Econometrica*, 63(3), 507–544.

Hori, H., and S. Kanaya (1989): "Utility Functionals with Nonpaternalistic Intergenerational Altruism," *Journal of Economic Theory*, 49(2), 241–265.

Jones, R. A. (1977): *Self-fulfilling Prophecies: Social, Psychological, and Physiological Effects of Expectancies*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–291.

Kőszegi, B. (2001): "Anxiety, Doctor-Patient Communication, and Health Education," Mimeo, UC Berkeley; URL: http://emlab.berkeley.edu/users/botond/doctors.pdf.

Kőszegi, B., and M. Rabin (2002): "Modeling Reference-Dependent Preferences," Mimeo, UC Berkeley; in progress.

Kimball, M. S. (1987): "Making Sense of Two-Sided Altruism," *Journal of Monetary Economics*, 20(2), 301–326.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 97, 666–684.

——— (1992): "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice," in *Choice Over Time*, ed. by G. Loewenstein, and J. Elster, pp. 19–20. Russell Sage Foundation, New York.

Manove, M., and A. J. Padilla (1999): "Banking (Conservatively) with Optimists," *Rand Journal of Economics*, 30(2), 324–350.

Mehra, R., and E. Prescott (1984): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Mellers, B. A. (2000): "Choice and the Relative Pleasure of Consequences," *Psychological Bulletin*, 126, 910–924.

Mellers, B. A., A. Schwartz, and I. Ritov (1999): "Emotion-Based Choice," *Journal of Experimental Psychology: General*, 128(1), 1–14.

Rabin, M., and J. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 114(1), 37–82.

Samuelson, P. (1939): "The Rate of Interest Under Ideal Conditions," *Quarterly Journal of Economics*, 53(2), 286–297.

Skiadas, C. (1998): "Recursive Utility and Preferences for Information," *Economic Theory*, 12, 293–312.

Weil, P. (1989): "The Equity Premium Puzzle and the Riskfree Rate Puzzle," *Journal of Monetary Economics*, 24(3), 401–421.