Sweet self-deception

Dag Einar Sommervoll
BI Norwegian Business School

This is the author's final, accepted and refereed manuscript to the article published in

Journal of Economics, **109(2013)1: 73-88**

DOI: http://dx.doi.org/10.1007/s00712-012-0308-2

Sweet Self-deception

JEL classification: A12, B49, C70, D11, D60, D74, D91, E21

Abstract

People have a tendency to procrastinate when faced with aversive tasks - but they also procrastinate in relation to beneficial matters whose rewards are instantaneous. If agents value present anticipations of future consumption, revision of consumption plans may be viewed as a benign form of self-deception. We consider a minimal generalization of the Samuelson discounted utility model to allow for utility linked to next period consumption. Agents are assumed to vary with respect to their sophistication. In this context, commitment and self-control are obstacles to the pursuit of increased utility. We also examine different environments that are likely to facilitate repeated revisions.

Key words: Intertemporal choice; self-deception; time inconsistency; naivete; self-control; discounted utility functions; anticipation; memory

1 Introduction

We all make plans. Plans involve actions of our future selves. Putting off vacuuming the house may please today's you, but may not go down too well with the you of tomorrow. Indeed, come tomorrow, you might well find yourself looking for an excuse to postpone the chore, letting some future you take responsibility. But as we all know, procrastination, even over seemingly menial tasks, can snowball into significant welfare losses.[1] This paper reverses this proposition, however. What if your failure to commit to yesterday's plan actually produces higher utility? If plans for future consumption offset anticipations of present value, revising a plan may increase utility all told. You may in fact find it better to consume less today than you had planned or anticipated yesterday. In this scenario, the well-known self-control problem of procrastination becomes one of self-deception. And the question is not how to improve your self-commitment record, but to find an environment which helps you revise your plans.

Now self-deception could be said to be a contradiction in terms. It seems to rely on a logical inconsistency, knowing and not knowing at the same time. And self-deception can be extremely harmful in a competitive environment. Trivers (2011) takes an opposite view. Self-deception, he contends, is widespread in the animal kingdom and there is an evolutionary push toward self deception. The main explanation for self-deception, he adds, is that you fool yourself for better to fool others. There is, moreover, an immunological upside of an overly positive self-evaluation. True or not, neither explanation has any immediate bearing on economic theory. What we want to do in this paper is to show that beneficial self-deception occur in one-agent intertemporal consumption scenarios. In other words, self-deception for your own good is not limited to the animal kingdom; we need to add homo economicus to the list as well.

[1] See O'Donoghue and Rabin (1999) for a discussion and extensive literature review.

The following example provides the intuition. Sheila plans to make a cake last three days. On the first day she cuts it into three pieces of different sizes. While eating the first day's ration, she savors the thought of tomorrow's helping, the larger of the two remaining two slices. But when tomorrow comes, Sheila discovers that she'd prefer to save the biggest piece for the final day because she'll have an extra day to mull over the enticing prospect of the big slice to come.

It is important to note the conflict of interest between the Sheila of Day 1 and the Day 2 counterpart. For Day 1-Sheila, her commitment to her plan is arguably essential for the utilities she enjoys on the first day. However, Day 2-Sheila's commitment to yesterday's plan is less evident. After all, Sheila may argue, yesterday is history, and so is Day 1-Sheila. What concerns her now is getting as much utility as possible from the two remaining slices of cake. She could save the biggest slice for third and final day, of course. But then again, her second day self may begin to waver in her commitment to the plan, or feel remorse for deviating from it. This may or may not be sufficient to induce Sheila to proceed as planned. The key observation is that whatever Day 2-Sheila does, Day 1-Sheila is not worse off.

In short, there is a potential upside of non-commitment in this intertemporal consumption problem. But that immediately raises the question of why the first day Sheila can believe that she will consume the biggest piece the following day, when it is better for the second day Sheila to deviate. There must be some kind of self-deception at work. This is where we are heading.

Our starting point is the much used Samuelson's discounted utility model (Samuelson 1937). As this is a discounted sum of instantaneous utility functions, it involves a premise of total amnesia, as if we were to plan a sequence of sensations of no value to us except in the heat of the moment. In medical parlance, the economic agent invoked by Samuelson's DU model is suffering from Korsakoff's syndrome.[2] The paper generalizes Samuelson's DU model to allow for utility from memories and anticipations. Two types of agents are considered. The first has preferences as implied by the standard Samuelson utility model (0-korsakoffs); the other a rudimentary memory (1-korsakoffs). A 1-korsakoff remembers a consumption plan the time period it is made, and may harbor (positive) anticipations regarding the next period's consumption. But she has no recollection of past consumption or yesterday's plan. A 1-korsakoff can be viewed as the first step towards an agent with full recollection. This rudimentary memory affects intertemporal choice. 1-korsakoffs tend to be time-inconsistent, and repeated revisions may cut a path to higher long-run utility.

This paper connects with the literature on self-deception and self-control. Gul and Pesendorfer's (2001) seminal contribution shifted attention of inconsistencies of intertemporal choice away from preferences over consumption bundles, to preferences over a class of decision problems. Their axiomatic treatment of preferences allows for "Set Betweenness," i.e. a preference for limiting future choices. This provides fertile ground for studying temptations and the cost of self-control. Noor (2007) found the implications of future temptations to be mixed. Agents who are aware of their self-control problems may not take advantage of commitment opportunities. The possibility of indulging temptation in the future is itself a source of temptation. Kopylov and Noor (2010) consider self-deception in a fairly explicit form, because their model allows for the agent to rationalize actions that eventually will lead to temptation. Arguing along a similar vein, Sarver's(2008) agents seeks to reduce potential regret. Halevy (2008) studies diminishing impatience and argues that positive time preference is deeply connected to uncertainty. Fudenberg and Levine

[2] Korsakoff's syndrome: Neurological disorder marked by severe amnesia despite clear perception and full consciousness, resulting from chronic alcoholism, head injury, brain illness, or thiamin deficiency. Affected persons typically fail to remember events in the recent or even immediate past; some retain memories for only a few seconds. Sufferers are also likely to forget longer periods –up to 20 years. (Encyclopedia Britannica)

(2012) show also that timing is essential when it comes to understanding the cost of self-control. An agent may resist a one shot temptation, but give in to temptation if it is permanent or extends over a number of time periods.

