

Article

A Markov Decision Process with Awareness and Present Bias in Decision-Making

Federico Bizzarri ¹, Chiara Mocenni ¹ and Silvia Tiezzi ^{2,*}

¹ Department of Information Engineering and Mathematics, University of Siena, Via Roma, 56, 53100 Siena, Italy; federico.bizzarri@student.unisi.it (F.B.); chiara.mocenni@unisi.it (C.M.)

² Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8, 53100 Siena, Italy

* Correspondence: silvia.tiezzi@unisi.it

Abstract: We propose a Markov Decision Process Model that blends ideas from Psychological research and Economics to study decision-making in individuals with self-control problems. We have borrowed a dual-process of decision-making with self-awareness from Psychological research, and we introduce present bias in inter-temporal preferences, a phenomenon widely explored in Economics. We allow for both an exogenous and endogenous, state-dependent, present bias in inter-temporal decision-making and explore, by means of numerical simulations, the consequences on well-being emerging from the solution of the model. We show that, over time, self-awareness may mitigate present bias and suboptimal choice behaviour.

Keywords: awareness; present bias; decision-making; Markov Decision Process

MSC: C61; C63; D91



Citation: Bizzarri, F.; Mocenni, C.; Tiezzi, S. A Markov Decision Process with Awareness and Present Bias in Decision-Making. *Mathematics* **2023**, *11*, 2588. <https://doi.org/10.3390/math11112588>

Academic Editor: Óscar Valero Sierra

Received: 30 April 2023

Revised: 26 May 2023

Accepted: 30 May 2023

Published: 5 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since Samuelson's [1] discounted utility model, Economics has modelled impatience in decision-making by assuming that agents discount future streams of utility exponentially over time. Exponential discounting means that people have the same preferences now over future behaviour as they will have when the future arrives. For example, if today we prefer to start spending less and saving more tomorrow, tomorrow we will want to start spending less and saving more immediately; if today we prefer to quit smoking tomorrow, then tomorrow we will want to quit smoking immediately; if today we prefer to go on a diet tomorrow, tomorrow we will want to go on a diet immediately; and if today we prefer to start exercising tomorrow, tomorrow we will want to start exercising immediately.

Behavioural Economics has built on the work of [2] to explore the consequences of relaxing the standard assumption of exponential discounting. Both [3] and [4] suggest that some basic features of inter-temporal decision-making (namely that many individuals value, e.g., consumption in the present more than any delayed consumption) may be explained by a particular type of time inconsistency, a quasi-hyperbolic discounting which formalizes the idea that today's "self" is impatient while tomorrow's "self" is much more patient. This shape of time discounting was originally proposed by Phelps and Pollak in 1968 [5], later employed by David Laibson [6,7], and further developed in the work of O'Donoghue and Rabin [8–10]. In the formulation of quasi-hyperbolic discounting adopted by [7], the taste for immediate gratification or present bias is captured by an extra discount parameter $\beta \in (0, 1]$. Accordingly, the consumption path planned at each time period for future periods may never be realized, due to inter-temporal trade-off changes over time. If people have a bias towards immediate gratification, when thinking about two generic future dates they care roughly equally about well-being on those two dates, but when the first of those two future dates becomes "today", people care more about well-being on the first date than about well-being the second date. So, they will save less, smoke more, eat more,

and exercise less tomorrow than they plan today they will do tomorrow. In other words, there is a conflict between what the individuals would like for themselves today and what they would like for themselves tomorrow. The implications of such self-control problems depend on the agents' awareness of their future preferences [8,11]. Extreme assumptions about such awareness, e.g., full awareness and full unawareness, identify two types of individuals usually considered in the Behavioural Economics literature [2,8,12]: naïve and sophisticated. Sophisticated persons are fully aware of what their future selves' preferences will be. Naïve persons are instead fully unaware of what their future preferences will be. They believe their future self's preferences will be identical to their current selves, not realizing that their tastes will have changed as they become closer to implementing their decisions.

To analyse the equilibrium behaviour of individuals with different time preferences, for example, equilibrium consumption behaviour, researchers in Economics [8,9,11] have formally modelled a consumer as a sequence of temporal selves making choices in a dynamic game. Hence, a T -period consumption problem translates into a T -period game, with T players or selves indexed by their respective periods of consumption decision. Individual behaviour is described by *perception-perfect strategies*, i.e., solution concepts describing the individual's optimal action in all periods given her current preferences and her perception of future behaviour. The naïve have present-biased preferences but believe that they are time consistent (TC). Therefore, the decision process of the naïve is identical to that of TCs, and amounts to choosing an optimal future consumption path. Thus, both naïve and TC equilibrium behaviour can be solved as an optimization problem [9]. At all times they maximize their expected utility given their current information

$$\text{Max } U_t + \beta \sum_{i=1}^{\infty} \delta^i U_{t+i} = U_t + \beta \delta U_{t+1} + \beta \delta^2 U_{t+2} + \dots \tag{1}$$

where $\delta = \frac{1}{(1+\tau)}$ is the long-run discount factor, τ is the discount rate, and the extra discount parameter $\beta \in (0, 1]$ is intended to capture the essence of quasi-hyperbolic discounting; namely, that the discount factor between consecutive future periods, δ , is larger than the discount factor between the current period and the next one, $\beta\delta$. If $\beta \neq 1$ preferences in Equation (1) are dynamically inconsistent, that is to say preferences at date t are inconsistent with preferences at date $t + 1$. If, instead, $\beta = 1$ the discount factor between consecutive future periods, δ , is the same as the discount factor between the current period and the next one, $\beta\delta$, and individuals are time-consistent.

