Overcompensating for Present Bias: A Note on Meta-Cognitive Adjustment in Intertemporal Choice

Yaakov Bayer *

Posted Date: 6 May 2025

doi: 10.20944/preprints202505.0215.v1

Keywords: Intertemporal choice

 doi:10.20944/preprints202505.0215.v1

Article

Overcompensating for Present Bias: A Note on Meta-Cognitive Adjustment in Intertemporal Choice

Ya'akov M. Bayer

Ben Gurion University of the Negev; ymebayer@gmail.com

Abstract: This paper presents a theoretical model of meta-cognitive adjustment in intertemporal choice. We consider agents who are aware of their present bias—captured by quasi-hyperbolic discounting—and attempt to correct for it by applying a self-chosen adjustment parameter, λ. While modest correction improves utility, we show that overcorrection, where the agent becomes excessively future-oriented, may result in outcomes inferior to those under the original biased baseline. The model characterizes optimal correction as occurring when the agent faithfully incorporates their time preferences. Attempts to fully "cancel out" present bias (e.g., by setting $\lambda=1/\beta$) often yield second-order inefficiencies due to nonlinear interactions between discounting and utility. These results challenge the common assumption that bias awareness necessarily enhances welfare. We discuss implications for behavioral interventions and policy design, emphasizing the need for calibrated support rather than rigid rationalization. Our findings highlight that self-awareness is necessary but not sufficient for improved intertemporal decision-making.

Keywords: intertemporal choice

1. Introduction

Intertemporal choice has long been central to economic theory, traditionally modeled through exponential discounting, which assumes consistent preferences over time. However, a substantial body of empirical and theoretical research has demonstrated that individuals systematically depart from this framework. In particular, *present bias*—the tendency to overweight immediate rewards relative to future ones—has emerged as one of the most robust deviations from the rational actor model (Laibson, 1997; Frederick, Loewenstein, & O'Donoghue, 2002).

Quasi-hyperbolic discounting models, which introduce a parameter $\beta<1$ to capture present bias alongside a standard exponential discount factor δ, have successfully accounted for behaviors such as procrastination, under-saving, and health-related inaction (Bayer & Osher, 2018; O'Donoghue & Rabin, 1999). In this framework, individuals may display dynamically inconsistent preferences, favoring immediate gratification despite long-term costs. Sophisticated agents—those aware of their future time inconsistency—may take preemptive steps to mitigate its effects (O'Donoghue & Rabin, 2001), for example, by using commitment devices or structuring incentives to align short-term actions with long-term goals (Ariely & Wertenbroch, 2002; Thaler, 1999).

Yet, while sophistication can enable corrective action, recent theoretical and empirical work suggests that the effectiveness of such self-correction depends critically on its calibration. Overly rigid commitment strategies can reduce flexibility and welfare (DellaVigna, 2009), and some commitment devices are misused or applied inappropriately (Karlan, Ratan, & Zinman, 2014). Furthermore, efforts to impose excessive future orientation—motivated by a desire to "act rationally"—may result in distorted decision-making that undermines personal well-being (Duckworth et al., 2019).

In this paper, we develop a simple model in which a present-biased agent is aware of their bias and attempts to correct for it through a self-chosen adjustment parameter. We analyze the utility implications of undercorrection, accurate correction, and overcorrection, showing that excessive adjustment can lead to worse outcomes than the original bias. We derive closed-form results and

 doi:10.20944/preprints202505.0215.v1

extend the model to cases where agents scale their correction based on the perceived severity of their bias. Our findings offer new insights into the limits of behavioral self-awareness.

2. Theoretical Framework

We consider a three-period model, t = 0, 1, 2, in which an agent allocates a fixed resource endowment (normalized to 1) across periods. Preferences follow a quasi-hyperbolic structure. Let consumption in period t be denoted by c_t, and instantaneous utility be $u(c_t) = \ln(c_t)$, with standard concavity.

The agent's present-biased utility is:

$U_0 = u(c_0) + \beta\delta u(c_1) + \beta\delta^2 u(c_2)$

where $\delta \in (0, 1]$ is the exponential discount factor and $\beta \in (0, 1]$ captures present bias.

We introduce a correction parameter $\lambda > 0$, representing the agent's metacognitive adjustment. The agent allocates consumption by maximizing an adjusted utility function:

$$\tilde{U}_0 = u(c_0) + \lambda\beta\delta u(c_1) + \lambda^2\beta\delta^2 u(c_2)$$

subject to the constraint $c_0 + c_1 + c_2 = 1$. Here, $\lambda = 1$ represents accurate self-correction, $\lambda < 1$ reflects undercorrection, and $\lambda > 1$ reflects overcorrection.

Clarifying the Interpretation of $\lambda = 1$:

While $\beta < 1$ reflects a structural deviation from exponential discounting, we interpret $\lambda = 1$ as the agent faithfully integrating their own preferences into decision-making. The agent does not eliminate the bias but makes consistent tradeoffs in light of it. Attempting to "correct" the bias by imposing a different discount structure—e.g., by neutralizing β entirely—risks generating utility losses due to internal inconsistency.