The economic agents that we shall be considering bear a closer resemblance to those modeled by Loewenstein(1987), Caplin and Leahy (2001), Bernheim and Thomadsen (2005), Epstein (2008) and Kőszegi (2010), inasmuch as anticipatory feelings affect intertemporal choice. The model presented here contrasts with this literature, as commitment and self-control may stand in the way of utility-increasing revisions of consumption plans. The self-deception considered here relies on the agent not fully understanding her future preferences and actions, and partly resembles in structure potential information aversion as considered by Epstein(2008). Moreover, our results also differ from the game theory approach of Asheim (1997). He introduces revision-proof strategies as a refinement of subgame-perfectness. If an agent becomes aware of her self-control problems, she can limit her decision problem by only considering plans she will actually follow. In our model, self-deceiving behavior is conditional on non-revision-proof consumption plans.

This paper's contribution to the study of self-deception and intertemporal choice is threefold. First, it extends the Samuelson DU model so as to facilitate a comparison of agents with differing mnemonic capacities. Second, it considers a simple, but critical distinction where anticipation of consumption is connected to plans of consumption, not future consumption per se. Third, it adds to the potential downside of commitment and self-control.

The paper is organized as follows. In section 2 we define k-korsakoff preferences in a T-period consumption scenario of one continuous good. Section 3 considers consumption of one indivisible good that is to be consumed over a given number of time periods. The discussion draws heavily on the agent types, naïfs and sophisticates, introduced by O'Donoghue and Rabin (1999).

Whereas naivete tends to yield poor long-run outcomes, the results presented here extend the potential downside of sophistication. Sophisticates have less opportunity to revise repeatedly, and as a result may to get lower long-run utility in comparison with naïfs. In section 4 we discuss a model, the life span uncertainty model, which provides an example of benign self-deception among sophisticates. The model is an abstraction of the following scenario. Upon retirement we decide to spend a year in Rio. But which year? While we know we will not live forever, we do not know how many years we have left. As time passes and our general health deteriorates, we want to go before it is too late. In a stylized version of this consumption problem, we show that sophisticates can achieve the same long-run utility as naïfs. In other words, a little uncertainty regarding the number of time periods facilitates utility increasing revisions even for sophisticates. Section 5 concludes.

2 The DU model with memory and anticipation

In this section we extend Samuelson's discounted utility model by including utility of memories of past consumption and anticipations of planned consumption.

Definition 2.1 *Any consumption vector* $(c_1, \ldots, c_T)$ *give rise to a vector of memories* $(0, m_2, \ldots, m_T)$*, where* m_i *is a function of past consumption, that is* $m_i = m_i(c_{<i})$*, where* $c_{<i} = (c_1, \ldots, c_{i-1})$ *for all* $i > 1$*.*

Anticipation utility is linked to expected consumption, thus in the context of intertemporal choice, *plans* of future consumption. We can formalize anticipations like this:

Definition 2.2 *Any consumption plan represented by a planned consumption* *vector* $(c_{1,plan}, \ldots, c_{T,plan})$[3] *give rise to a vector of anticipations* $(a_1, \ldots, a_T)$ *where* a_i *is a function of future planned consumption, that is* $a_i = a_i(c_{>i,plan})$ *where* $c_{>i,plan} = (c_{i+1,plan}, \ldots, c_{T,plan})$ *for all* $i \geq 1$.

We can view the anticipations of Definition 2.2 as first order anticipations, i.e. anticipations about consumption. It could be argue that people also have anticipations about memories ($a_i(m_j)$), memories of anticipations ($m_j(a_i)$) and anticipations of anticipations ($a_i(a_j)$) et cetera. These we can view as second or higher order memories and anticipations. Here, we consider only first order memories and anticipations.

A straightforward generalization of a Samuelson's DU model with anticipations and recollections is:

$$U = \sum_{i=1}^{T} \delta^{i-1} u(c_i, m_i(c_{<i}), a_i(c_{>i,plan})). \tag{2.1}$$

where u is interpreted as an instantaneous utility function depending on present consumption, memories of past consumption, and anticipations related to future consumption.

The utility function (2.1) places no limitation on utility from anticipation or recollection. In the following we define preference types by the extent to which they receive utility from memory and anticipation.

Let a person consider an T-period consumption scenario. We will assume that the preferences of all agents can be represented by a DU utility function, but may differ according to their (korsakoff) type. A k-korsakoff has a consumption horizon of k periods into the future and a recollection of the last k periods of consumption, including the present time period. This is formalized in the

[3] The reason for the subscript 'plan' is to make the distinction between consumption and planned consumption explicit.

following definition:

Definition 2.3 *An agent is k-korsakoff, k a nonnegative integer, if the following the following two conditions hold:*

I. $m_i(c_1, \ldots, c_{i-1}) = m_i(c_{max(1,i-k+1)}, \ldots, c_{i-1})$ *if* $k, i > 1$ *and* $m_i = 0$ *otherwise.*

II. $a_i(c_{i+1,plan}, \ldots, c_{T,plan}) = a_i(c_{i+1,plan}, \ldots, c_{min(T,i+k),plan})$ *if* $k \geq 1$ *and* $a_i = 0$ *otherwise.*

A few comments are in order. A 0-korsakoff has $a_i = m_i = 0$, and is the economic agent implied by the standard DU function(DU_0). A 1-korsakoff has an instantaneous utility function given by $u = u(c_i, 0, a_i(c_{i+1}))$. It represents the crudest sense of memory, remembering the consumption plan in the time period in which it was made. In other words, if a 1-korsakoff plans on eating one hamburger every day, she cherishes the thought of tomorrow's burger, while eating today's. In the next step up, the 2-korsakoff remembers last periods consumption, enjoys this period's consumption and looks forward to the planned consumption of two next time periods. In this paper we shall largely be concerned with 0-korsakoffs and 1-korsakoffs. [4]