For sophisticates, *perception-perfect strategies* imply that they are playing a game against their future selves. Their behaviour partly reflects "strategic" reactions to predicted behaviour by future selves that they cannot directly control, and partly reflects attempts to induce better behaviour from their future selves.

In this strand of literature, the decision process of naïve and TC individuals amounts to just choosing an optimal future consumption path, but the equilibrium behaviour of sophisticated individuals is the solution to a dynamic game in which the decision maker is modelled as a sequence of temporal selves making choices.

The Economics literature summarized here assumes that present-biased individuals are endowed with a given, exogenous, degree of present bias and with an exogenous level of awareness about their self-control problems. References [10,13] make assumptions on individuals' awareness of their future self-control problems. People could be fully aware of their future self-control problems and correctly predict how they will behave in the future; by indicating the perceived level of self-control with $\hat{\beta}$, full sophisticates will have $\hat{\beta} = \beta < 1$. People could be fully unaware of their future self-control problems and wrongly predict how they will behave in the future. Fully naïve people will have $\beta < \hat{\beta} = 1$. Further, people could be partially naïve, aware that they will have self-control problems, but underestimate their magnitude: $\beta < \hat{\beta} < 1$. Finally, people with standard time-consistent preferences will have $\hat{\beta} = \beta = 1$.

In this strand of literature, the degree of individual present bias is exogenous and not affected by one's level of awareness. However, it is not unreasonable to assume that agents have the power to affect their discount factor through various means, such as investing in education [14], and/or to affect their present bias through self-awareness. In addition, this literature defines self-awareness as the exogenously given difference between perceived and actual present bias, leaving no scope for learning.

A recent paper in Psychological research proposed a Markov Decision Process Model of Decision-Making which embeds awareness as a dynamic process, allowing the decision maker (DM) to switch from habitual to optimal behaviour [15]. In this model, the level of awareness impacts individual well-being at each time period, and agents discount future streams of well-being exponentially over time.

The purpose of this paper is to offer a generalization of that model and allow for a taste of immediate gratification, i.e., present bias, in inter-temporal decision-making. The evolution of well-being obtained by solving the model using numerical simulations shows that, for a given magnitude of the present-bias parameter, β , a higher level of individual awareness leads to higher well-being. Consider, as an example, present-biased DMs whose retirement savings are too low. By increasing their awareness level, they might decide to commit to a savings contract forcing them to save a given amount per month. Under the setup of [15], the decision to subscribe a savings contract—a commitment device driven by higher awareness—improves decisions and well-being. We then make β endogenous by letting the present-bias parameter vary with the awareness level and show that even individuals suffering from severe present bias may learn over time to partly overcome their self-control problems.

This paper does not have the presumption nor the ambition to contribute to the Economics literature on present bias but aims to highlight the possible relationship between self-awareness and present bias using, for this purpose, a simple mathematical framework proposed in Psychological research. We offer the following contributions to the literature.

First, we use a mathematical framework—a Markov Decision Process (MDP)—for modelling inter-temporal decision-making in situations where outcomes are partly random and partly under the control of the decision maker, and where there is scope for learning, as the probability that the decision process moves into its new state is influenced by the previously chosen actions. Therefore, self-awareness is naturally embedded in this setup. The essence of a MDP is that a decision maker's action generates a response impacting the individual state, and consequently affects the immediate reward obtained by the agent, as well as the probabilities of future state transitions. The agent's objective is to select actions to maximize a long-term measure of total reward. We add individual traits such as self-awareness and present bias to this mathematical tool. Second, we embed awareness and present bias in a dual-process model of decision-making, relying on the intuition that decision-making is not made by a unique coherent entity, but by individuals who may operate in a "cold" or "hot" mode (Dual process theories of decision-making rely on the concept of multiple interacting brain systems in neuroscience, i.e., on the existence of several brain systems interacting with each other to make a variety of choices. The idea that choices can be explained by decision-making not made by a unique coherent entity has been developed both in Psychology [16,17] and in Behavioural Economics [18,19]; see [20] for a survey of dual process theories of decision-making in Economics). Our model can thus predict heterogeneity in awareness and present bias across individuals and within individuals over their life cycle. Third, we explore how individuals can affect their present bias through self-awareness in a general setting. Specifically, we explore how a psychological factor—awareness—can impact self-control problems implied by present bias, an additional discounting parameter. Finally, our setup is general, not centred on economic choices. The main implication is that we do not include the usual elements of choice in economic models, e.g., preferences, prices, and a budget constraint that enter the dynamic programming algorithm.

The model can find application in a variety of fields, such as strategic managerial decisions [21,22], social psychology [23] and health research [24], where the combination of intuitive–analytic reasoning propensities and present bias in inter-temporal decision-making may contribute to understanding and promoting new views of yet unexplained processes.

The paper proceeds as follows: the generalized Sequential Decision Process is outlined in Section 2, we present numerical simulations of the model in Section 3, behavioural patterns emerging from the simulations are discussed in Section 4, and Section 5 concludes.

2. The Model

In this Section, we develop a generalization of [15]’s model. The details and the background of the model are reported in the companion paper. Here we introduce the basic formalism necessary to incorporate the idea of present bias introduced in the previous Section.