2.1. Solution

We solve the problem by maximizing $\tilde{U}_0$ using Lagrangian methods:

$$L = \ln(c_0) + \lambda\beta\delta lnu(c_1) + \lambda^2\beta\delta^2 lnu(c_2) + \mu(1 - c_0 - c_1 - c_2)$$

First − order conditions yield:

$$\frac{1}{c_0} = \mu, \frac{\lambda\beta\delta}{c_1} = \mu, \frac{\lambda^2\beta\delta^2}{c_2} = \mu$$

Solving gives optimal allocations:

$$c_0 = \frac{1}{S}, c_1 = \frac{\lambda\beta\delta}{S}, c_2 = \frac{\lambda^2\beta\delta^2}{S}$$

where $S = 1 + \lambda\beta\delta + \lambda^2\beta\delta^2$.

Substituting back into the original (true) utility function U_0 gives:

$U_0(\lambda) = -(1 + \beta\delta + \beta\delta^2)\ln S + (\beta\delta + 2\beta\delta^2)\ln\lambda + (\beta\delta + \beta\delta^2)\ln\beta + (\beta\delta + 2\beta\delta^2)\ln\delta.$

Utility U_0 is maximized at $\lambda = 1$. For sufficiently large $\lambda > 1$, utility declines below the level achieved with no correction ($\lambda = 0$).

Proof Sketch: The function $U_0(\lambda)$ is strictly concave in $\ln\lambda$ around $\lambda = 1$ under regularity conditions. Overcorrection increases future weightings nonlinearly, leading to excessive deferred consumption and welfare loss. Full proof omitted for brevity (see Appendix A).

2.2. Extension: Bias-Neutralizing Correction

Suppose the agent attempts to neutralize their bias entirely by setting:

$$\lambda = \frac{1}{\beta}$$

This yields adjusted weights of:

$$\lambda\beta = 1, \lambda^2\beta = \frac{1}{\beta}$$

Substituting into the allocation formulas:

 doi:10.20944/preprints202505.0215.v1

$c_0 = \frac{1}{S'}$, $c_1 = \frac{\delta}{S'}$, $c_2 = \frac{\delta^2/\beta}{S'}$

where $S' = 1 + \delta + \delta^2/\beta$.

Although this correction restores standard exponential discounting for period 1, it overweights period 2 by a factor of $1/\beta$, which may be large. Substituting into the true utility $U_{0,}$ one can show that for low values of β, the utility is lower than at $\lambda = 1$.

Corollary 1: Bias-neutralizing correction ($\lambda = 1/\beta$) may produce lower utility than consistent correction ($\lambda = 1$) for all $\beta < 1$.

3. Discussion

The results of the model presented here suggest that self-awareness of behavioral biases does not inherently lead to better outcomes—indeed, in some cases, it can produce new inefficiencies. This finding adds nuance to the literature on sophisticated agents and behavioral self-control (O'Donoghue & Rabin, 2001; Bénabou & Tirole, 2004), highlighting that the degree of correction plays a critical role in determining welfare.

In practice, agents often face uncertainty not only about their future preferences but also about the *appropriate* adjustment to make once their biases are recognized. While models of sophistication assume internal rationality over biased preferences, our framework shows that agents who act "too rationally"—by imposing a correction that overemphasizes future utility—may violate their own subjective welfare criterion. This insight is particularly relevant in domains such as health, savings, or education, where overplanning or excessive self-discipline can backfire, reducing current well-being or long-run adherence (Duckworth et al., 2019).

The policy implications are twofold. First, interventions aimed at helping individuals overcome present bias, such as nudges, commitment devices, or financial incentives, must be calibrated carefully. A one-size-fits-all correction strategy may be ineffective or even harmful for individuals with varying degrees of bias or self-control capacity. Second, behavioral education that promotes awareness of cognitive distortions should be paired with tools for calibrating correction intensity. Simply telling agents they are biased does not ensure better choices.

Our extension, in which agents apply a correction proportional to the perceived size of their bias (e.g., $\lambda=1/\beta$), reflects a common intuition in behavioral policy: that bias should be canceled out analytically. However, the model demonstrates that such proportional correction can lead to nonlinear overweighting of distant outcomes, due to the interaction between exponential discounting and utility curvature. This highlights the need for flexible, empirically informed behavioral interventions that support gradual and self-consistent improvement rather than sharp rationalization.

4. Conclusion

This paper develops a formal model of metacognitive adjustment in intertemporal decision-making. We show that agents who are aware of their present bias and attempt to correct for it may not always improve their welfare. While moderate correction—faithfully incorporating one's quasi-hyperbolic preferences—leads to optimal outcomes, overcorrection can distort consumption or effort allocation to such an extent that utility falls below the level associated with uncorrected bias.