Definition 2.4 *(Discounted utility functions for k-korsakoffs) The intertemporal utility function of a k-korsakoff is given by*

$$DU_k = \sum_{i=1}^{T} \delta^{i-1} u(c_i, m_i(c_{max(1,i-k-1)}, \ldots, c_{i-1}), a_i(c_{i+1,plan}, \ldots, c_{min(T,i+k+1),plan}) \tag{2.2}$$

where $\delta < 1$ *and* u*,* m_i *and* a_i *are continuously differentiable unbounded func-*

[4] Example A.1 in the appendix is of a 3-korsakoff. Bernheim and Thomadsen (2005) study economic decision making with memory and anticipation. Their economic agents may forget their past actions, and in this sense resemble our 1-korsakoffs.

tions with positive first order partial derivatives and negative second order partial derivatives with respect to all arguments.

Definition 2.4 states that the time invariant utility function u satisfies standard assumption of nonsatiation and diminishing marginal utility of consumption.[5] It also states that the same preference structure applies to consumption, memories and anticipations. Note that a 1-korsakoff takes account of (discounted) future anticipations, since the utility function u, is in all terms of the discounted sum. To increase readability and limit notational complexity, we suppress the subscript i in most cases, and write $DU_k = DU_k(c, m, a)$.

We use the following definition of time-consistent preference:

Definition 2.5 *Let $c_{plan} = (c_{1,plan}, \ldots, c_{T,plan})$ and $c^{\star}_{plan} = (c^{\star}_{1,plan}, \ldots, c^{\star}_{T,plan})$ be two consumption vectors that agree up to time j, that is $c_{i,plan} = c^{\star}_{i,plan}$ for $i \leq j$. Consider a consider at a discounted utility function DU_k, and let $DU_k(i) = DU_k(c, m(c), a(c))$ and $DU^{\star}_k(i) = DU_k(c^{\star}, m(c^{\star}), a(c^{\star}))$ denote the utility of c_{plan} and $c^{\star}_{plan}$ at time i respectively. A function, DU_k, is said to a represent time-consistent preference if $DU_k(i) > DU^{\star}_k(i)$ for some $i \leq j$ imply $DU_k(i) > DU^{\star}_k(i)$ for all $i \leq j$.*

This definition is just a formalized way of saying the following. Imagine comparing two consumption plans, A and B, for a week starting Monday. They are exactly the same until Thursday. You find A better than B on Monday. If you have time-consistent preferences, you will find A better than B on Tuesday, and Wednesday as well. Preferences that are not time-consistent are said to be time-inconsistent.

The following theorem states that 1-korsakoffs are time-inconsistent.

[5] Note that this utility function implies that all selves has the same utility function over the stream of physical outcomes and expectations. In this respect the utility function considered here is connected to the dynamic model of Kőszegi (2010).

Theorem 2.1 *An agent with DU_1-preferences is time-inconsistent.*

A proof of the theorem is given in the Appendix. The dynamic inconsistency for a 1-korsakoff comes into play because she picks a plan that involves her consuming more in the next period than she actually ends up consuming. [6] Compared to a 0-korsakoff, who is always time-consistent, a 1-korsakoff skews consumption towards the future. [7] The time inconsistencies that arise for a 1-korsakoff are structurally similar to those implied by hyperbolic discounting. The following example illustrates this point.

Example 2.4 An intertemporal utility function for a 1-korsakoff

Let $u(c_i, a(c_{i+1})) = c_i^{\frac{1}{2}} + a(c_{i+1}) = c_i^{\frac{1}{2}} + kc_{i+1}^{\frac{1}{2}}$, where k is a (positive) constant. This gives the following n-period DU_1-utility function:

$$DU_1(c_{plan}) = \sum_{i=1}^{n} \delta^{i-1} u(c_i, a(c_{i+1,plan})) = c_{1,plan}^{\frac{1}{2}} + (\delta + k) \sum_{i=2}^{n} \delta^{i-1} c_{i,plan}^{\frac{1}{2}}$$

The last sum is $\delta(1 + k)$ times the DU_1 at time two (provided) $c_1 = c_{1,plan}$ in period 1. This relation allows us to make two observations. In this easy case, where utility from anticipation is proportional to utility from consumption, the utility function is structurally equal to an intertemporal utility function with hyperbolic discounting. Present-biased preferences tend to be modeled by an intertemporal utility function of the following type: $U_i(u_i, \ldots, u_T) = \delta^i u_i + \beta \sum_{\tau=i+1}^{n} \delta^\tau u_\tau$. [8] In our model $(\delta + k)$ plays the role of β. Present-

[6] Loewenstein (1987) consider a formal model where a person's instantaneous utility function takes the form $u(c_\tau; c_{\tau+1}, c_{\tau+2}, \ldots)$ where the partial derivatives with respect to $c_{\tau'}$ is positive for all $\tau' > \tau$. Loewenstein proposes the following functional form: $u(c_\tau; c_{\tau+1}, c_{\tau+2}, \ldots) = v(c_\tau) + \alpha(\gamma v(c_{\tau+1}) + \gamma^2 v(c_\tau) + \ldots)$ for some $\gamma < 1$. DU anomalies may also occur in Loewenstein's model.

[7] Loewenstein(1987) reports from a study of undergraduates. They were asked to state the 'most they would pay now' for a kiss from their favorite movie star. They could receive the kiss immediately or later (four possible delays). The students went for the 'three day delayed' kiss. A more recent study by Shu and Gneezy (2010) includes experiments and field studies showing procrastination of enjoyable experiences such as visits to fashionable restaurants and going to the movies.

[8] See O'Donoghue and Rabin (2001), for a discussion and an extensive literature review.

biased preferences and potentially myopic behavior arise when $\beta < 1$. In contrast, since we assume $k > 0$ are 1-korsakoffs potentially future-biased, and hyperopic.[9]

A valuation of anticipation of next period consumption implies giving more weight to future consumption, compared to absence of valued anticipation. The effect of the discount factor goes the other way: less weight is put on future consumption compared to present consumption. If $k + 1 = \delta^{-1}$, then anticipation completely balances the effect of discounting for consumption in time period 2, but k has no effect on the weight put on consumption in time period 3 compared to consumption in time period 4.