Using MDPs to model the mechanisms and the dynamics of human decision-making seems natural for many reasons. The first reason is that MDPs describe processes changing over time, an essential element of inter-temporal choice. Secondly, the stochastic terms in the model allow us to embed internal and external sources of uncertainty in human decision processes. Finally, the actions obtained by maximizing the reward function correspond to the solution of an optimal control problem, so the model can mine information on the main factors to address when dealing with present bias. Moreover, the optimal policies embed the objective, expressed in terms of the expected reward at the final time, to be achieved by the DM.

Definition 1. *An MDP is a tuple $M = (S, A, P, r)$ where S is a finite set of states, A is a finite set of actions, $P : S \times S \times A \rightarrow [0, 1]$ is a transition probability function, and $r : S \times S \times A \rightarrow R$ is a reward function.*

Each individual makes a decision in the present by taking into account the outcomes realized over a time horizon of length T .

Definition 2. *The time horizon is the discrete and finite set $t = 0, 1, 2, \dots, T$.*

At each time period, the individual experiences a particular state, s_t , which belongs to the state space, S , representing the finite set of available states.

Definition 3. *The state $s_t \in S = [0, 1]$ of the individual is defined as her level of awareness at time t .*

The state of the individual provides information on the level of awareness they adopt when making decisions at each time period, and belongs to the state space S defined in the closed interval $[0,1]$ so that the higher the value of s at time t , the higher the level of awareness. The state and, particularly, its evolution over time, is the main aspect we focus on in this work. The DM’s level of awareness has an impact on their overall well-being as well as on their choices. By considering a Markov Decision Process, the current state incorporates the entire history of the DM so that awareness is a state embodying, to some extent, all aspects affecting individual decision-making: from personality to values and beliefs developed over a lifetime, to education and past experiences.

At each time period, the individual makes a choice a_t , which belongs to the action space, A , denoting the set of available actions.

Definition 4. *The action $a_t \in A = (0, 1)$ chosen by the individual at time t is defined as their decision, and embodies the level of analyticity adopted in solving a particular decision-making problem.*

In our framework, each choice is the solution to a decision problem. The individual faces a generic decision problem over a set of possible actions, representing different trade-offs between an intuitive ($a_t = 0$) and an analytical approach ($a_t = 1$), which are the two

extremes of a continuum of values. In this way, the focus is maintained on the process underlying the decisions without focusing on the specific decision in itself.

Psychological research suggests that decision-making is characterized by a dual-process mechanism [16,17,25,26] where the individual reasoning propensity can be hot or cold. The authors of [27] provide empirical evidence that the two systems are, in fact, complementary and necessary for optimal decision-making. We introduce a dual-process mechanism by defining the reasoning propensity in the following way.

Definition 5. *The reasoning propensity $p_r \in (0, 1)$ is an individual trait defining the DM’s attitude in processing the information used in the decision problem embedding the trade-off between two reasoning modalities: intuitive and analytical.*

The reasoning propensity takes values in a continuum between the two extreme attitudes [28] called intuitive ($p_r = 0$) and analytical ($p_r = 1$), assuming that both are always involved, to different degrees, in any decision. The fact that the interval is open, i.e., the extremes are not in the domain of p_r , remarks the idea that both mechanisms are always involved with a different trade-off, and one does not completely exclude the other.

The agent does not have perfect knowledge of the outcomes of the decision due to the interaction with the external environment which embeds uncontrollable factors. This uncertainty affects the evolution of the state over time and the reward resulting from the decision a_t . Regarding the first aspect, the DM’s state, s_t , evolves according to the following non-deterministic dynamic rule:

$$s_{t+1} = f(s_t, a_t, w_t). \tag{2}$$

The future level of awareness of the individual depends on the current one, s_t , and on the choices they made, a_t , but it is also subject to some uncertainty represented by a stochastic variable, w_t , related to the state transition. We assume, for simplicity, that the state can remain the same, increase or decrease by only one step at each time period, thus preventing “jumps” in its dynamic. This reflects the consideration that awareness-raising is a continuous process over time. By convention, considering the current state s_t , we write $s_t + z$ to indicate that the state increases by one step z , and $s_t - z$ to indicate that the state decreases by one step. With these assumptions, the stochastic variable w_t belongs to the set $W = \{z, 0, -z\}$, indicating the possibility for the state to increase, remain constant, or decrease, respectively. The probability with which w_t takes on each of these values depends on the decision a_t (it does not have, in this case, any explicit dependence on the state s_t).

The system dynamics can equivalently be written in terms of probabilities. We assume that for each state s_t , only three possibilities exist: $s_t + z$, s_t , and $s_t - z$, and we define a function P with three components specifying the probability of each of the three cases depending on the decision a_t . P is represented by a matrix of dimensions $[|A| \times 3]$, in which the three columns embed the different probabilities:

$$P = [(P^F(a_t) \ P^S(a_t) \ P^B(a_t))], \tag{3}$$

specifying the probability for the state to increase (*Forward probability*), remain constant (*Stationary probability*), or decrease (*Backward probability*), respectively, explicitly depending only on the action a_t made at time t . Alternatively we can say that w_t assumes a value equal to z with probability $P^F(a_t)$, 0 with probability $P^S(a_t)$, and $-z$ with probability $P^B(a_t)$.

The last element defining an MDP is a reward function.

