An approximate dual-self model and paradoxes of choice under risk

Drew Fudenberg [a], David K. Levine [b], Zacharias Maniadis [c,*]

[a] Harvard University, Department of Economics, Littauer Center 310, 1805 Cambridge Street, Cambridge, MA 02138, United States
[b] Washington University of St. Louis, Department of Economics, Campus Box 1208, St. Louis, MO 63130-4899, United States
[c] University of Southampton, Economics Division, School of Social Sciences, Southampton SO17 1BJ, United Kingdom

ARTICLE INFO

ABSTRACT

We derive a simplified version of the model of Fudenberg and Levine (2006, 2011) and show how this approximate model is useful in explaining choice under risk. We show that in the simple case of three outcomes, the model can generate indifference curves that "fan out" in the Marschak–Machina triangle, and thus can explain the well-known Allais and common ratio paradoxes that models such as prospect theory and regret theory are designed to capture. At the same time, our model is consistent with modern macroeconomic theory and evidence and generates predictions across a much wider set of domains than these models.

1. Introduction

Fudenberg and Levine (2006, 2011, 2012) develop a model of costly self-control that can explain many ways that observed individual choice departs from the predictions of the "standard model" of maximizing expected discounted utility. Their self-control model is based on the idea that a more rational "long run self" controls the impulses of a "short run self" that is very tempted by immediate rewards.[1] Fudenberg and Levine (2006) points out that the self-control model can explain "time-domain" phenomena, such as a preference for commitment and time-inconsistent choice, and that when the model is enriched with the assumption of mental accounts or "pocket cash constraints" it can also explain the very high levels of small-stakes risk aversion seen in the lab, a quantitative puzzle that has become known as the Rabin paradox, after Rabin (2000). Fudenberg and Levine (2011) show that the same model can also explain the interaction of risk and delay seen in such experiments as Baucells and Heukamp (2010) and Keren and Roelofsma (1995). Moreover they move beyond the qualitative

---

* Corresponding author.
 E-mail addresses: dfudenberg@harvard.edu (D. Fudenberg), david@dklevine.com (D.K. Levine), Z.Maniadis@soton.ac.uk (Z. Maniadis).
 [1] While the model is not intended as a very precise model of the internal processes underlying self-control, at a very rough level the model is consistent with fMRI evidence, since the "long run self" is identified with activation in the prefrontal cortex while the short run self corresponds to more primitive and faster-acting parts of the brain. See e.g. McClure, Laibson, Loewenstein, and Cohen (2004, 2007) for fMRI evidence that support this general idea.

matching of theories and facts that is typical in this literature to a quantitative calibration of the model to both Rabin-paradox data and the Allais paradox.[2]

Unfortunately the model of Fudenberg and Levine (2011) is fairly complex, which may obscure some of the key insights and make it difficult for others to apply the model. Our purpose here it to develop an approximation to the FL model that is easier to work with yet still accurate enough to be useful in applied work. After developing this approximation, we characterize its theoretical properties and show how it helps explain observed behavior in the Allais and common ratio paradoxes and examine the implications of the theory for intransitivity.

To study decision makers who act as if tempted by money winnings but also manage to save, as well as to explain the level of risk aversion observed in lab experiments, Fudenberg and Levine (2006) use the idea of mental accounting: A decision maker reduces the cost of restraining impulsive decisions by using mental accounting to commit to daily expenditures – the idea is that the mental account is set when the decision maker is in a "cool state" and not subject to temptation. By assumption, the commitment is to net expenditures, and not to consumption *per se,* so small losses must be born out of daily expenditures and small gains create a self-control problem, which leads the marginal propensity to consume out of small gains to be equal to 1.[3] Because small losses and small gains are applied entirely to daily expenditures and not spread over the lifetime, the decision maker is much more risk averse over small unexpected lotteries than under the classical model, where any change in wealth results in a much smaller permanent change in consumption over the individual's lifetime.

A second consequence of this theory is that if there is an increasing marginal cost of self-control then the decision maker's utility is not linear in probabilities. Moreover, while any form of nonlinearity makes the model depart from expected utility, the increasing-marginal-cost specification predicts the particular violations of the independence axiom seen in, for example, the Allais, common-ratio and other related paradoxes, as detailed in Fudenberg and Levine (2011).

The approximate version of the theory that we develop here uses several simplifying assumptions. A key simplification is the assumption that the long-run value function is risk neutral, that is, the marginal utility of savings is a constant. This is a good approximation to decisions that have little impact on lifetime wealth; it simplifies the model by replacing an unknown non-linear value function with a known linear value function. We also assume that the interest received over a single period (the "temptation horizon" of the short run self) is small enough to be ignored; this fits with the usual calibration of this period length to be one to three days. We use the simplified model to explain how the theory ranks general small-stakes money lotteries, and illustrate this in the context of lotteries with only three possible outcomes in the gains domain using the classic Marschak–Machina depiction of indifference curves in the corresponding probability simplex. We also illustrate how the model leads to intransitivity.

The structure of the paper is as follows. In Section 2 we derive the approximate version of the dual-self model. In Section 3 we study its properties using a series of propositions. In Section 4 we examine the special case of a single gamble with a unique positive prize. Section 5 addresses the very interesting case of choices in menus of two lotteries, with three possible outcomes. We show how the model predicts well-known paradoxes that violate expected utility and illustrate this in the Marschak–Machina triangle. Section 6 provides a general discussion and concludes.

2. Deriving an approximate dual-self model

We will remind the reader of the main ingredients of the Fudenberg and Levine (2006, 2011) model, and then show how the approximation of risk neutrality for wealth leads to a much more tractable model. There is an individual who makes a consumption-savings decision, with a short-run utility function $u^x(x + c; x)$ each period, where $x + c$ represents total consumption, x is the planned level of consumption under the mental account ("pocket cash"), and c denotes "incremental consumption": the additional (possibly negative if money is lost) consumption made possible by unexpected windfalls.

When studying a fixed individual and holding fixed that individual's initial wealth and preferences, we can suppress the dependence on x and take short-run utility to be $u(c) = u^x(x + c; x)$, where $u' > 0$ and $u'' < 0$. In the remainder of the paper, the term "consumption" will refer to this incremental consumption.

We are primarily interested in how the agent chooses lotteries Z from a fixed menu $\mathfrak{I}$, where each of the lotteries resolves in the current period. For this, an important intermediate step is to analyze the agent's optimal consumption *ex post* after a particular lottery has been chosen.[4] Let $u^* = \max_{Z \in \mathfrak{I}} Eu(Z)$ be the greatest available short-run utility. This "temptation" represents what the short-run self would like – to spend all the gains immediately. If the lottery Z has n outcomes, the choice of optimal consumption entails choosing an optimal random consumption plan $\bar{c}$ with outcomes $(c_1, \ldots, c_n)$, specifying one consumption level for each possible lottery realization. Overall first period utility is given by $Eu(\bar{c}) - g(u^* - Eu(\bar{c}))$ where g is

---