From a welfare perspective, time-inconsistent behavior could have some serious implications. O'Donoghue and Rabin (1999) show that welfare losses can be associated both with procrastination as well as preproperation. An analysis of long-run consequences requires some notion of long-run utility of consumption which allows us to compare different paths of consumption either ex ante or ex post. In the following analysis we rely on the same long-run utility definition as O'Donoghue and Rabin (1999). Long-run utility from consumption is here defined as the sum of utility in all time periods without discounting. This may be viewed as an ex ante utility where all future and past selves are considered equal and this utility is assigned to a time period 0.[10] This utility concept captures the possibility of achieving higher long-run utility by having wrong expectations of future behavior.

Example 2.5 Long-run utility for a 1-korsakoff

Assume that a 1-korsakoff has made a plan to eat a cake over a four day period. Her preferences are given by the intertemporal utility function of Example

[9] If we allow for negative anticipations, that is $k < 0$, then this 1-korsakoff model corresponds to a standard hyperbolic DU model with $\beta = \delta + k$ $(k > -\delta)$xxxxtenkigjennom dette med retting.

[10] See Goldman (1979) for a brief discussion of different approaches regarding long-run utility

2.4. We assume that $k = 1$ and $\delta = \frac{1}{2}$. In other words, this 1-korsakoff has preferences for anticipations, while at the same time discounting the future. The utility maximizing plan at Day 1 is given in the first row of Table 2.1. The next row gives the best consumption plan at Day 2, given that our 1-korsakoff ate as much cake as planned the first day. Likewise row 3, given that she consumed on Day 1 and Day 2 in accordance with the best intertemporal plan at Day 1 and Day 2 respectively.

Table 2.1 Consumption plans for 1-korsakoff eating a cake over a four day period. Consumption is given in percent.

Consumption	Day 1	Day 2	Day 3	Day 4
Best plan Day 1	58	32	8	2
Best plan Day 2	58	25	14	3
Best plan Day 3	58	25	11	6

Now the best plan at Day 1 prescribes higher consumption on Day 2 than actually is chosen on Day 2, and likewise for Day 2 and Day 3. Table 2.2 gives the corresponding utility levels for each time period. Some of the utility of 13.3 on the first day is anticipation utility (5.7). The following day, consumption is not 32 but 25. The anticipation utility associated with consumption of 25 the following day is 5.0. In other words, our Day 1 1-korsakoff enjoys an excess utility of $5.7 - 5.0 = 0.7$ due to the difference between actual and planned consumption in time period 2.

Table 2.2 Utility associated with consumption and consumption plans for a 1-korsakoff intending to eat cake over a four day period. The long-run utility is the sum of the time period utility levels. (xxxusikker p om denne formuleringen holder.

Consumption	Day 1	Day 2	Day 3	Day 4	Long-run utility
Best plan Day 1	13.3	8.5	4.3	1.4	27.5
Best plan Day 2	13.3	8.7	7.3	1.9	31.2
Best plan Day 3	13.3	8.7	5.9	2.5	30.4

We see also that long-run utility is up from 27.5 to 31.2. The revision at Day 3, does not, however, give higher long-run utility, illustrating the potential downside of delayed consumption that affects the last day. The Day 4 1-korsakoff has no cake-eating future. She cannot enjoy consumption anticipation, and she cannot, unlike her previous selves, postpone consumption. In other words, she is at the mercy of the Day 3 1-korsakoff, who can save, and in this case indeed does save, more cake to the last day than is optimal from a long-run perspective. In this numerical example the effect of the second-to-last day revision is small, but can be considerable for small δ's and large k's.

The above example also highlights a general property of these revisions of consumption plans. Revisions tend to increase long-run utility. Whereas the long-run utility measure can be challenged, the immediate benefit of the revisions cannot. The revisions themselves constitutes Pareto improvements. Saving more cake than initially planned, leaves more cake for subsequent selves and increases the utility of the present self.

3 One indivisible good and degrees of sophistication

O'Donoghue and Rabin (1999) consider agent types differing with respect to sophistication. Self-deception in their model occurs because naive agents fail to fully take into account the preferences and actions of their future selves. Sophisticated agents, on the other hand, do. In short, self-deception is attributed to a kind of limited cognitive ability. In the following discussion we shall adopt this notion of self-deception and draw heavily on the framework developed in O'Donoghue and Rabin (1999).

We start with a motivating example:

Example 3.1 A 1-korsakoff going to Rio.

Sheila has won a trip to Rio and can choose either to go this year, next year or the year after. We assume her preferences are given by a DU_1-model with anticipation that is $DU_1 = \sum_i \delta^{i-1} u(c_{i,plan}, a(c_{i+1,plan}))$. We also assume for simplicity's sake $u = c_i + a(c_{i+1,plan})$. At the beginning of year 1, she faces three possible consumption plans with the corresponding utilities:

$DU_1(c, 0, 0) = c$

$DU_1(0, c, 0) = a(c) + \delta c$

$DU_1(0, 0, c) = \delta a(c) + \delta^2 c,$

If $c > a(c) + \delta c$, going the first year is the preferred alternative. If not, that is, if $c < a(c) + \delta c$, postponing the Rio trip to the second year is the better alternative. Since $\delta < 1$, going in the second year is always better than waiting to the third year. Moreover, if $c < a(c) + \delta c$, and she does not go the first year, she may postpone yet again, and go the third and last year. She may do this even if going the first year is strictly better ($c > \delta a(c) + \delta^2 c$) than going the third year.

This example spurs the following questions. What if Sheila performs the above computation? Why compare the first year and second year alternatives, if she already knows she will not be going the second year? To address these question we introduce two types of agent, naïfs and sophisticates.

The definitions of naïfs and sophisticates are given in terms of strategies. A strategy is an assignment of an action for every contingency. In the case of consumption of one indivisible good, this is either consume (Y) or not consume (N).