Definition 6. *The reward function $r : S \times S \times A \rightarrow R$ is a stochastic function where $r(s, s', a)$ gives the reward the agent obtains by performing the action a from state s and transitioning to state s' .*

The appropriate formulation of the reward function, r_t , depends on its relationship with awareness. The authors of [29] recently proposed a theory of self-connection and

showed how self-connection is an important potential contributor to a person’s well-being. Accordingly, it is reasonable to assume that awareness has a positive influence on all aspects of an individual’s life so that living with a higher level of awareness can improve their general well-being (physical, psychological, emotional, as well as economic). Therefore, we assume the reward function to be positively affected by the current level of awareness, such that the higher the level of awareness of an individual, s_t , the higher their general well-being. The reward function also depends on the reasoning propensity. Rational/analytical reasoning is resource-consuming because it requires the acquisition of information about the problem and the possible solutions, and it is time-consuming because time is needed to sort, analyse, and elaborate the collected data. Therefore, the more analytical a decision is, the more resources are needed in terms of time, personal energy, and monetary resources. This translates into a negative dependence of the reward function on choice, a_t : the higher a_t , the more analytical the DM’s reasoning and the more their resources are consumed. Our instantaneous reward is, therefore, a stochastic function involving both the current state and the next state the DM may reach, with a given probability, by choosing an action. Specifically, in the formulation $r_t(s_t, a_t)$ we make explicit the dependence of the reward function on the state s_t and the action a_t , while it also involves some stochastic component.

The long-run discount factor δ is the current value of an expected reward, i.e., it is the weight the individual assigns to a future reward when they have a non-null probability to transition to state s' from state s as a consequence of choice a . We assume that the discount factor is $\delta \in [0, 1]$. When $\delta = 0$ the future is not considered. Otherwise, the higher the value of δ , the higher the weight to future rewards given, until the case $\delta = 1$, in which there is no distinction between present and future values.

2.1. Exogenous Present Bias

The main purpose of this paper is to generalize the above setup by introducing present bias, signalling the presence of self-control problems, and analysing the impact of both an exogenous and a state-dependent present bias on the evolution of the state. Most experimental research aimed at quantifying present bias has focused on a single specific reward type, namely money. However, estimates of β in other domains, such as real effort [30], and healthy and unhealthy foods [31] also exist. Experimental estimates of the present-bias parameter range from 0.69 for healthy food and 0.71 for unhealthy food [31], to 0.65 for money, to 0.88 for real effort [30]. A recent meta-analysis of empirical estimates of present bias measured with the experimental method called the Convex Time Budget, [32] reports an average present bias between 0.95 and 0.97, but finds that estimates differ depending on the type of reward. Specifically, while monetary reward estimates are close to one, studies with non-monetary rewards produce an average present bias of 0.88.

Definition 7. *The factor $\beta \in (0, 1]$ embeds a taste for immediate gratification, i.e., a present bias affecting the DM’s evaluation of the future stream of rewards.*

The objective function the DM wants to maximize is given by the sequence of expected rewards incurred so that at each time instant t the DM’s maximization problem is

$$\max_{a_t \in A} (r_t(s_t, a_t) + \beta \sum_{\tau=t+1}^T \delta^\tau [r(s_\tau, s_\tau + z, a_\tau)P^F(a_\tau + r(s_\tau, s_\tau, a_\tau)P^S(a_\tau) + r(s_\tau, s_\tau - z, a_\tau)P^B(a_t))], \tag{4}$$

considering the state dynamics defined in (2). The second part of the equation exploits the expected value of the future rewards.

First, the presence of an expected value is due to the presence of a source of external uncertainty, making the evolution of the state stochastic. Second, δ , the weight given to future rewards, is subject to an exponential decrease (because it is smaller than 1), whereas

β is constant for all future time periods but is not considered in the current time period t in which the decision is made.

The maximization problem is solved with a Dynamic Programming algorithm [33] exploiting a mechanism of backward induction in which, starting from a fixed terminal value, $r_T(s_T)$, the sequence of all optimal decisions is reconstructed step by step for the entire time horizon.

2.2. Endogenous Present Bias

With exogenous present bias, the lower the value of β , the stronger the bias, the worse the state. However, it is reasonable to think of personal awareness as a mechanism to contrast and mitigate individual present bias. A higher level of awareness could weaken self-control problems, leading to a higher β . In other words, the higher the DM's level of self-awareness, the lower the negative impact of present bias on the system's dynamics. We assume β to be a function of the individual's current level of self-awareness, which acts as a deterrent of self-control problems. This assumption translates into a state-dependent $\beta(s_t)$ such that

$$\beta(s_t) = s_t \bar{\beta} + (1 - s_t) \underline{\beta}. \tag{5}$$

$\beta(s_t)$ is now a linear combination of s_t and its complement to one, $(1 - s_t)$, weighted by the two constants $\bar{\beta} > \beta$. They are both in the range $(0, 1]$ and correspond to weak and severe present bias, respectively, (for example, in the simulations in Section 3 they are set to $\bar{\beta} = 0.9$ $\underline{\beta} = 0.2$). The final value of β is not constant over time, but it is allowed to evolve according to the state level of the individual. From (5), we see that at $s_t = 0$ the value of β coincides with $\underline{\beta}$, corresponding to a maximum level of bias, whereas at $s_t = 1$, the value of β coincides with $\bar{\beta}$, a minimum level of bias. In the range between these two values, the higher the state, the higher the resulting value of β (the weaker the degree of present bias). By modifying their evaluation of the present/future trade-off in terms of well-being, the agent affects the entire representation of their future rewards, which includes both a present bias parameter and a long-run discount parameter.

The idea of an endogenous present bias is not new. In Economics, [14] argued that economic agents have the power to affect their discount factor, δ , through various means, such as investing in education. More recently, [34,35] allow the agent to exert some control over the magnitude of their present bias by investing time, effort, and resources in human capital which allows them to affect the future well-being implied by current actions. In our setup, we allow the present bias parameter to depend on self-awareness so that the agent has some control over its magnitude.