By introducing a flexible correction parameter λ, we characterize a continuum of behavioral adjustment: from undercorrection (remaining biased), through accurate correction ($\lambda=1$), to overcorrection (becoming overly future-oriented). Surprisingly, attempts to fully neutralize the present bias ($\lambda=1/\beta$) can lower utility when the distant future becomes excessively weighted. This result challenges the implicit assumption that self-awareness always improves outcomes.

The findings emphasize that **behavioral sophistication is not binary**: it is not enough to recognize one's biases—individuals must also accurately calibrate their correction. This has implications for behavioral policy and personal decision support tools. Future research should focus on measuring the distribution of correction behavior in real populations, testing interventions that

 doi:10.20944/preprints202505.0215.v1

aid in adjustment calibration, and designing default mechanisms that promote effective but bounded behavioral correction.

Appendix A. Mathematical and Numerical Analysis

A.1 Proof that Utility is Maximized at λ=1

We begin with the agent's indirect utility function after optimizing over consumption, given quasi-hyperbolic preferences and a self-correction parameter λ:

$$U_0(\lambda) = -(1 + \beta\delta + \beta\delta^2)\ln S + (\beta\delta + 2\beta\delta^2)\ln\lambda + \text{const},$$

where

$$S = 1 + \lambda\beta\delta + \lambda^2\beta\delta^2.$$

We compute the derivative of $U_0(\lambda)$ with respect to λ:

$$\frac{dU_0}{d\lambda} = \frac{\beta\delta}{\lambda} \cdot \frac{-\beta\delta^2\lambda^2 + \beta\delta^2\lambda - 2\delta\lambda^2 + 2\delta - \lambda + 1}{\beta\delta^2\lambda^2 + \beta\delta\lambda + 1}.$$

Substituting λ=1 yields:

$$\left.\frac{dU_0}{d\lambda}\right|_{\lambda=1} = 0.$$

Since the function is strictly concave in lnλ, this critical point corresponds to **a maximum**.

A.2 Clarification: When Overcorrection Can Be Worse than No Correction

When λ=0, the agent allocates all consumption to the present: c_0=1c, c1=c2=0, which yields:

$$U_0(\lambda = 0) = \ln(1) + \beta\delta \cdot \ln(0) + \beta\delta^2 \cdot \ln(0) = -\infty.$$

This shows that zero correction yields extremely low utility. However, when λ is very large (e.g., λ=1/β), the agent overweights the distant future, causing over-deferral of consumption and again reducing utility significantly. The model shows that **both extremes**—no correction and overcorrection—are inefficient, and optimal utility is achieved at a moderate correction level.

A.3 Numerical Illustration

Using the parameter values:

- β=0.7
- δ=0.9

 doi:10.20944/preprints202505.0215.v1

We compute consumption allocations and utility under three values of λ:

λ	C_0	c_1	c_2	$U_0(\lambda)$
$\lambda = 0$	0.9999	0.00006	5.67e-9	–16.86
$\lambda = 1$	0.4552	0.2868	0.2581	–2.34
$\lambda = 1/\beta$	0.3271	0.2944	0.3785	–2.44

At $\lambda = 1$, utility is highest.

At $\lambda = 0$, almost no future utility is considered, leading to utility collapse.

At $\lambda = 1/\beta \approx 1.43$, overcorrection shifts too much utility to the distant future.

Thus, **excessive correction is harmful even when** compared to no correction. This validates our theoretical insight that optimal behavior occurs at calibrated, not extreme, levels of adjustment.

References

1. Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science, 13*(3), 219–224. https://doi.org/10.1111/1467-9280.00441
2. Bayer Y. M., & Osher, Y.(2018). Time preference, executive functions, and ego-depletion: An exploratory study. *Journal of Neuroscience, Psychology, and Economics, 11*(3), 127-134. http://dx.doi.org/10.1037/npe0000092 .
3. Bénabou, R., & Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy, 112*(4), 848–886. https://doi.org/10.1086/421167
4. DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature, 47*(2), 315–372. https://doi.org/10.1257/jel.47.2.315
5. Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-control and academic achievement. *Annual Review of Psychology, 70*, 373–399. https://doi.org/10.1146/annurev-psych-010418-103230
6. Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351–401. https://doi.org/10.1257/002205102320161311
7. Karlan, D., Ratan, A. L., & Zinman, J. (2014). Savings by and for the poor: A research review and agenda. *Review of Income and Wealth, 60*(1), 36–78. https://doi.org/10.1111/roiw.12101
8. Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112*(2), 443–478. https://doi.org/10.1162/003355397555253
9. O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review, 89*(1), 103–124. https://doi.org/10.1257/aer.89.1.103
10. O'Donoghue, T., & Rabin, M. (2001). Choice and procrastination. *Quarterly Journal of Economics, 116*(1), 121–160. https://doi.org/10.1162/003355301556365
11. Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*(3), 183–206.