Definition 3.1 *A perception-perfect strategy for naïfs at time i is a strategy, $s^n = (s^n_1, s^n_2, \ldots, s^n_T)$ that satisfies $s^n_i = Y$ if $DU_k(i,i) = max(DU_k(i,j))$ for all $j \geq i$, where $DU_k(i,j)$ denotes the utility at time i of consuming the indivisible good at time j.*

Definition 3.2 *A perception-perfect strategy for sophisticates is a strategy, $s^s = (s^s_1, s^s_2, \ldots, s^s_T)$ that satisfies for all $i < T$: $s^s_i = Y$ if and only if $DU_k(i,i) \geq DU_k(i,j)$ for $j > i$ such that $j = min_{k>i}\{k|s^s_k = Y\}$.*

Sophisticates only compare the utility of consuming today with later consumption times, for which consumption has been planned if reached. That is, they ignore irrelevant alternatives. Naïfs, on the other hand, make comparisons with all later consumption dates whether consumption is likely or not at those times. Added to this, in a T period setting, the construction of the game requires that $s^n_T = Y$. Consumption must occur in the final period, if not before.[11]

[11] These definitions are the same as those in O'Donoghue and Rabin (1999). However, the implementation in T-period games differs for our sophisticates. An O'Donoghue-Rabin sophisticate does not discount as assumed under the Samuelson DU model. She is present-biased in the sense of only differentiating between now or later, but not between two later periods. Later is just later. In this analysis we use the standard discounting inferred by the DU model, and in this case (see Theorem 3.2), is a sophisticate time-consistent. In other words, a perception-perfect strategy for sophisticates has more bite in the case of standard discounting.

In Example 3.1, if Sheila is a naïf, she will choose to postpone travelling until the third year if $c < a(c) + \delta c$. What she fails to realize, though, is that while year two is better for Year-1 Sheila, it is not the best for Year-2 Sheila. Year-2 Sheila will transfer the sandy beaches of Rio to Year-3 Sheila. But if Sheila were a sophisticate, she would know that Year-2 Sheila won't be going to Rio. So she will only compare going the first year against going the third year. If $c > \delta a(c) + \delta^2 c$, she will go the first year. The results of this example are special cases of the two following theorems.

Theorem 3.1 *Let C be one indivisible good that is to be consumed in one of the T time periods. Consider a 1-korsakoff agent with utility function $DU_1 = \sum_i^T \delta^{i-1} u(c_i, a(c_{i+1,plan}))$. If she is a naïf, she will be time-consistent if and only if $u(C, 0) \geq u(0, a(C)) + \delta u(C, 0)$.*

Theorem 3.2 *Let C be one indivisible good that is to be consumed in one of the T time periods. Consider a 1-korsakoff agent with utility function $DU_1 = \sum_i^T \delta^{i-1} u(c_i, a(c_{i+1,plan}))$. If she is a sophisticate, she will be time-consistent.*

In sum, these theorems tell us that although sophisticates may have time-inconsistent preferences, they make time-consistent plans when they take into account the preferences of their future selves. The case of a 1-korsakoff naïf and one indivisible good allows for an interesting contingency. She can be oblivious to potential future revisions and still be time-consistent. This case relies on a weak preference for anticipations compared to the discount rate. On the other hand, if anticipations are not outweighed by the discount rate, naïfs may harvest higher long-run utility from repeated delays of consumption. The following theorem states this formally.

Theorem 3.3 *Let C be one indivisible good that is to be consumed in one of the T time periods. Consider a 1-korsakoff agent with utility function $DU_1 = \sum_i^T \delta^{i-1} u(c_i, a_i(c_{i+1,plan})$. Let $a = u(0, a(C))$ and $c = u(C, 0)$. If she is a naïf, her long-run utility will be:*

i. $U_{\textit{long-run}} = c$ *if* $a < (1-\delta)c$.

ii. $U_{\textit{long-run}} = c + (T-1)a$ *if* $a > (1-\delta)c$.

Sophisticates, though, given their sophistication, are prevented from having repeated anticipations of next period consumption:

Theorem 3.4 *Let C be one indivisible good, that is to be consumed in one of the T time periods. Consider a 1-korsakoff agent with utility function $DU_1 = \sum_i^T \delta^{i-1} u(c_i, a_i(c_{i+1,plan}))$. Let $a = u(0, a(C))$ and $c = u(C, 0)$. If she is a sophisticate, her long-run utility will be*

i. $U_{\textit{long-run}} = c$ *if T is odd.*

ii. $U_{\textit{long-run}} = a + c$ *if* $a > (1-\delta)c$ *and T is even.*

According to condition $a > (1-\delta)c$, the utility of anticipation outweighs the discounted utility of next period consumption. If it doesn't, consuming in the first period gives the highest utility. The even and odd condition is driven by the backward induction which follows from a perception-perfect strategy for sophisticates.[12]

Theorems 3.3 and 3.4 are, in sum, a bit discouraging. The 1-korsakoff naïfs can revise consumption plans and enjoy benign self-deception. They never realize their next period self will make the same calculations as they are doing today. Sophisticates are prevented from utility increasing revisions. At a more technical level, a sophisticate's perception-perfect strategy relies on backward induction, and this prevents unforeseen revisions. In the next section we will briefly discuss an extension to the model that facilitates repeated revision and

[12] This dependence of the parity of time periods is also present in O'Donoghue and Rabin (1999), in the case of time-consistent agents, TCs. However, in their Example 1 p. 109, they consider only the case of an even number of time periods ($T = 4$).

benign self-deception also for sophisticates.

4 The life span uncertainty model

In the previous section sophisticates obtained lower long-run utility than naïfs. In this section we consider an extension that facilitates repeated revisions by sophisticates as well as naïfs. The following scenario illustrates the extension. Upon retirement we decide to spend a year in Rio. Knowing we will not live forever, we need to decide when to go, now or later. As each year passes and we observe our health (h_i), we know that the probability of enjoying good health throughout the next year (p_i) is increasing in h_i. Agents do not know p_i, but form subjective beliefs $q(h_i)$ of the probability of a healthy year ahead. For notational convenience we define $p_0 = 1$.