3. Simulations

In this Section we show, by means of numerical simulations (parameter setting is detailed in Appendix A) of the model, the impact on the state dynamics of both an exogenous and a state-dependent present-bias parameter, β . As mentioned before, the state variable is a representation of the level of awareness of the individual, and, according to [15], it displays monotonic increasing dynamics. The baseline is the simulation of the original model with no present bias, which is equivalent to fixing $\beta = 1$ in Equation (4).

We then introduce the present-bias parameter $0 < \beta \leq 1$, considering an intuitive ($p_r = 0.2$) and an analytical ($p_r = 0.8$) individual, respectively.

Figure 1 shows how the introduction of exogenous present bias has a slowdown effect on the state dynamics for both intuitive and analytical individuals. The effect is stronger for analytical individuals for whom even the increasing monotonic characteristic of the state is lost. When β is set below a threshold, the state eventually falls to zero. The state dynamics of the intuitive individual also display a slowdown effect, though a less prominent one which only smooths the growth speed of the state.

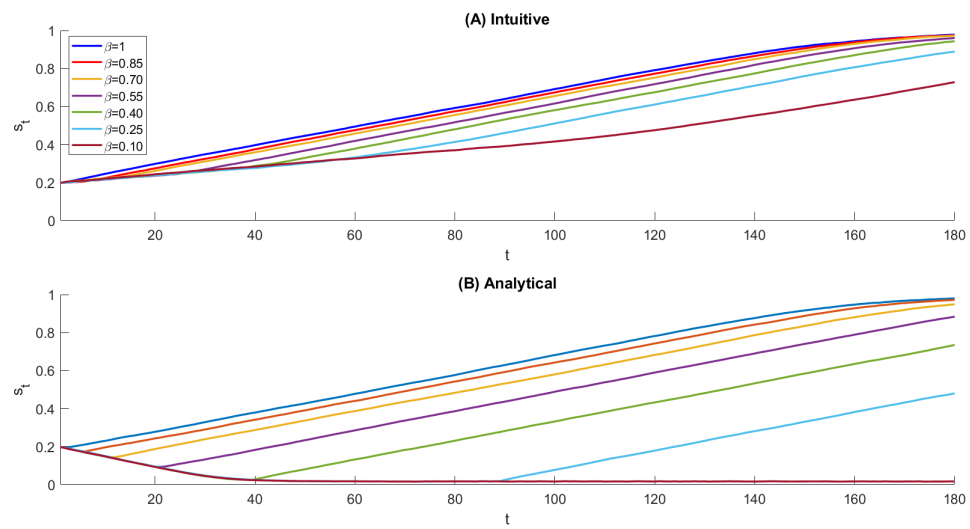


Figure 1. State dynamics, exogenous present bias. The figure shows state dynamics at different levels of present bias for an intuitive (A) and an analytical (B) DM. In both cases, the initial state is low ($s_0 = 0.20$), while the values of β vary in the interval $(0, 1]$ with a step of 0.15.

With an endogenous present bias, there is, in the long run, an improvement in all cases (Figure 2A,B, where the red lines are always above the blue ones) in terms of growth speed or state values. The state dynamics when β is exogenous and extremely low (Figure 2A, blue line, $\beta = 0.2$), is now overcome. The initial decreasing state dynamics in Figure 2A (red line), can be interpreted as a learning period during which the individual has not yet found ways of addressing their self-control problems.

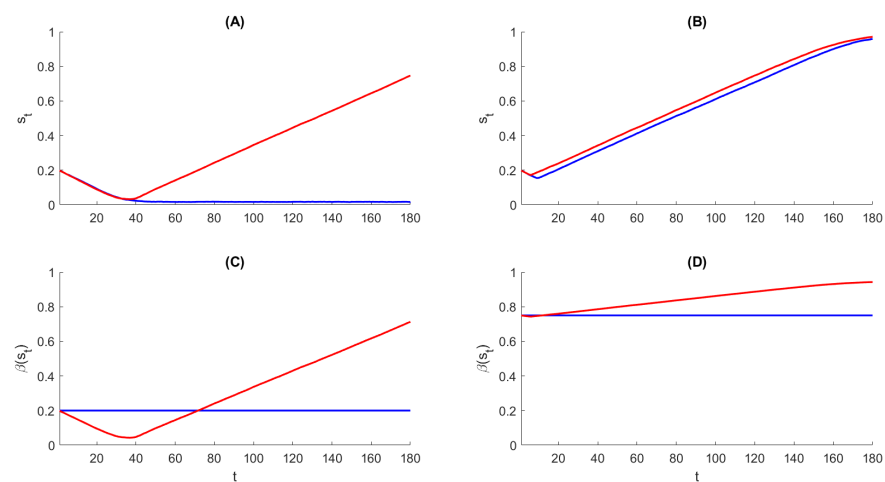


Figure 2. State dynamics (A,B) and present bias dynamics (C,D), analytical individual. Panels (A) and (B) show the state evolution for an analytical individual, $p_r = 0.8$, in the case of endogenous present bias (red lines) and exogenous present bias (blue lines) for low and high initial levels of the parameters, respectively. In Panel (A), $\bar{\beta} = 0.95$, $\beta = 0.0125$, and $\beta = 0.2$. In Panel (B), $\bar{\beta}$ is the same as Panel (A), while $\beta = 0.7$ and $\beta = 0.75$. Panels (C) and (D) show, instead, the dynamics of endogenous present bias (red lines) and exogenous constant present bias (blue lines). In particular, the parameter setting for Panels (C) and (D) is the same as Panels (A) and (B), respectively.

