



# General time consistent discounting



Tor Lattimore, Marcus Hutter

Research School of Computer Science, Australian National University, Australia

## ARTICLE INFO

### Keywords:

Rational agents  
Sequential decision theory  
General discounting  
Time-consistency  
Game theory

## ABSTRACT

Modeling inter-temporal choice is a key problem in both computer science and economic theory. The discounted utility model of Samuelson is currently the most popular model for measuring the global utility of a time-series of local utilities. The model is limited by not allowing the discount function to change with the age of the agent. This is despite the fact that many agents, in particular humans, are best modelled with age-dependent discount functions. It is well known that discounting can lead to time-inconsistent behaviour where agents change their preferences over time. In this paper we generalise the discounted utility model to allow age-dependent discount functions. We then extend previous work in time-inconsistency to our new setting, including a complete characterisation of time-(in)consistent discount functions, the existence of sub-game perfect equilibrium policies where the discount function is time-inconsistent and a continuity result showing that “nearly” time-consistent discount rates lead to “nearly” time-consistent behaviour.

Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.

## 1. Introduction

A rational agent, by definition, should choose its actions to maximise its expected utility [17,18]. Discounting is used to construct a simple model of global utility as a weighted sum of local utilities (well-being experienced at each time-step). The weighting usually assigns greater value to earlier, rather than later, consumption.

The discounted utility (DU) model, first introduced by Samuelson [21], provides a framework for making decisions about inter-temporal consumption. Samuelson assumed that the global utility of a sequence of local utilities could be expressed as follows

$$V_k = \sum_{t=k}^{\infty} d_{t-k} r_t \quad (1)$$

where  $r_k, r_{k+1}, \dots$  is an infinite sequence of expected local rewards (utilities),  $V_k$  is the global utility at time  $k$  and  $d_{t-k}$  is the weight assigned to consumption at time-step  $t$  when in time-step  $k$ .

This model has a number of consequences:

1. *Utility independence*: It assumes that global utility can be represented as a discounted sum of local utilities, which removes the possibility of preferring one utility structure over another. For example, there is no way to distinguish between a relatively flat well-being profile and a roller-coaster of ups and downs [4].
2. *Consumption independence*: In the simplistic models commonly used in economic theory, the instantaneous utility of consumption at time  $k$  is independent of previous consumption choices [13,22]. This means that if pizza is preferred to Chinese on one day then it will be preferred every day. It is not possible to account for getting sick of pizza.

E-mail addresses: [tor.lattimore@anu.edu.au](mailto:tor.lattimore@anu.edu.au) (T. Lattimore), [marcus.hutter@anu.edu.au](mailto:marcus.hutter@anu.edu.au) (M. Hutter).

3. *Age independence*: The DU model denies the possibility that the discount function may change over time.
4. *Time-inconsistency*: An agent choosing a plan to maximise its discounted utility for the future may continuously change its plan over time despite receiving no new information [24].

The first and third are limitations of the DU model while the last is a rational consequence of an agent acting to maximise its discounted utility (in some environments and with some discount functions). Despite these limitations, the simple DU model is widely used in both computer science and economics.

In this paper we address all points above while updating the work of [24] on time-inconsistency to our more general setting. The first two limitations are largely eliminated by using a more general model than typically considered by economists (see the first example in Section 2). The third limitation is removed by allowing the discount function to change over the lifetime of the agent. For each time-step  $k$  we assume the agent uses a (possibly) different discount function  $d^k$ . Utility can then be written

$$V_k = \sum_{t=k}^{\infty} d_t^k r_t. \tag{2}$$

This allows an agent to become more (or less) farsighted over time, which is important for two reasons. First, because we frequently wish to model humans that operate in this way, i.e. children plan only a few months, or at most years ahead, whilst adults think also of retirement. Some studies have shown this experimentally by empirically estimating discount rates [9]. The second reason comes from computer science where we wish to construct agents that behave in certain ways. Allowing an increasing effective horizon may allow an agent to explore more effectively [11]. For example, it is well known that the Bayesian policy for learning unknown stochastic bandits with geometric discounting suffers from linear regret, while other algorithms enjoy logarithmic regret [7,16]. This occurs because a rational agent discounting geometrically has no incentive to explore more than a certain amount as the reward it receives from the extra knowledge occurs too far in the future. Under certain conditions, however, a Bayesian agent with a more farsighted discount function suffers sub-linear regret in the bandit setting, as well as more general environment classes [10].

It has been remarked that time-inconsistent behaviour can be a rational consequence of discounting. This has been used to explain inter-temporal preference reversals observed in humans. For example, many people express a preference for \$50 in three years and three weeks over \$20 in three years, but favour \$20 now rather than \$50 in three weeks [6]. This effect is a natural consequence of some discount functions (hyperbolic) but not others (exponential).

Strotz showed that if the same discount function is used in each time-step as in Eq. (1), then *only* exponential discounting is guaranteed to lead to time-consistent rational agents [24]. Strotz worked with continuous time and assumed the utility  $V$  at time  $k$  could be written as

$$V_k = \int_k^{\infty} d_{t-k} r(t) dt \tag{3}$$

where  $r$  is now a continuous function. Formally he showed that if  $d_{t-k}$  is not proportional to  $\gamma^{t-k}$  for some  $\gamma \in (0, 1)$ , then there exists an environment where the policy that maximises utility changes over time.

We extend this work to a complete classification of time-consistent discount functions in the more general case where the discount function is permitted to change with age. However, rather than using deterministic choices of continuous consumption profiles as in [24], we use (stochastic) Markov decision processes to model our environments [3,23]. This has a number of implications:

1. Discrete time, rather than continuous.
2. Arbitrary utility and infinitely many consumption profiles choices.

The limitation of discrete time should not be seen as too problematic for two reasons. First, because it is possible to use discrete time to approximate continuous time. Second, because our main theorem regarding the characterisation of time-consistent discount rates in discrete environments is transferable to the continuous case with minimal effort.

Given that an agent may operate using a time-inconsistent discount function, it is reasonable to ask how they will behave, and also how they should behave. If the agent is unaware that its discount function is time-inconsistent then it will (if rational) take action to maximise its expected discounted utility at the present time (ignoring that it may not follow its own plan due to changing preferences). For time-inconsistent discount functions this can lead to extremely bad behaviour in some environments (see Section 3 for an example).

On the other hand, if the agent knows its discount function is time-inconsistent, then blindly acting as if it were not is irrational. In this case it may be optimal to take a course of action that restricts its own choices in the future to ensure its future self does not act poorly according to its current preferences. To illustrate the idea, a recovering alcoholic who knows they become myopic in the evenings may choose to pour their alcohol away in the morning and so remove temptation in

Download English Version:

<https://daneshyari.com/en/article/438406>

Download Persian Version:

<https://daneshyari.com/article/438406>

[Daneshyari.com](https://daneshyari.com)