We formalize this with the following definition.

Definition 4.1 *(Life Span Uncertainty Game (LSU game)) An LSU game is a one player game. At every node, the player has to choose between two actions, Y or N (consume or not consume an indivisible good). At any given node i, there is a positive probability* $(1 - p_i)$ *that this node is the last. The probability of continuation* $p_i = p(h_i)$ *is an increasing function in* h_i. *The player observes* h_i, *and assigns a subjective probability* $q(h_i)$ *for a next time period. The player knows that the true probability* p_i *is a function of* h_i *and that* h_i *is decreasing over time.*

Definition 4.2 *(naïf in LSU game) A perception-perfect decision rule for naïfs in the LSU game is a rule that assign* $s_i^n = Y$ *if and only if the expected utility* $EDU_k(i,i) = max(EDU_k(i,j))$ *for* $j > i$, *where* $EDU_k(i,j) = q(h_j)DU_k(i,j)$ *and* $q(h_j)$ *is the subjective probability for reaching j given* h_j.

In discussing the LSU game we assume that both 1-korsakoff naïfs and sophisticates are risk neutral and maximize expected utility given their observation of the health parameter h_i and subjective probabilities $q(h_i)$ for the next time periods. [13]

A perception-perfect strategy for sophisticates in this case, requires a refinement of Definition 3.2. A refinement may be achieved in the following way:

Let $a = u(0, a(C))$ and $c = u(C, 0)$. If $a + q(h_i)c > c$, then it is better to go the next year. The sophisticate realizes how this reasoning is conditional on going the next year. Moreover, if $a + q(h_{i+1})c > c$ the next-year sophisticate will not go. In this model there is a critical health level h_c such that $a + q(h_c)c = c$. When h_c is reached, there is no point in postponing consumption. In other words, if the sophisticate believes that $h_{i+1} \leq h_c$, she may enjoy anticipation at year i. As h_i is decreasing, an h_i close to h_c may be read as a high probability of $h_{i+1} \leq h_c$. We formalize this into the following termination criterion:

Definition 4.3 *(ϵ-criterion)*

If the player observes that $h_i - h_c < \epsilon$, then $q(h_{i+1} < h_c) = 1$, else $q(h_{i+1} < h_c) = 0$.

We can use this termination criterion to formulate perception-perfect strategies for 1-korsakoff sophisticates in the LSU game.

Definition 4.4 *(ϵ-sophisticate in a LSU game)*

A perception-perfect decision rule for an ϵ-sophisticate in an LSU game is a rule that assigns

1. $(s_i^s, s_{i+1}^s) = (N, Y)$ if $q(h_i)(a + c) > c$ and $h_i - h_c < \epsilon$ ($q(h_{i+1} < h_c) = 1$ by the ϵ-criterion).

[13] We do not assume that agents assign lower q's for more distant time periods.

2. $(s_i^s, s_{i+1}^s) = (Y, Y)$ *if* $q(h_i)(a+c) \leq c$ *or* $h_i - h_c \geq \epsilon$ $(q(h_{i+1} < h_c) = 0$ *by the* ϵ*-criterion).*

for every t*.*

Theorem 4.1 *1-korsakoff naïfs and 1-korsakoff* ϵ*-sophisticates have the same expected long-run utility in the LSU game provided* $h_i - h_c < \epsilon$ *for all* i*. The expected long-run utility is in this case:* $E(U) = a \sum_{i=1}^{\infty} (\pi_{j=0}^{i-1} p_j) q_i$*.*

The theorem tells us that the LSU game allows sophisticates as well as for naïfs to perform utility increasing revisions. That being the case, they would both also achieve the same long-run utility. The conditions for this to happen are restrictive. The health parameter needs to remain low but high enough, all the same, to exceed h_c for all time periods. Sophisticates as well as naïfs end up consuming nothing. Their long-run utility is pure anticipation.

The model can be extended by giving the agent a choice prior to period 1 between knowing and not knowing the number of time periods. As Epstein(2008) points out, the agent may well display information aversion. This as adds to the conflict of interest between present you and future you, since knowing the number of time periods may give the present you higher expected utility , but at the cost of future you. A thorough discussion of this extended model is beyond the scope of this paper.

5 Conclusion

We have set out in this paper a framework for beneficial self-deception. We show that utility from anticipation allows the agent to revise plans with immediate and long-runl utility benefits. This adds to the downside of commitment and self-control, and as such stands in contrast with much of the recent literature on time inconsistencies and self-deception.

Basically, the paper explores an extension to Samuelson's discounted utility model that allows agents to remember their plans and value future anticipation. These agents, called 1-korsakoffs, are very different from those inferred by Samuelson's standard discounted utility model (0-korsakoffs). 1-korsakoffs may achieve higher immediate and long-run utility from repeated revisions of consumption plans.

If we assume that agents vary with respect to sophistication, our results are structurally similar to those of O'Donoghue and Rabin (1999). A naïf fails to realize that her future self may decide to deviate from her original plan. Repeated revisions provide an opportunity for higher long-run utility. A sophisticate, on the other hand, doesn't have as many opportunities to engage in repeated revisions, because she considers only the consumption plans she is likely to follow. At a technical level, backward induction in consumption scenarios with a fixed number of time periods prevents a sophisticate from anticipating higher consumption levels of future selves than actually occur.

In sum, the results presented here illustrate the difficulties of uniting sophistication and self-deception even in the pursuit of higher immediate and long-run utility. This somewhat discouraging insight may be sweetened by a conjecture. Benign self-deception may be viable for sophisticates in consumption scenarios with a higher degree of complexity and uncertainty than the stylized consumption scenarios considered here. The final model extension where uncertainty regarding the number of time periods facilitates repeated revisions also for sophisticated agents can be read as modest evidence in favor of this conjecture.