Panels C and D in Figure 2 show the endogenous present bias, $\beta(s_t)$, dynamics over time compared to a constant β at low (Panel C) and high (Panel D) initial levels of the parameters. In Panel C, after an initial decreasing phase, $\beta(s_t)$ is monotonously driven towards higher values corresponding to weaker present bias.

Figure 3 shows the dynamics of an intuitive individual. The introduction of an endogenous present bias accelerates the state dynamics, especially when β is extremely low (0.2). As before, Panels C and D refer to the evolution of $\beta(s_t)$ over time compared to a constant β .

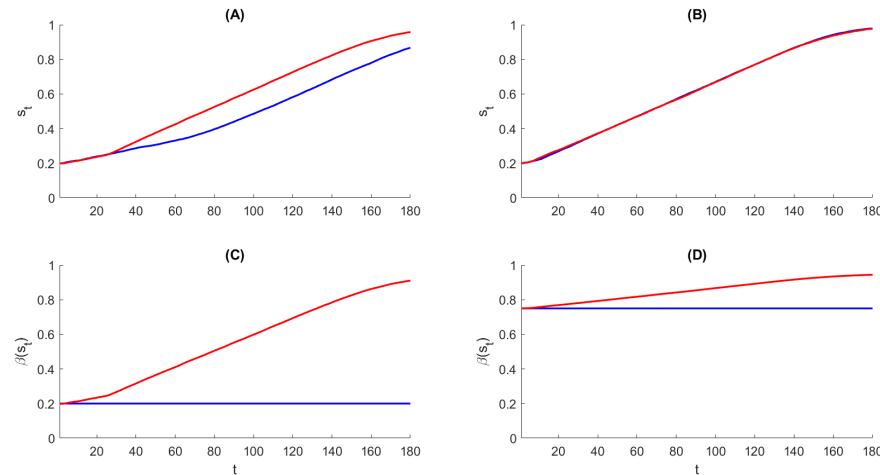


Figure 3. State dynamics (A,B) and present bias dynamics (C,D), intuitive individual. Panels (A) and (B) show the state evolution for an intuitive individual, $p_r = 0.2$, in the case of endogenous present bias (red lines) and exogenous present bias (blue lines) for low and high initial levels of the parameters, respectively. In Panel (A), $\bar{\beta} = 0.95$, $\beta = 0.0125$, and $\beta = 0.2$. In Panel (B), $\bar{\beta}$ is the same as Panel (A), while $\beta = 0.7$ and $\beta = 0.75$. Panels (C) and (D) show, instead, the dynamics of endogenous present bias (red lines) and exogenous constant present bias (blue lines). In particular, the parameter setting for Panels (C) and (D) is the same as Panels (A) and (B), respectively.

4. Discussion

Varying the degree of present bias in the original setup produces the expected dynamics for an intuitive individual’s reward; increasing β (i.e., decreasing the severity of present bias) has a positive effect on the evolution of the state (Figure 1). With an analytical reasoning propensity we observe counter-intuitive dynamics, particularly at low values of β . While a more analytical reasoning propensity should lead to better decisions and outcomes, we observe, instead, a drop in the state variable. The drop is more severe when β is smaller. When β takes on its lower bound the state dynamics falls to zero. As counter-intuitive as it may seem, this dynamic is suggestive to the co-existence of extremely costly decision-making and severe self-control problems, which eventually lead to negative outcomes. Past decision research has indeed pointed towards the downsides of overthinking [36,37] and at how, e.g., careful analysis of all available options might be taken too far. Here, such costly decision-making interacts with severe present bias.

In Figure 2, we compare the state dynamics for low (Panel A) and high (Panel B) values of β (blue line) and $\beta(s_t)$ (red line) for individuals with an analytical reasoning propensity. When β is high and present bias is weak (Panel B), the dynamic is very similar in the two cases. However, when β is low and present bias is severe (Panel A), learning to reduce one’s own self-control problems through self-awareness may trigger an increase in well-being compared to the case in which the degree of present bias is constant. Panels C and D in Figure 2 show, instead, the evolution of β . The learning process triggered by self-awareness is especially beneficial when present bias is severe (Panel C), while it does not play a crucial role in well-being when present bias is already negligible (Panel D).

Finally, Figure 3 depicts the case of individuals with an intuitive reasoning propensity. As in Figure 2, the state dynamics are presented in Panels A and B, while Panels C and D show the present bias dynamics. When decision-making is driven by a more intuitive reasoning propensity, the interaction with severe present bias (Panels A and C) is not conducive to the downward trends observed in Figure 2 for the first 40 time periods.

5. Conclusions

In recent years, decision-making researchers in a variety of settings have introduced self-awareness as a crucial ingredient that may boost optimal decisions: [38] in consumer decision-making; [39,40] in leadership; [41] in a consumption-savings environment. In addiction models, for example, being unaware or aware of one’s self-control problems (a consequence of present bias) determines whether or not people will consume more of an addictive product (such as cigarettes or alcohol) than they would like to consume from a long-run perspective [42].