Appendix

Proof of Theorem 2.1

It is enough to prove the result for $T = 3$. Consider two consumption plans:

$c_{plan} = (0,1,0)$ and $c^{\epsilon}_{plan} = (0,1-\epsilon,0)$. Since u, and a_i are all strictly increasing in their arguments is $DU_1(0,1,0) > DU_1(0,1-\epsilon,0)$, where we by slight abuse of notation let $DU_1(0,1,0)$ denote the total discounted utility at time 1 associated with the consumption plan $(0,1,0)$, and likewise for the consumption plan $(0,1-\epsilon,0)$. Furthermore, c_{plan} is also preferred at time 2: $DU_1(1,0) > DU_1(1-\epsilon,0)$. Again, since all functions are continuous, increasing and unbounded in their arguments, there exists an ϵ_2 such that $DU_1(1,0) = DU_1(1-\epsilon,\epsilon_2)$. (That is, there exists a planned consumption level at time 3 that fully compensates reduced consumption at time 2, evaluated at time 2.) By the same line of reasoning, an ϵ_3 exists such that $DU_1(0,1,0) = DU_1(0,1-\epsilon,\epsilon_3)$. (That is, there exists a planned consumption level at time 3 that fully compensates reduced consumption at time 1, evaluated at time 1.) Assume that $\epsilon_3 > \epsilon_2$ and consider an $\epsilon_4 \in <\epsilon_2,\epsilon_3>$. Since u and a_i are all strictly increasing in their arguments, we get: $DU_1(1,0) < DU_1(1-\epsilon,\epsilon_4)$ and $DU_1(0,1,0) > DU_1(0,1-\epsilon,\epsilon_4)$. In other words, the consumption plan $(0,1,0)$ is strictly preferred to $(0,1-\epsilon,\epsilon_4)$ at time 1, whereas the latter is strictly preferred to the former at time 2. It remains to prove whether $\epsilon_3 > \epsilon_2$. This is equivalent to proving that $DU_1(0,1,0) > DU_1(0,1-\epsilon,\epsilon_2)$, but this follows from a direct comparison of the following two equalities:

$I.\ DU_1(0,1,0) = u(0,1) + \delta DU_1(1,0)$

$II.\ DU_1(0,1-\epsilon,\epsilon_2) = u(0,1-\epsilon) + \delta DU_1(1-\epsilon,\epsilon_2)$

Since $DU_1(1,0)$ is equal by construction to $DU_1(1-\epsilon,\epsilon_2)$ by construction, by taking the difference between these two equations we get

$$DU_1(0,1,0) - DU_1(0,1-\epsilon,\epsilon_2) = u(0,1) - u(0,1-\epsilon) > 0$$

The right hand side is greater than zero, thus $DU_1(0,1,0) > DU_1(0,1-\epsilon,\epsilon_2)$, which is the desired inequality.

$\square$

Proof of Theorem 3.1

A perception-perfect strategy for naïfs at time i is a strategy, $s^n = (s_1^n, s_2^n, \ldots, s_T^n)$ that satisfies $s_i^n = Y$ if $DU_1(i,i) = \max(DU_1(i,j))$ for $j \geq i$ That is, she will delay consumption if and only if there exists a j, $j > i$ that gives higher utility. Since she is a 1-korsakoff, and the instantaneous utility function is equal for all periods, at any given time i, the present utility of time periods $i+2, \ldots T$ is strictly less than $DU_1(i,i)$ (Note that $\delta < 1$ implies $DU_1(i,j) < max\{DU_1(i,i), DU_1(i,i+1)\}$ for $j \geq i+2$.) In other words, consumption at time i, $s_i^n = Y$, if and only if $DU_1(i,i) = u(C,0) \geq u(0,a(C)) + \delta u(C,0)$.

Proof of Theorem 3.2

A perception-perfect strategy for sophisticates is one in which $s^s = (s_1^s, s_2^s, \ldots, s_T^s)$ that satisfies for all $i < T : s_i^s = Y$ if and only if $DU_1(i,i) \geq DU_1(i,j)$ for $j > i$ such that $j = \min_{k>i}(k|s_k^s = Y)$. That is, she will only choose action Y if the present utility of Y gives a higher utility than the present utility of the next (planned) Y. Since she is a 1-korsakoff, and the instantaneous utility function is equal across all periods, at any given time i, the present utility of time periods $i+2, \ldots T$ is strictly less than $DU_1(i,i)$ (Note that $\delta < 1$ implies $DU_1(i,j) < max\{DU_1(i,i), DU_1(i,i+1)\}$ for $j \geq i+2$.) In other words, consumption at time i, $s_i^n = Y$, if and only if one of the two following conditions holds:

1. $DU_1(i,i) = u(C,0) \geq u(0,a(C)) + \delta u(C,0)$ and $s_{i+1}^n = Y$

2. $s_{i+1}^n = N$.

As the action at time i is uniquely determined by the time invariant condition $u(C,0) \geq u(0,a(C)) + \delta u(C,0)$ and the action at time $i+1$, the one unique perception-perfect strategy will be determined by backward induction. In other words, she will be time-consistent.

□

Proof of Theorem 3.3

In the first time period she compares c versus $a + \delta c$. If $a + \delta c \leq c$, $s_1^n = Y$, (since all later consumption times will be discounted ($c > \delta^i(a + \delta c)$ for all $i > 0$). $s^n = (s_1^n, s_2^n, \ldots, s_T^n)$=$(Y, Y \ldots, Y)$. This gives a long-run utility of c. If $a + \delta c > c$, then $s_1^n = N$, since consuming the next period is better. Then $s^n = (s_1^n, s_2^n, \ldots, s_T^n) = (N, Y, Y \ldots, Y)$ defines a perception-perfect strategy. The same reasoning also applies to a naïf who has reached time period i. That is, $s_i^n = (s_i^n, s_2^n, \ldots, s_T^n) = (N, Y, Y, \ldots, Y)$ defines a perception-perfect strategy at time i. The long-run utility is $c + (T - 1)a$, since in every time period consumption in the next period gives the highest utility except when the final period T is reached.