In this paper, we generalize a Markov Decision Process Model, proposed in Psychological research to study the interaction between analytic and intuitive decision-making, by introducing present bias, a phenomenon that causes self-control problems in the form of an impulse for immediate gratification. We allow for both an exogenous present bias and for an endogenous, state-dependent, present bias and explore the consequences on well-being emerging from the solution of the model. One interesting finding is that self-awareness, implied by the model, may mitigate endogenous present bias and suboptimal choice. The implication is that individuals can exert some control over the magnitude of their present bias and of their time inconsistency. Our model could in principle explain some of the heterogeneity in behaviour and in present bias observed in quantitative estimates of discounting [43]. One limitation of this research is that we do not account for preferences and budget constraints of the underlying individuals, nor for other ingredients of economic models. Therefore, we cannot make any specific behavioural prediction. One interesting and natural avenue for future research is to adapt our setup to study the consumption behaviour of addictive goods.

Author Contributions: Formal analysis, F.B., C.M. and S.T.; Investigation, F.B., C.M. and S.T.; Writing—original draft, F.B., C.M. and S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The following parameter settings are used in the simulations.

$$s_{t+1} = \alpha_b s_t - \alpha_c a_t, \text{ where } \alpha_b, \alpha_c > 0. \tag{A1}$$

Equation (A1) describes the DM’s reward at time t. This is linearly dependent on the DM’s current state with a positive coefficient α_b (set to 10) and negatively dependent on the level of analyticity adopted in solving the decision problem with a positive coefficient α_c (set to 2).

As said before, w_t can take three different values z , 0 or $-z$ depending on forward, stationary, and backward probabilities, $P^F(a_t)$, $P^S(a_t)$, and $P^B(a_t)$. The forward probability results from the linear combination of the two functions $P^A(a_t)$ and $P^I(a_t)$, representing the individual propensity for analytic or intuitive reasoning [27].

$$P^A(a_t) = \frac{(\bar{a}_t - a) + b}{(\bar{a}_t - a)^2 + c} + d, \tag{A2}$$

$$P^I(a_t) = \frac{(\bar{a}_t) + b}{(\bar{a}_t)^2 + c} + d. \tag{A3}$$

Their functional forms are described by Equations (A2) and (A3), where $\bar{a}_t = 10(1 - a_t)$ is introduced to resize the variable a_t so that it belongs to the set $A = (0, 1)$, and to allow

the propensities $P^A(a_t)$ and $P^I(a_t)$ to lay in the interval $[0, 1]$. The propensity to intuitive reasoning is shaped similarly to the one proposed by [27] in their experimental research. $P^A(a_t)$ and $P^I(a_t)$ are combined using as coefficient the reasoning propensity p_r , thus giving rise to Equation (A4).

$$P^F(a_t) = p_r P^A(a_t) + (1 - p_r) P^I(a_t). \quad (\text{A4})$$

The stationary transition probability is assumed to be constant (Equation (A5))

$$P^S(a_t) = k. \quad (\text{A5})$$

Finally, the backward transition probability in Equation (A6) is the difference between the previous two.

$$P^B = 1 - (P^F(a_t) + P^S(a_t)) \quad (\text{A6})$$

The chosen values of the initial state s_0 and reasoning propensity p_r used in the simulations are detailed in Section 3. All other parameters are set to the following values in all simulations: $a = 1$, $b = 7$, $c = 10$, $d = 0.1$, $k = 0.1$, and $z = 0.1$.

References

- Samuelson, P.A. A Note on Measurement of Utility. *Rev. Econ. Stud.* **1937**, *4*, 155–161. [CrossRef]
- Strotz, R. Myopia and Inconsistency in Dynamic Utility Maximization. *Rev. Econ. Stud.* **1956**, *23*, 165–180. [CrossRef]
- Ainslie, G. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*; Cambridge University Press: Cambridge, UK, 1992.
- Loewenstein, G.; Elster, J. *Choice Over Time*; Russell Sage: New York, NY, USA, 1992.
- Phelps, E.S.; Pollak, R.A. On Second-Best National Saving and Game-Equilibrium Growth. *Rev. Econ. Stud.* **1968**, *35*, 185–199. [CrossRef]
- Laibson, D. Hyperbolic Discounting and Consumption. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1994.
- Laibson, D. Golden Eggs and Hyperbolic Discounting. *Q. J. Econ.* **1997**, *62*, 443–478. [CrossRef]
- O'Donoghue, T.; Rabin, M. Doing it Now or Later. *Am. Econ. Rev.* **1999**, *114*, 102–124. [CrossRef]
- O'Donoghue, T.; Rabin, M. Incentives for Procrastinators. *Q. J. Econ.* **1999**, *114*, 769–816. [CrossRef]
- O'Donoghue, T.; Rabin, M. Choice and Procrastination. *Q. J. Econ.* **2001**, *116*, 121–160. [CrossRef]
- O'Donoghue, T.; Rabin, M. *Addiction and Present Biased Preferences*; Working Paper Department of Economics; University of California at Berkeley: Berkeley, CA, USA, 2002; E02-312.
- Pollak, R. Consistent Planning. *Rev. Econ. Stud.* **1968**, *35*, 201–208. [CrossRef]
- O'Donoghue, T.; Rabin, M. Self-awareness and self-control. In *Now or Later: Economic and Psychological Perspectives on Intertemporal Choice*; Baumeister, R., Loewenstein, G., Daniel, R., Eds.; Russell Sage Foundation Press: New York, NY, USA, 2003.
- Becker, G.; Mulligan, C. The Endogenous Determination of Time Preference. *Q. J. Econ.* **1997**, *112*, 729–758. [CrossRef]
- Bizzarri, F.; Giuliani, A.; Mocenni, C. Awareness: An empirical model. *Front. Psychol.* **2022**, *13*, 1–20. [CrossRef]
- Schneider, W.; Shiffrin, R.M. Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychol. Rev.* **1977**, *84*, 1–66. [CrossRef]
- Schneider, W.; Shiffrin, R.M. Controlled and automatic human information processing: 2. Detection, search, and attention. *Psychol. Rev.* **1977**, *84*, 127–190. [CrossRef]
- Bernheim, D.; Rangel, A. Addiction and Cue-Triggered Decision Processes. *Am. Econ. Rev.* **2004**, *94*, 1558–1590. [CrossRef] [PubMed]
- Loewenstein, G.; O'Donoghue, T. Animal Spirits: Affective and Deliberative Processes in Economic Behavior. Available online: <https://ssrn.com/abstract=539843> (accessed on 5 April 2023).
- Brocas, I.; Carrillo, J.D. Dual-process theories of decision-making: A selective survey. *J. Econ. Psychol.* **2014**, *41*, 45–54. [CrossRef]
- Brockmann, E.N.; Anthony, W.P. Tacit knowledge and strategic decision making. *Group Organ. Manag.* **2002**, *27*, 436–455. [CrossRef]
- Dane, E. Exploring intuition and its role in managerial decision making. *Acad. Manag. Rev.* **2007**, *32*, 33–54. [CrossRef]
- Alberts, H.; Martijn, C.; de Vries, N.K. Fighting self-control failure: Overcoming ego depletion by increasing self-awareness. *J. Exp. Soc. Psychol.* **2011**, *47*, 58–72. [CrossRef]
- Beaulieu-Jones, B.K.; Yuan, W.; Brat, G.A.; Beam, A.L.; Weber, G.; Ruffin, M.; Kohane, I.S. Machine learning for patient risk stratification: Standing on, or looking over, the shoulders of clinicians? *NPJ Digit. Med.* **2021**, *4*, 1–6. [CrossRef]
- Kahneman, D. *Thinking, Fast and Slow*; Farrar, Strauss and Giroux: New York, NY, USA, 2011.
- Daw, N.; Niv, Y.; Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **2005**, *8*, 1704–1711. [CrossRef]

27. Moerland, T.M.; Deichler, A.; Baldi, S.; Broekens, J.; Jonker, C.M. Think Neither Too Fast Nor Too Slow: The Computational Trade-off Between Planning And Reinforcement Learning. In Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS), Nancy, France, 16–20 October 2020.
28. Allinson, C.; Hayes, J. The cognitive style index: A measure of intuition-analysis for organizational research. *J. Manag. Stud.* **1996**, *33*, 119–135. [[CrossRef](#)]
29. Klussman, K.; Curtin, N.; Langer, J.; Nichols, A. The Importance of Awareness, Acceptance, and Alignment With the Self: A Framework for Understanding Self-Connection. *Eur. J. Psychol.* **2022**, *18*, 120–131. [[CrossRef](#)]
30. Augenblick, N.; Niederle, M.; Sprenger, C. Working over time: Dynamic inconsistency in real effort tasks. *Q. J. Econ.* **2015**, *130*, 1067–1115. [[CrossRef](#)]
31. Cheung, S.; Tymula, A.; Wang, X. Present bias for monetary and dietary rewards. *Exp. Econ.* **2022**, *25*, 1202–1233. [[CrossRef](#)]
32. Taisuke, I.; Rutter, T.A.; Camerer, C.F. A meta-analysis of present-bias estimation using convex time budgets. *Econ. J.* **2021**, *131*, 1788–1814.
33. Bellman, R.E.; Dreyfus, S.E. *Applied Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 2015; Volume 2050.
34. Galperti, S.; Strulovici, B. *From Anticipations to Present Bias: A Theory of Forward-Looking Preferences*; Working Paper; Northwestern University: Evanston, IL, USA, 2014.
35. Galperti, S.; Strulovici, B. *Anticipations and Endogenous Present Bias*; Discussion Paper; Kellogg School of Management: Evanston, IL, USA, 2015; Volume 1582.
36. Ariely, D.; Norton, M. From thinking too little to thinking too much: A continuum of decision making. *Wiley Interdiscip. Rev. Cogn. Sci.* **2011**, *2*, 39–46. [[CrossRef](#)]
37. Nordgren, L.; Dijksterhuis, A. The Devil Is in the Deliberation: Thinking Too Much Reduces Preference Consistency. *J. Consum. Res.* **2009**, *36*, 39–46. [[CrossRef](#)]
38. Goukens, C.; Dewitte, S.; Warlop, L. Me, Myself, and My Choices: The Influence of Private Self-Awareness on Choice. *J. Mark. Res.* **2009**, *46*, 682–692. [[CrossRef](#)]
39. Caldwell, C.; Hayes, L. Self-efficacy and self-awareness: Moral insights to increased leader effectiveness. *J. Manag. Dev.* **2016**, *35*, 1163–1173. [[CrossRef](#)]
40. Higgs, M.; Rowland, D. Emperors with Clothes On: The Role of Self-awareness in Developing Effective Change Leadership. *J. Change Manag.* **2010**, *10*, 369–385. [[CrossRef](#)]
41. Ali, S. Learning Self Control. *Q. J. Econ.* **2011**, *126*, 857–893. [[CrossRef](#)]
42. O'Donoghue, T.; Rabin, M. Addiction and self-control. In *Addiction: Entries and Exits*; Elster, J., Ed.; Russell Sage: New York, NY, USA, 1999.
43. O'Donoghue, T.; Rabin, M. Present Bias: Lessons learned and to be learned. *Am. Econ. Rev. Pap. Proc.* **2015**, *105*, 273–279. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.