□

Proof of Theorem 3.4

Using backward induction we get $s_{T-1}^s = Y$ if and only if $c \geq a+\delta c$. That is $a \leq (1 - \delta)c$. In this case (by induction again) $s^s = (s_1^s, s_2^s, \ldots, s_T^s) = (Y, Y, \ldots, Y)$ is a perception-perfect strategy. This gives long-run utility c. If $a < (1 - \delta)c$ then $s_{T-1}^s = N$. In this case s_{T-2}^s must be equal to Y, since $c > \delta a + \delta^2 c$ (Note: $a < (1 - \delta)c$ implies $\delta a + \delta^2 < \delta(1 - \delta)c + \delta^2 c = \delta c < c$ (only next period anticipation for 1-korsakoffs, and $s_{T-1}^s = N$ implies no planned consumption in time period $(T - 1)$). By induction we get that $s^s = (s_1^s, s_2^s, \ldots, s_T^s) = (Y, N, Y, \ldots, N, Y)$ if T is odd, and $s^s = (s_1^s, s_2^s, \ldots, s_T^s) = (N, Y, N, \ldots, N, Y)$ if N is even. By construction these strategies are perception-perfect strategies for the odd and even cases respectively. The long-run utility associated with these strategies is c if T odd, and $a + c$ if T even.

□

Proof of Theorem 4.1

If $h_i - h_c < \epsilon$ for all i, the perception-perfect decision rule for ϵ-sophisticates and naïfs agrees at every achieved decision node (time interval). Furthermore, their beliefs regarding future behavior are the same in that both believe in next period consumption. The long-run utility is given by $q_1 a + p_1 q_2 a + p_2 p_1 q_3 a + \cdots = a \sum_{i=1}^{\infty} (\pi_{j=0}^{i-1} p_j) q_i$.

□

Example A.1

Going to Rio, the case of a 3-korsakoff

Sheila has won a trip to Rio. She has three options. Go this year, next year or the year after. We assume for expository purposes that her preferences are given by an additive DU_3-model with memory and anticipation, $U_i = c_i + m_i(c_{<i}) + a_i(c_{>i,\ \text{plan}})$, with a slight notational abuse measuring all three in utility directly.

Sheila has three possible consumption plans $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.[14] Their corresponding utility evaluated at the start of year one is:

$U(1, 0, 0) = 1 + \delta m^1 + \delta^2 m^2$

$U(0, 1, 0) = a^1 + \delta + \delta^2 m^1$

$U(0, 0, 1) = a^2 + \delta a^1 + \delta^2$,

[14] The consumption utility of the Rio trip is normalized to 1

where $m^1 = m_2(c_{<2}) = m_2((1,0,0))$, the superscript tells us that it is referring to memories of consumption in the previous period and $m^2 = m_3(c_{<3}) = m_3((1,0,0))$, the superscript tells us that it is memories of consumption two periods back. We adopt the same notation for anticipation; that is, a^1 is anticipation of next period anticipation, and a^2 anticipation of consumption two periods later.

If she does not travel the first year, she will compare, $a^1 + \delta$ to $1 + \delta m^1$ the second year. Assume that $a^1+\delta > 1+\delta m^1$, then she will rate traveling the second year as better than the first year, provided that $m^1 > m^2$, i.e. she values first order memories higher than second order. She will be time-inconsistent if $a^1 + \delta + \delta^2 m^1 > a^2 + \delta a^1 + \delta^2$. In other words, she is time-inconsistent if $a^1 > \max(1-\delta+\delta m^1, (a^2-\delta+(1-\delta^2)m^1)/(1-\delta))$ and $m^1 > m^2$. In this case, time inconsistency relies on her rating next period consumption higher than combination of next period memories and later periods anticipation. The question of time inconsistency under full recollection is shown to be nontrivial by this example, and dependent on the relative strength of utility of anticipation and memories as well as the discount factor.

References

Asheim, Geir 1997. Individual and Collective Time-Consistency, Review of Economic Studies, 64(3), 427–443.

Douglas Bernheim, B. and Raphael Thomadsen, R. 2005. Memory and Anticipation, Economic Journal, 115(305), 271–304.

Caplin, Andrew and Leahy, John 2001. Psychological Expected Utility Theory and Anticipatory Feelings, Quarterly Journal of Economics, 116(1), 55–79.

Epstein, Larry G 2008. Living with risk, Review of Economic Studies 75, 2008, pp. 1121–1141.

Fudenberg, Drew, Levine, David K. 2012. Timing and Self-Control, Econometrica, 80(1), pp. 1–42.

Goldman, Steven M 1979. Intertemporally inconsistent preferences and the rate of consumption, Econometrica, 47, pp. 621–626.

Gul, Faruk and Pesendorfer, Wolfgang 2001. Temptation and Self-Control, Econometrica, 69(6), pp. 1403–1435.

Halevy, Yoram. 2008. Strotz Meets Allais: Diminishing Impatience and the Certainty Effect, American Economic Review, 98:3, pp. 1145–1162.

Kőszegi, B., 2010 Utility from anticipation and personal equilibrium, Economic Theory 44, pp. 415–444.

Loewenstein, George. 1987. Anticipation and the Valuation of Delayed Consumption, Economic Journal, 97(387), pp. 666–684.

Kopylov, Igor., Noor, Jawwad 2010. Self-Deception and Choice, mimeo, Boston University.

Noor, Jawwad 2007. Commitment and self-control, Journal of Economic Theory, 135, pp. 1–34.

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

Peleg, Bezalel and Yaari, Menahem E. 1973. On the existence of a consistent course of action when tastes are changing, The Review of Economic Studies, 40(3), pp. 391–401.

Samuelson, Paul. 1937. A Note on Measurement of Utility, Review of Economic Studies 4, pp. 155–161.

Sarver, Todd. 2008. A Note on Measurement of Utility, Econometrica 76(2), pp. 263–305.

Shu, Suzanne B., Gneezy, Ayelet, 2010. Procrastination of Enjoyable Experiences, Journal of Marketing Research XLVII, pp. 933–944.

Trivers, Robert 2011. Deceit and Self-Deception, Allan Lane, Penguin books Ltd, London, England.